
LES NOUVEAUX PRINCIPES MÉTHODOLOGIQUES DE L'ENQUÊTE PERMANENTE SUR L'UTILISATION DES VÉHICULES ROUTIERS DE MARCHANDISES

Nathalie CARON (*), Philippe BRION (**)

(*) MEN-DEPP¹

(**) Consultant

nathalie.caron@education.gouv.fr

philippe.brion55@gmail.com

Mots-clés : Échantillonnage, redressement, enquête transport de marchandises, charge statistique

Résumé

L'enquête permanente sur l'utilisation des véhicules routiers de marchandises, régie par un règlement européen, porte sur le transport de marchandises par route réalisé par les poids lourds, qui représente en France près de 90 % du transport de marchandises. La méthodologie de l'enquête française (plan de sondage et redressement) en vigueur aujourd'hui est celle mise en place au début des années 2000. Les unités échantillonnées y sont des poids lourds qui sont sélectionnés par un plan de sondage à probabilités inégales combinant le type de poids lourd, l'âge du véhicule, la charge utile et l'activité de l'établissement. L'échantillon d'une taille globale de 80 000 véhicules est renouvelé par moitié chaque année et est réparti sur l'ensemble des semaines.

Le nouveau plan de sondage repose globalement sur les mêmes principes. Cependant, outre la révision des paramètres déterminés lors de la précédente mise au point de la méthodologie, il prend en compte l'optimisation de la base de sondage de l'enquête en tenant compte du rapprochement de deux bases disponibles (fichiers des immatriculations et fichiers des contrôles techniques), une taille d'échantillon (utilisant la base de sondage optimisée) permettant de conserver une précision globale des résultats du même ordre de grandeur qu'actuellement, ainsi qu'une meilleure prise en compte de la dimension entreprise avec un lissage de la charge statistique, que ce soit au niveau de la charge globale ou de la charge « individuelle », à savoir le nombre de véhicules enquêtés pour une entreprise en possédant plusieurs.

La méthode proposée conduit ainsi à une réduction de la taille d'échantillon, qui est de 80 000 pour passer à 65 000 véhicules enquêtés par an. Par ailleurs, le redressement de l'enquête a été entièrement revu.

Abstract en anglais

The French survey of road freight transport answers to a European regulation, and uses methodological principles which were elaborated at the beginning of the 2000's. These principles have been reviewed, taking into account the experience of the present survey, as the availability of new information, especially a file concerning the motor vehicle controls. The resulting sampling plan

¹Au moment de la rédaction de ce papier, N. Caron travaillait au ministère de la transition écologique et solidaire, Service de la Donnée et des Études Statistiques.

Les auteurs remercient Lise Dervieux, Bruno Lutinier et Germain Marteau qui ont contribué à une partie des travaux présentés dans le papier.

allows a reduction of the size of the sample, as a better distribution of the statistical burden among enterprises. The post-collection treatments, to produce the results, have also been reviewed.

1. Introduction

L'enquête permanente sur l'utilisation des véhicules routiers de marchandises (dite enquête TRM) est la seule source d'information sur le transport de marchandises par route, qui représente en France près de 90 % du transport de marchandises. L'historique de cette enquête remonte à 1952, première année où elle a été réalisée pour aider à la coordination des investissements sur le transport et leur partage entre rail et route (Lutinier, B., 2017). S'inscrivant dans le dispositif européen de connaissance du transport de marchandises depuis 1978 (directive européenne), elle est actuellement régie par le règlement 70/2012 et permet ainsi de connaître les transports routiers effectués par chaque pavillon européen avec leurs caractéristiques. La précision des indicateurs demandée par Eurostat est donnée dans l'article 2 du rapport [Commission Regulation \(EC\) 642/2004](#). Il est ainsi précisé que « *In general, the standard percentage error (95% confidence) of annual estimates for tonnes transported, tonne-kilometres performed and total kilometres travelled loaded shall not be greater than 5 % for total goods road transport and for national goods road transport* ». Les résultats sont attendus par les autorités publiques, pour l'analyse économique et dans le cadre de l'établissement ou de l'évaluation des politiques publiques (implication des flux de marchandises par origine-destination sur la circulation, la sécurité routière, la pollution), ainsi que par les organismes représentatifs de la profession, les universitaires et chercheurs et les organisations internationales (UE, ONU, OCDE).

L'enquête porte sur le transport, pour compte d'autrui et pour compte propre, réalisé par les véhicules utilitaires poids lourds de marchandises immatriculés en France : camions et tracteurs routiers sous certaines limites d'âge et de gabarit. Elle est effectuée par sondage. L'unité échantillonnée est le véhicule et les données collectées correspondent à une semaine donnée (l'échantillon est réparti sur l'ensemble des semaines de l'année et est renouvelé par moitié chaque année). Les poids lourds suivis sont interrogés sur l'ensemble de leurs trajets sur une semaine de référence, que ceux-ci se situent sur le territoire national ou à l'étranger.

La méthodologie en vigueur jusqu'à ces dernières années (base de sondage, plan de sondage et redressement) est celle mise en place au début des années 2000 (Gallais, A., 2002). Elle a été entièrement revue durant l'année 2016. Outre la révision des paramètres déterminés lors de la précédente mise au point de la méthodologie, la recherche du nouveau plan de sondage devait également permettre de prendre en compte les considérations suivantes : l'optimisation de la base de sondage de l'enquête en tenant compte du rapprochement de deux bases (fichiers des immatriculations et fichiers des contrôles techniques), un choix de taille d'échantillon sur la base de sondage optimisée permettant de conserver une précision globale des résultats du même ordre de grandeur qu'actuellement, autorisant ainsi les mêmes utilisations au niveau national et au niveau régional de l'enquête tout en s'assurant qu'elle soit compatible avec la baisse du nombre de gestionnaires de l'enquête, et une meilleure prise en compte de la dimension entreprise avec un lissage de la charge statistique, que ce soit au niveau de la charge globale ou de la charge « individuelle », à savoir le nombre de véhicules enquêtés pour une entreprise en possédant plusieurs.

Les travaux de mise au point du nouveau plan de sondage (estimation de précision, étude du biais pour les extensions éventuelles du champ) nécessitaient d'une part de disposer d'informations sur les variables d'intérêt et d'autre part de données sur les immatriculations fusionnées avec celles des contrôles techniques. Pour le premier point, nous avons utilisé l'échantillon des 58 000 répondants à l'enquête TRM de l'année 2014 avec les pondérations associées. Ensuite, pour estimer la précision du nouveau plan de sondage sur les variables d'intérêt, nous avons aussi utilisé le fichier des résultats de TRM en affectant à chaque véhicule une pondération comme s'il avait été sélectionné avec le nouveau plan de sondage. Pour le second point, nous avons constitué spécifiquement une base de

sondage 2014 à partir des données de contrôles techniques des poids lourds des trois dernières années.

Cet article détaille l'optimisation de la base de sondage et du champ de l'enquête, ainsi que les études préliminaires et le plan de sondage retenu dans les parties 2 et 3. La partie 4 précise la mise en œuvre du plan de sondage et la dernière partie donne les principaux éléments sur les traitements post-collecte de l'enquête.

2. Optimisation de la base de sondage et étude du champ de l'enquête

La base de sondage utilisée jusqu'à présent est le répertoire statistique des véhicules routiers (RSVERO), qui est alimenté à partir des informations transmises par le ministère de l'intérieur et issues du système d'immatriculation des véhicules (informations contenues sur la carte grise). Les véhicules sortis du parc (détruits ou vendus à l'étranger) ne sont pas forcément tous bien indiqués comme tels dans ce fichier ; ainsi sur un échantillon actuel de près de 80 000 véhicules sélectionnés par an, 58 000 sont « répondants » mais parmi ceux-ci 23 000 sont déclarés « non roulants » (détruit, vendu à l'étranger, panne, période de congé de l'entreprise...). On peut se demander si l'utilisation d'autres informations disponibles ne pourrait pas permettre d'avoir une base de sondage plus efficace.

Par ailleurs, le champ de l'enquête, jusqu'à présent, est basé sur l'ensemble des véhicules appartenant aux entreprises utilisatrices de poids lourds (pour compte d'autrui et/ou pour compte propre). A ce titre, il inclut des entreprises loueuses de véhicules, qui indiquent alors l'utilisateur du véhicule à enquêter. Seuls les poids lourds (camions et tracteurs routiers) immatriculés en France métropolitaine de 15 ans et moins entrent dans le champ de l'enquête². Une extension du champ, aux véhicules plus âgés ou à ceux des DOM, est à étudier.

2.1. Utilisation des fichiers de contrôle technique pour optimiser au mieux le fichier du parc roulant

Le nouveau processus RSVERO opérationnel fin 2018 intégrera l'appariement, au niveau des véhicules, des données des contrôles techniques avec celles transmises par le ministère de l'intérieur. Ainsi, la confrontation de ces deux informations permet de « nettoyer la base de sondage » et d'améliorer sa qualité en écartant de celle-ci les véhicules ne circulant plus : plus précisément, les véhicules absents pendant trois années du fichier des contrôles techniques seront considérés comme cessés, sachant que les poids lourds doivent passer un contrôle technique chaque année. La confrontation des données des enquêtes TRM récentes avec le fichier du contrôle technique a confirmé la pertinence de cette règle : les véhicules détectés comme cessés d'après leur absence de contrôle technique étaient également mentionnés comme tels dans les réponses à l'enquête.

Ceci a permis d'éliminer de l'ordre de 30 % des véhicules de la base de sondage utilisée jusqu'à présent, pour les véhicules de 15 ans ou moins, passant ainsi d'un effectif d'environ 582 000 à 410 000 (voir *tableau 1*) ; concernant l'inclusion des véhicules de 16 à 25 ans, le nettoyage de la base de sondage via les données du contrôle technique s'avère encore plus efficace : plus de 75 % des véhicules de 16 à 25 ans, principalement des tracteurs routiers, sont éliminés, passant ainsi de 221 000 véhicules à 50 000.

Tableau 1 : Effectifs du parc et du parc roulant (avec confrontation avec les données de contrôle technique) des poids lourds au 1/1/2015

Moins de 25 ans - champ métropole

² Concernant l'âge, il est nécessaire de définir précisément la définition choisie : le critère retenu dans le plan de sondage pour les véhicules correspond à l'âge en différence de millésime entre l'année de réalisation de l'enquête et l'année de la première immatriculation ; l'âge reste donc constant durant toute l'année de réalisation de l'enquête, indépendamment du jour d'anniversaire de la mise en circulation.

	RSVERO1 au 01/01/2015	Parc roulant 2014
Camions	407 528	270 517
Tracteurs routiers	401 422	189 615
	808 950	460 132

15 ans ou moins - champ métropole

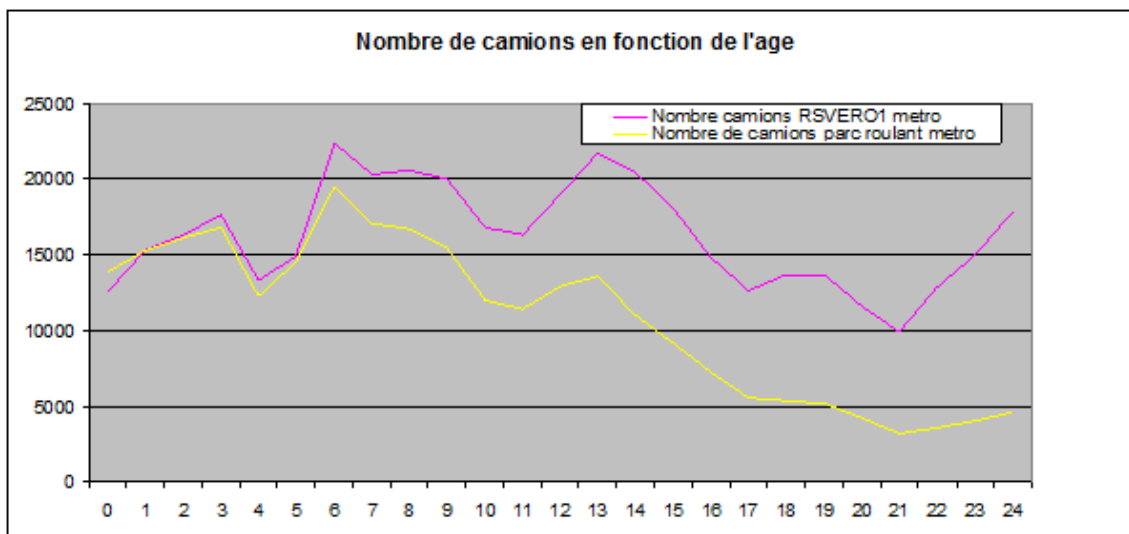
	RSVERO1 au 01/01/2015	Parc roulant 2014
Camions	285 670	270 517
Tracteurs routiers	302 209	189 615
Total	581 879	410 188

De 16 ans à moins de 25 ans - champ métropole

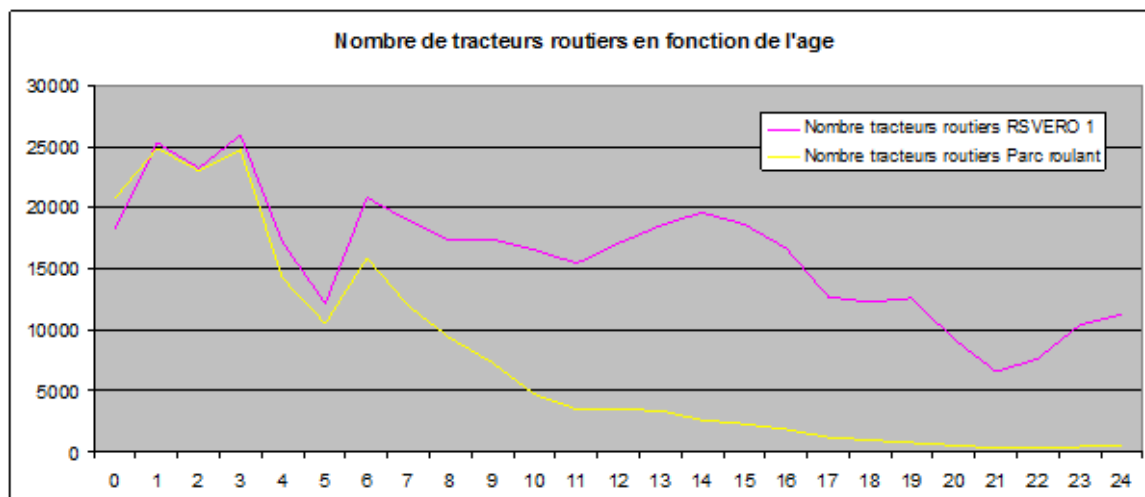
	RSVERO1 au 01/01/2015	Parc roulant 2014
Camions	121 858	43 114
Tracteurs routiers	99 213	6 830
Total	221 071	49 944

A partir de 6 ans, on commence à voir une différence entre le parc et le parc roulant, celle-ci ne fait qu'augmenter en fonction du vieillissement des véhicules. Le décrochage est beaucoup plus net pour les tracteurs routiers (voir *graphiques 2 et 3*).

Graphique 2 : Répartition des camions du parc et du parc roulant selon l'âge (champ métropole)



Graphique 3 : Répartition des tracteurs routiers du parc et du parc roulant selon l'âge (champ métropole)



L'utilisation des fichiers de contrôles techniques conduit donc à une avancée majeure dans la constitution de la base de sondage : la base est ainsi optimisée en éliminant les poids lourds ne circulant plus.

2.2. Étude du champ de l'enquête : prise en compte ou non des DOM, des véhicules de 16 à 25 ans et de l'exclusion ou non de certaines entreprises en fonction de leur activité

La refonte de la méthodologie a aussi été l'occasion de revoir le champ de l'enquête.

2.2.1. Interrogation sur l'exclusion ou non de certaines catégories dans la base de sondage

Certaines entreprises se plaignent et/ou répondent « ne pas être concernées par l'enquête ». Nous avons donc examiné la question de l'inclusion, ou non, de certains types d'entreprises en fonction de leur activité principale exercée (codes APE) dans le champ de l'enquête (en particulier, les entreprises de commercialisation de véhicules, camion-écoles, etc.) en regardant à partir des réponses à l'enquête TRM de 2014, pour chaque code APE de l'entreprise possédant le véhicule, le nombre de véhicules enquêtés dans TRM et le nombre de ceux qui ont déclaré avoir roulé à l'enquête. Ainsi, par exemple, le pourcentage de véhicules pour la division 45 de la NAF (commerce automobile) ayant roulé selon l'enquête est de 14 %, donc sensiblement différent de zéro et ne justifiant pas, à l'heure actuelle, l'exclusion de l'ensemble des véhicules appartenant à des entreprises situées dans ce code NAF. Les résultats obtenus conduisent donc à être prudents sur l'élimination possible de certains codes APE du champ de l'enquête. La seule exclusion proposée concerne le code APE 84.11Z (administration publique générale) du champ de l'enquête, le règlement européen ne demandant pas de le prendre en compte.

2.2.2. Extension ou non du champ de l'enquête aux DOM et aux véhicules de 16 à 25 ans

L'extension du champ aux véhicules de 16 à 25 ans ainsi que la collecte dans les DOM ont été analysées dans le cadre de la rénovation méthodologique de l'enquête. Les effectifs globaux du parc et du parc roulant des poids lourds en métropole et dans les DOM pour les moins de 25 ans sont précisés dans le tableau 4.

Tableau 4 : Effectifs du parc et du parc roulant des poids lourds en métropole et dans les DOM pour les moins de 25 ans

Métropole moins de 25 ans	RSVERO1 au 01/01/2015	Parc roulant 2014
Camions	407 528	270 517
Tracteurs routiers	401 422	189 615

DOM moins de 25 ans	RSVERO1 au 01/01/2015	Parc roulant 2014
Camions	11 819	6 679
Tracteurs routiers	3 069	1 730
Part DOM	1,80%	1,80%

L'extension du champ de l'enquête peut avoir deux objectifs distincts :

- meilleure prise en compte de la population concernée par l'enquête pour la production de résultats agrégés ;

- production de résultats spécifiques sur des sous-populations ; si l'on peut penser que ce n'est pas spécifiquement le cas pour la population des véhicules de plus de 15 ans, il y a en revanche de fortes présomptions qu'il y ait une demande ciblée concernant la production d'estimations sur les DOM, voire sur chacun des départements les constituant.

Ces deux objectifs peuvent conduire à des stratégies très différentes en matière de taille d'échantillon.

Si l'on ne s'intéresse qu'au premier objectif, l'apport des extensions (inclusion ou non dans le champ de l'enquête des DOM et les véhicules de 16 à 25 ans) s'apprécie en comparant l'erreur quadratique moyenne obtenue pour l'estimation du nombre de tonnes-kilomètres en sélectionnant dans les deux bases de sondage (actuelle et étendue au nouveau champ) un échantillon de même taille.

Pour les deux extensions envisagées, l'erreur quadratique moyenne sur la base de sondage étendue est plus faible que celle de la base réduite (voir *annexe 1*). La perte de variance due à l'extension du champ serait compensée par l'annulation du biais dû à la non-enquête d'une partie des véhicules. En conséquence, une extension du champ couvert, à taille d'échantillon équivalente, permet une amélioration de la qualité des résultats. Les calculs réalisés légitiment l'extension du champ à la fois aux DOM et aux véhicules de plus de quinze ans.

Mais, concernant les DOM, on se heurte au deuxième objectif mentionné précédemment, à savoir la volonté de produire des résultats spécifiques sur une sous-population (sur l'ensemble des DOM, voire sur chacun d'entre eux) ; dans cette perspective de produire des résultats spécifiques, il serait nécessaire de gonfler la taille de l'échantillon dédié aux DOM, pour passer par exemple à une taille de 15 000 véhicules afin d'avoir des résultats fiables (alors que le nombre de véhicules concernés par l'enquête est égal, pour les DOM, à 1,8 % de l'ensemble de véhicules dans la base de sondage).

Une telle taille d'échantillon conduirait donc à un taux de sondage moyen beaucoup plus élevé dans les DOM. Ceci est contraire aux objectifs de refonte de l'enquête, qui mettent en avant la réduction de la taille de l'échantillon ; de plus, même avec cette taille d'échantillon, on pourrait avoir des résultats de qualité moindre, du fait qu'on réaliserait dans cette optique l'enquête pour la première fois dans les DOM, ce qui pourrait y conduire à des taux de non-réponse plus élevés.

En conséquence, et également parce que le fait de ne pas enquêter les DOM pèse pour moins de 0,4 % d'un agrégat comme le total tonnes-kilomètres, les DOM ne seront pas inclus dans le champ de l'enquête. En revanche, l'estimation produite pour le biais dû à la non-enquête des véhicules de 16 à 25 ans (de l'ordre de 2,7 %) légitime l'extension du champ à cette population.

En résumé, le champ retenu de l'enquête est constitué des camions et tracteurs routiers appartenant aux entreprises utilisatrices de poids lourds (camions et/ou tracteurs routiers), hors administration publique (secteur 84.11Z), ce secteur n'étant pas demandé dans la réglementation européenne. Seuls les poids lourds immatriculés en France métropolitaine entrent dans le champ de l'enquête, les départements d'outre-mer étant exclus. Les véhicules de 26 ans et plus dans l'année d'interrogation sont également exclus du champ.

3. Études préliminaires et plan de sondage retenu

Les véhicules sélectionnés actuellement dans la base de sondage le sont par un tirage à probabilités inégales combinant le genre (camion/tracteur routier), l'âge du véhicule, sa charge utile (utilisation du poids total autorisé en charge - PTAC - pour les camions et du poids total roulant autorisé - PTR - pour les tracteurs routiers) et l'activité de l'établissement avec prise en compte de spécialisation éventuelle (transport international, transports marchandises dangereuses).

3.1. Étude des critères potentiels pour une stratification

A partir des résultats de l'enquête TRM 2014, nous avons regardé la répartition des véhicules pour lesquels on a une réponse en distinguant les camions des tracteurs routiers avec en abscisse les

différentes variables qui interviennent (âge, capacité, activité...) dans la définition des probabilités inégales du plan de sondage actuel et en ordonnée la distance parcourue ou la charge transportée qui correspondent aux principales variables d'intérêt de l'enquête.

3.1.1. Age

On constate très nettement une décroissance quasi linéaire entre la variable âge et la distance parcourue exprimée en logarithme, le lien est moins net entre l'âge et la charge. L'âge ainsi que le genre de poids lourd constituent donc des critères essentiels à prendre en compte dans le plan de sondage.

3.1.2. Capacité

Pour les camions, la capacité en termes de poids total autorisé en charge (PTAC) en tonnes a été considérée comme une variable pertinente pour la définition du « futur » plan de sondage. En revanche, la variable PTRAC pour les tracteurs routiers ne l'est pas, car peu discriminante puisque près de 99 % des tracteurs routiers ont une valeur de PTRAC égale à 44. Elle a été remplacée par la puissance fiscale. Le critère de spécialisation éventuelle (transport international, transport de marchandises dangereuses) dont la qualité n'était pas bonne ciblait très peu de véhicules et n'a donc pas été retenu.

3.1.3. Activité de l'entreprise ou activité de l'établissement

La question s'est posée de savoir s'il faut utiliser la variable caractérisant l'activité principale exercée (APE) de l'entreprise à laquelle appartient le véhicule (appelée APEN) ou la variable caractérisant l'activité principale exercée de l'établissement (appelée APET). Outre le fait que les codes APEN sont a priori mieux suivis dans les répertoires SIRENE / SIRUS, le tableau concernant le lien entre APEN et APET montre une bonne « diagonalité » de la matrice croisant les deux critères. De plus, le SIRENE est directement disponible dans le répertoire RSVERO, ce qui facilite la sirénisation au niveau de l'entreprise. Effectuer une sirénisation au niveau de l'établissement nécessiterait de partir des adresses, ce qui risque d'être trop coûteux en termes de reprise manuelle des échos et, d'autant plus que la qualité des données de l'enquête dépendra aussi de cette phase. Pour toutes ces raisons, il est proposé de privilégier le critère APEN et la sirénisation au niveau des entreprises.

In fine, les variables retenues pour définir des « strates » pour un futur plan de sondage sont donc : genre - camion / tracteur -, l'âge du véhicule, PTAC pour les camions, puissance fiscale pour les tracteurs routiers et l'activité de l'entreprise à laquelle appartient le véhicule - résultant de son code APE.

A partir de travaux réalisés sur ces différentes variables, nous avons retenu pour les variables numériques les tranches suivantes pour définir ces « strates » :

- Pour le PTAC, 4 tranches :]3,5-15] ;]15-20] ;]20-30] ; plus de 30 tonnes : en effet, on distingue sur la répartition des camions par PTAC 4 modes, les 4 tranches choisies permettent d'isoler chaque mode dans une seule tranche.
- Pour la puissance fiscale, 2 tranches : « moins de 32 CV » et « supérieure ou égale à 32 CV » : la variable puissance fiscale est très concentrée et peu discriminante pour les tracteurs routiers. La valeur 32 choisie correspond à la médiane dans la plupart des tranches d'âge retenues.
- Pour l'âge du véhicule, 4 tranches : 0 à 5 ans, 6 à 10 ans et 11 à 15 ans, et on ajoute à ces trois tranches la tranche 16-25 ans lié à l'élargissement du champ. Les 3 premières tranches correspondent à celles qui existent dans le plan de sondage actuel ; elles permettraient, dans l'ancien système, de séparer la base de sondage en 4 parties grosso modo de taille égale et ces tranches sont « entières » et parlantes.

Cependant, l'utilisation de la base optimisée conduit à concentrer un peu plus de 62 % du parc roulant dans la première tranche. Les travaux réalisés sur la variable âge pourront donc être poursuivis sur la base optimisée pour éclater la première tranche en plusieurs tranches.

En ce qui concerne l'activité, 9 regroupements de postes ont été effectués : 4941B (Transports routiers de fret de proximité), 4941A (Transports routiers de fret interurbains), 4941C regroupée avec 4942Z (location de camions avec chauffeur, services de déménagement), autres transports (donc APE appartenant aux codes 49 à 53, sauf les précédentes), agriculture (APE 01, 02, 03, 10, 11 et 12), construction (APE 41, 42, 43), commerce (APE 45, 46, 47), APE 7712Z (Location et location-bail de camions) qui a été ajoutée aux catégories utilisées dans la publication TRM afin de l'isoler, et autres (hors 8411Z - administration publique générale qui a été retirée du champ).

En croisant l'ensemble de ces critères, on arrive à 216 « strates », dont 54 concernent les véhicules de 16 à 25 ans.

Une fois les strates définies, on va, du moins dans un premier temps, s'intéresser à la mise au point d'un plan de sondage stratifié, en décidant d'une allocation de questionnaires par strate. A partir de la variance statistique de la variable « tonnes-kilomètres » estimée pour chaque strate à partir du plan de sondage actuel (la formule d'estimation de variance statistique utilisée est précisée en annexe 2), on a réparti un ensemble de 58 000 questionnaires (véhicules) sur les 162 strates (strates ne concernant que les véhicules de moins de 16 ans pour lesquels un calcul est possible à partir des résultats de l'enquête TRM2014) selon une allocation de Neyman utilisant comme variable-objectif les tonnes-kilomètres, en utilisant la base de sondage actuelle (c'est-à-dire non optimisée car ne prenant pas en compte le fichier du contrôle technique).

Le plan de sondage stratifié ainsi défini donne de bons résultats puisque le coefficient de variation de l'estimation d'un certain nombre de variables-cibles (tonnes-kilomètres, kilométrage, tonnes-kilomètres pour compte propre et pour compte d'autrui, tonnes-kilomètres selon la nature de la marchandise) est toujours plus faible que celui obtenu avec le plan de sondage actuel (voir annexe 3).

3.2. Détermination de taux de sondage à partir de l'allocation de Neyman

Pour déterminer de nouvelles allocations « arrondies », on est reparti des valeurs calculées ci-dessus pour les taux de sondage, et on a affecté à chacune des 162 strates un taux de sondage « entier », en termes de pourcentage, en se rapprochant de valeurs « centrales » ; on arrive ainsi à des taux de sondage par strate variant de 2 % à 30 % (valeurs utilisées : 2, 3, 5, 10, 15, 20, 30).

Le tableau 5 donne ces taux de sondage arrondis selon les strates déterminées à partir de travaux menés sur l'ancienne enquête, dont le champ était limité aux véhicules de 15 ans et moins.

Tableau 5 : Taux de sondage en % selon les strates concernant les poids lourds de 15 ans et moins.

Genre	Age(ans)	Capacité	4941B	4941A	4941C 4942Z	Autres transports	Agriculture	Construction	Commerce	7712Z	Autres hors 8411Z
Camions	0-5]3,5-15]	3	3	3	2	2	2	2	2	2
Camions	0-5]15-20]	10	10	10	10	3	2	5	5	10
Camions	0-5]20;30]	10	10	10	10	10	5	5	5	10
Camions	0-5	>30	10	10	10	10	10	5	5	5	10
Camions	6-10]3,5-15]	2	2	2	2	2	2	2	2	2
Camions	6-10]15-20]	5	5	5	5	5	2	2	5	5
Camions	6-10]20;30]	5	5	5	5	5	5	5	5	5

Camions	6-10	>30	5	5	5	5	5	5	5	5	5
Camions	11-15]3,5-15]	2	2	2	2	2	2	2	2	2
Camions	11-15]15-20]	5	5	2	2	5	2	2	2	2
Camions	11-15]20;30]	5	5	5	5	5	5	5	5	5
Camions	11-15	>30	5	5	5	5	5	5	5	5	5
Tracteurs routiers	0-5	< 32	30	30	30	20	20	15	20	20	20
Tracteurs routiers	0-5	>=32	30	30	30	20	20	15	30	30	30
Tracteurs routiers	6-10	< 32	15	15	15	15	15	10	10	10	10
Tracteurs routiers	6-10	>=32	20	20	20	15	15	10	10	10	15
Tracteurs routiers	11-15	< 32	10	10	10	5	5	5	5	10	5
Tracteurs routiers	11-15	>=32	15	10	5	5	5	5	5	15	5

En termes de précision, les résultats (voir annexe 3 - dernière colonne du tableau) de ce nouveau plan de sondage sont à peu près équivalents à ceux du plan fondé sur un Neyman strict, ce qui est logique car les taux retenus sont proches de ceux déterminés par l'allocation de Neyman.

Pour les véhicules de 16 à 25 ans, aucun calcul ne peut être fait car ces véhicules n'entrent pas dans le champ de l'enquête actuelle. Par convention, on prendra la même valeur que celle des véhicules de 11 à 15 ans du même croisement genre / taille, soit les taux de sondage présentés dans le tableau 6.

Tableau 6 : Taux de sondage en % selon les strates concernant les poids lourds de 16 à 25 ans

Genre	Age (ans)	Capacité	4941B	4941A	4941C 4942Z	Autres transports	Agriculture	Construction	Commerce	7712Z	Autres hors 8411Z
Camions	16-25]3,5-15]	2	2	2	2	2	2	2	2	2
Camions	16-25]15-20]	5	5	2	2	5	2	2	2	2
Camions	16-25]20;30]	5	5	5	5	5	5	5	5	5
Camions	16-25	>30	5	5	5	5	5	5	5	5	5
Tracteurs routiers	16-25	< 32	10	10	10	5	5	5	5	10	5
Tracteurs routiers	16-25	>=32	15	10	5	5	5	5	5	15	5

Ce choix correspond à une solution de prudence. En effet, on aurait pu mettre pour toutes les catégories des véhicules de 16 à 25 ans des taux de sondage plus faibles. Cependant, la base optimisée cible a priori des véhicules de plus de 15 ans qui roulent ; de plus, connaissant mal le comportement en termes d'activité de ces véhicules, on ne souhaitait pas leur donner une trop forte pondération par rapport aux autres véhicules de l'échantillon, et ainsi éliminer le risque d'extrapoler trop fortement des comportements atypiques. Ces coefficients seront peut-être à revoir d'ici quelques années, en exploitant les données d'activité de ces véhicules.

3.3. Les inconvénients liés à l'utilisation d'un plan de sondage stratifié simple pour le tirage de l'échantillon

La réalisation d'un plan de sondage stratifié simple pour l'enquête TRM dans laquelle l'unité d'échantillonnage est le véhicule a un inconvénient important : ce plan de sondage ne tient pas compte, lors du tirage, de l'entreprise à laquelle appartient chaque véhicule tiré. Une entreprise pouvant avoir des véhicules classés dans différentes strates et les tirages se faisant de manière indépendante dans chaque strate, aucun « lissage » de la charge statistique n'est possible, si l'on considère le point de vue de l'entreprise.

On aurait pu envisager de limiter le tirage stratifié aux entreprises ayant un seul véhicule (voire peu de véhicules), et d'avoir un tirage plus complexe pour les entreprises multi-véhicules. On a donc étudié la répartition des entreprises selon leur flotte en termes de nombre de véhicules, pour la population concernée par l'ancien plan de sondage, et pour celle prévue pour le futur, donc en incluant les DOM (même si, in fine, il a été décidé qu'on ne réalise pas l'enquête dans ces départements) et les véhicules de plus de 15 ans. Plus de la moitié des véhicules du champ de l'enquête est concentrée dans la sous-population des entreprises d'au moins 40 véhicules (même si plus de 50 % des entreprises concernées par le champ ont un seul véhicule). La proposition de plan de sondage envisagée précédemment, à savoir un plan de sondage stratifié simple pour les entreprises ayant peu de véhicules (un, voire deux ou trois) complété par un plan de sondage plus complexe (par exemple à deux degrés, où le premier degré serait un tirage exhaustif d'entreprises) pour les « grandes », doit donc être abandonnée, vu que les « petites » entreprises, en nombre de véhicules, représentent une part faible de la population étudiée.

Il est donc décidé de ne pas utiliser un plan de sondage stratifié « classique » dont les strates dépendraient des caractéristiques du véhicule, mais plutôt de s'orienter vers un plan de sondage à probabilités inégales utilisant les caractéristiques en question (comme c'est d'ailleurs le cas actuellement).

En revanche, il est possible d'utiliser en amont de la réalisation du sondage à probabilités inégales une « stratification » qui pourrait être réalisée par un tri judicieux à partir de variables du niveau entreprises, ce qui permettrait de classer tous les véhicules d'une même entreprise dans une et une seule strate.

Les probabilités inégales retenues sont celles issues des tableaux 5 et 6. Les taux sont, de manière attendue, plus élevés pour les tracteurs, les véhicules récents et ceux pour lesquels la variable de capacité (PTAC/puissance) a une valeur plus élevée. En appliquant ces taux pour un plan de sondage à probabilités inégales, on considère que ces taux, qui auraient été appliqués dans le cadre d'un sondage stratifié, donnent un ordre de grandeur des probabilités « optimales » à affecter à chaque véhicule qui devraient être proportionnelles à la variable d'intérêt. L'hypothèse sous-jacente est que le « S » calculé par strate pour une allocation de Neyman est proportionnel à la variable d'intérêt (activité).

3.4. Détermination de la taille de l'échantillon

Le règlement européen en vigueur 70/2012³ fixe une contrainte, concernant la précision, qui est que le coefficient de variation de l'estimation de trois variables principales (tonnes transportées, tonnes-kilomètres et total des kilomètres parcourus, que ce soit pour le total ou le total national) doit rester en deçà de 5 %, contrainte largement respectée actuellement.

Lors de la rénovation du processus, l'objectif était d'essayer au maximum de conserver la même précision globale des résultats afin de continuer à diffuser les mêmes séries de données qu'actuellement, largement reprises au niveau national.

³ The requirements are set out in Article 2 of [Commission Regulation \(EC\) 642/2004](#): In general, the standard percentage error (95% confidence) of annual estimates for tonnes transported, tonne-kilometres performed and total kilometres travelled loaded shall not be greater than 5% for total goods road transport and for national goods road transport.

Avec l'optimisation de la base de sondage qui devrait limiter les hors-champ, la taille de l'échantillon retenu est 65 000 véhicules par an, ce qui devrait conduire à un nombre proche de répondants qu'actuellement, soit 58 000 répondants. Le nombre de questionnaires attendu correspond à une charge de traitement compatible avec le nombre de gestionnaires en charge de cette enquête. Les simulations qui ont été menées conduisent à estimer que le coefficient de variation obtenu sur une variable comme le total tonnes-kilomètres serait de l'ordre de 0,7 % à 0,8 %, ce qui reste bien en deçà des normes demandées par Eurostat.

4. Mise en œuvre du plan de sondage à probabilités inégales

La question du lissage de la charge statistique est considérée comme une des priorités de la rénovation du plan de sondage ; il est ainsi nécessaire de choisir une méthode de tirage qui permette de gérer correctement, lors du tirage, l'appartenance des véhicules à une même entreprise (grâce à la variable Siren permettant d'identifier les entreprises).

Par ailleurs, la diffusion standard des résultats de l'enquête distingue en particulier la nature du transport (compte d'autrui/compte propre à partir de la variable activité) et les origines/destinations par région. Il est donc important que l'échantillon conserve une certaine « représentativité » pour ces deux critères. Les deux variables région et activité interviennent dans la méthodologie actuelle dans la phase de tri préalable des véhicules avant de tirer l'échantillon à probabilités inégales par un tirage systématique, ce qui équivaut à une stratification implicite sur ces critères.

Il a donc été décidé de conserver ces variables région et activité comme critères de stratification (ou équivalents). A noter qu'une entreprise se trouve dans une région donnée et dans un regroupement d'APE donné.

4.1. Méthode de tirage

Pour les raisons indiquées précédemment sur le lissage de la charge, le plan de sondage n'est pas un plan de sondage stratifié, mais un plan de sondage à probabilités inégales dont les probabilités de tirage sont proportionnelles à une mesure de taille X_i , où X_i est lié aux taux de sondage obtenus dans le cadre du plan de sondage stratifié avec allocation de Neyman, et avec un tri préalable de la base de sondage. De façon plus précise, le tirage de l'échantillon est réalisé de la manière suivante :

- on affecte à chaque véhicule un coefficient numérique (qui correspond à la probabilité de tirage) variable selon différentes catégories de véhicules, catégories déterminées en croisant le genre (camion / tracteur), l'activité de l'entreprise, l'âge et la capacité (PTAC / puissance) selon les tableaux 5 et 6 ;
- on trie la base de sondage selon les différents regroupements des postes d'activité de l'entreprise retenus, la région (découpage de 2016), le n° Siren, puis le n° de classe du véhicule (les classes étant définies par le croisement genre, âge et capacité, ce qui correspond aux lignes du tableau 5 et 6) ;
- on tire à probabilités inégales selon la méthode du tirage systématique dans le fichier ainsi constitué.

On a donc un tirage à probabilités inégales dans une base triée selon différents critères, sans qu'on ait appliqué une stratification à proprement parler. Cependant, le tri de la base de sondage et le tirage par « sauts » a un impact sur la « représentativité » de l'échantillon : il est ici proposé de trier, au sein d'un regroupement de code APEN puis d'une région donnée, d'abord par Siren, puis ensuite selon le genre, puis par l'âge, et enfin par le critère PTAC/puissance. Ceci se fait au travers de la numérotation des classes définies : ainsi, la classe n°2 est constituée des camions de 0 à 5 ans et de PTAC compris dans la tranche]15;20] ; la catégorie n°5 est constituée des camions de 6 à 10 ans et de PTAC compris dans la tranche]3,5;15]. L'ordre selon lequel on opère pour le tri ayant des conséquences sur la robustesse des résultats détaillés selon les variables qui ont servi de critères de

tri, on peut penser que les résultats seront plus robustes si on s'intéresse à des déclinaisons par secteur ou région que par PTAC/puissance.

La méthode de tirage par sauts donne donc une probabilité de tirage, pour chaque véhicule, proportionnelle à la valeur fournie dans les tableaux 5 et 6, et permet en même temps d'assurer un lissage de la charge par entreprise (les véhicules d'une même entreprise étant rangés les uns derrière les autres dans le fichier, le tirage par sauts assure que l'on ne va pas avoir une concentration de véhicules tirés pour certaines entreprises et aucun véhicule sélectionné pour d'autres ayant un parc de véhicules non négligeable). Des tableaux sont prévus, à l'issue du passage du programme de tirage, pour vérifier que la charge affectant les entreprises ayant un nombre de véhicules important, est bien de l'ordre de celle attendue (sécurisation du processus).

Par ailleurs, il est prévu **un renouvellement annuel par moitié de l'échantillon** ; afin de séparer la base de sondage en deux parties équivalents, il a été décidé d'utiliser la parité du n° Siren : ceci classe donc tous les véhicules d'une même entreprise dans la même catégorie (Siren pair / Siren impair), et rend la gestion de la problématique de la charge plus facile qu'avec l'ancienne méthode : celle-ci utilisait la parité du n° de série du véhicule, ce qui faisait que deux véhicules d'une même entreprise pouvaient être classés dans les deux « parités », et que la gestion de la charge statistique au sein d'une même entreprise s'avérait ensuite très difficile. Enfin, il faut noter qu'aucune coordination négative de l'échantillon renouvelé avec celui tiré deux ans avant n'est opérée.

4.2. Une meilleure prise en compte de la dimension entreprise

Le nouveau plan de sondage proposé prend donc mieux en compte la dimension entreprise au sens où :

- le tirage effectif des véhicules est un tirage systématique réalisé sur une base triée par Siren (les véhicules d'une même entreprise sont rangés les uns derrière les autres et le tirage systématique assure que l'on n'a pas une concentration du tirage sur certaines entreprises).
- Pour le renouvellement annuel par moitié, la base de sondage est répartie en deux parties en utilisant la parité du n° Siren .

Les différentes simulations réalisées montrent qu'on arrive ainsi à un lissage de la charge des entreprises, avec en conséquence une augmentation du nombre d'entreprises concernées par l'enquête (de l'ordre de 20 %). La part des entreprises concernées par une seule enquête par an évoluerait de 56,7 % à 66,9 %, celle concernée par plus de 10 véhicules par an passerait de 7,3 % à 4 % (voir tableaux 7 et 8).

Tableau 7 : Répartition des entreprises de l'enquête « TRM » actuelle selon leur parc (en colonne) et le nombre de véhicules enquêtés en 2016 (en ligne)

Parc/ Véhicules enquêtés ⁽²⁾	Non identifié ⁽¹⁾	1	2	3	4	5	6 à 10	11 à 50	51 à 100	101 et plus	Total	En %
1	1 296	2 126	1 455	1 062	781	557	1 571	1 380	69	18	10 315	56,7
2	157	167	204	189	205	192	669	828	62	21	2 694	14,8
3	44	42	59	72	55	64	297	535	64	27	1 259	6,9
4	18	26	21	32	30	37	175	405	64	27	835	4,6
5	10	10	11	13	25	18	103	281	57	18	546	3,0
6 à 10	21	19	20	20	22	28	156	672	157	105	1 220	6,7
11 à 50	16	5	12	9	7	10	50	554	270	247	1 180	6,5

51 à 100	1	0	0	0	0	0	1	8	15	72	97	0,5
101 et plus	0	0	0	0	0	0	0	2	5	40	47	0,3
Total	1 563	2 395	1 782	1 397	1 125	906	3 022	4 665	763	575	18 193	100,0

(1) La colonne « non identifié » représente le nombre d'entreprises dont le Siren n'a pas été retrouvé (une partie des 1 296) ou est erroné.

(2) Dans le tableau 7, la répartition des entreprises de l'enquête « TRM » actuelle selon leur parc et le nombre de véhicules enquêtés a été réalisée après la réaffectation des véhicules loués à l'entreprise locataire (lorsque son « identité » via le loueur a été communiquée). Ceci explique que l'on puisse avoir un nombre de véhicules enquêtés supérieur au parc (les véhicules dont l'entreprise est propriétaire).

Tableau 8 : Simulation de la répartition des entreprises sélectionnées par le plan de sondage futur selon leur parc (en colonne) et le nombre de véhicules enquêtés (en ligne)

Parc/ Véhicules enquêtés	1	2	3	4	5	6 à 10	11 à 50	51 à 100	101 et plus	Total	En %
1	5 154	3 067	2 055	1 406	1 017	2 057	765	2	0	15 523	66,9
2	0	86	241	358	366	1 138	721	9	0	2 919	12,6
3	0	0	6	15	55	641	624	15	0	1 356	5,8
4	0	0	0	1	1	226	498	19	1	746	3,2
5	0	0	0	0	0	38	453	21	6	518	2,2
6 à 10	0	0	0	0	0	0	1 057	124	17	1 198	5,2
11 à 50	0	0	0	0	0	0	344	328	178	850	3,7
51 à 100	0	0	0	0	0	0	0	0	54	54	0,2
101 et plus	0	0	0	0	0	0	0	0	26	26	0,1
Total	5 154	3 153	2 302	1 780	1 439	4 100	4 462	518	282	23 190	100,0

4.3. Répartition de l'échantillon sur l'année

L'échantillon ainsi obtenu doit être ensuite réparti sur l'ensemble des semaines de l'année. Mais du fait que la base de sondage évolue puisque de nouveaux véhicules sont immatriculés tous les jours ouvrés, le dispositif de tirage est plus complexe qu'un tirage d'échantillon classique.

4.3.1. Répartition de l'échantillon entre les trimestres

La base de sondage évoluant au cours d'année, il est alors nécessaire, pour être représentatif du parc, de tirer des compléments d'échantillon trimestriels en cours d'année, et de les répartir sur les trimestres restants ; les tailles d'échantillons ne peuvent donc pas être les mêmes au fur et à mesure que l'on avance dans l'année. De plus, une moitié d'échantillon est reconduite chaque année. Les compléments trimestriels d'échantillons doivent donc être sélectionnés dans les deux parties de la base de sondage (les Siren pairs et les Siren impairs). Autrement dit, les différents compléments d'échantillons concernent également la partie conservée. A noter que fin novembre d'une année, on sélectionnera l'échantillon principal de la partie renouvelée dans la base de sondage actualisée à cette date ainsi qu'un échantillon complémentaire pour actualiser l'échantillon de la partie conservée dans la base de sondage des nouveaux véhicules avec un « Siren » de la partie conservée des mois de septembre, octobre et novembre.

Concernant le calibrage des échantillons complémentaires, des calculs menés sur les bases actuelles montrent qu'on a environ 2 % de véhicules nouveaux par trimestre (rapport entre le nombre de nouveaux véhicules d'un trimestre donné divisé par le nombre total de véhicules), mais que ceux-ci pèsent 3 % en termes de probabilités de tirage car ils sont récents, par définition (le 3 % est obtenu comme le rapport entre la somme des probabilités des nouveaux véhicules et la somme des probabilités de l'ensemble des véhicules). Cette valeur a servi à déterminer les tailles d'échantillons complémentaires indiquées ci-dessous.

Afin de synthétiser l'ensemble du processus, le tableau 9 précise la manière dont l'échantillon est tiré chaque trimestre et réparti sur les trimestres restants, en utilisant une grandeur N qui est un paramètre servant de référence, et pas la taille globale de l'échantillon (qui est plutôt de l'ordre d'un peu plus de 2N) ; on se place dans le cas où la partie conservée concerne les Siren pairs.

Tableau 9 : Schéma de tirage utilisé

Date de tirage	Taille échantillon	1 ^{er} trimestre	2 ^{ème} trimestre	3 ^{ème} trimestre	4 ^{ème} trimestre
Échantillon principal (tiré fin novembre)	<p><u>Siren pairs :</u></p> <ul style="list-style-type: none"> - N véhicules conservés (tirages année précédente) x - 0,03N pour les nouveaux véhicules arrivés en fin d'année dans la base de sondage (depuis le tirage de l'échantillon complémentaire 3) x <p><u>Siren impairs :</u> 0,955N véhicules tirés</p>	x	x	x	x
Échantillon complémentaire 1 tiré fin février (dans le fichier des véhicules immatriculés depuis le tirage de l'échantillon principal)	<p><u>Siren pairs :</u> 0,0225N</p> <p><u>Siren impairs :</u> 0,0225N</p>		x	x	x
Échantillon complémentaire 2 tiré fin mai (dans le fichier des véhicules immatriculés depuis le tirage de l'échantillon complémentaire 1)	<p><u>Siren pairs :</u> 0,015N</p> <p><u>Siren impairs :</u> 0,015N</p>			x	x
Échantillon complémentaire 3 tiré fin août (dans le fichier des véhicules immatriculés depuis le tirage de l'échantillon complémentaire 2)	<p><u>Siren pairs :</u> 0,0075N</p> <p><u>Siren impairs :</u> 0,0075N</p>				x
Taille totale de l'échantillon	<p><u>Siren pairs :</u></p> <p>$N+0,03N+0,0225N+0,0015N+0,0075N = 1,075N$</p> <p><u>Siren impairs :</u></p> <p>$0,955N+0,0225N+0,015N+0,0075N = N$</p>				

Par exemple, pour l'échantillon complémentaire 2, on affectera donc 0,015N véhicules enquêtés aux troisième et quatrième trimestres pour les Siren impairs (soit 0,0075N pour le troisième trimestre et 0,0075N pour le dernier) et également 0,015N

aux troisième et quatrième trimestres pour les Siren pairs. La décroissance de la taille des échantillons complémentaires au cours du temps est due au fait que ceux-ci doivent être répartis sur le nombre de trimestres restants.

In fine, on aura à ce moment, pour la partie d'échantillon relative aux Siren impairs un nombre de véhicules tirés égal à $0,955N+0,0225N+0,015N+0,0075N= N$. La partie d'échantillon relative aux Siren pairs sera, elle, de 1,03 au départ puis évoluera à la hausse avec les compléments tirés chaque trimestre pour atteindre $(1,03+0,045)N$, soit $1,075N$. Au total, la taille de l'échantillon est donc de $2,075N$. On peut alors déterminer le N qui permet de s'ajuster à un volume global d'échantillon, par exemple 65 000 véhicules.

L'échantillon principal est réparti de façon homogène entre les différentes semaines. Cependant, avec ce principe d'ajout au fur et à mesure d'échantillons complémentaires, la taille de l'échantillon à traiter par les gestionnaires augmente au fil des trimestres (pour être maximale au 4ème trimestre qui intègre les nouveaux véhicules des 9 derniers mois échantillonnés).

Cependant, on peut penser que le phénomène d'attrition de l'échantillon qui joue en sens inverse, même si on ne dispose pas ici de données quantifiées sur le sujet, sera plus fort sur les derniers trimestres : un certain nombre de véhicules tirés initialement ne sont plus en activité et également les sorties de champ liées à des sorties de territoire – ventes à l'étranger – s'accroissent selon le trimestre d'enquête : ainsi, au T4, un véhicule a une plus forte probabilité d'être sorti du champ durant l'année qu'au T1.

On aurait pu utiliser un plan de sondage utilisant des probabilités fixées a priori, ce qui aurait conduit à avoir des tailles des échantillons complémentaires aléatoires (en fonction du volume des fichiers de « nouveaux véhicules » arrivés dans la base de sondage). On a préféré l'option présentée, qui permet de contrôler la taille de l'échantillon.

4.3.2. Répartition de l'échantillon selon les semaines d'enquête

4.3.2.1. Affectation des semaines à un trimestre (et à une année)

Avant de répartir l'échantillon tiré sur les semaines, il est d'abord nécessaire de rattacher toutes les semaines à un et un seul trimestre : il est proposé d'affecter chaque semaine au trimestre où elle a le plus de jours, pour les journées réparties entre le lundi et le vendredi et non fériés. En effet, la répartition de l'activité, en termes de tonnes-kilomètres, montre que le poids du samedi et surtout du dimanche est très faible (tableau 10).

Tableau 10 : Répartition des tonnes kilomètres selon le jour de la semaine (source : enquête TRM2015)

Jour de la semaine	% du total tonnes-kilomètres
Lundi	21,4
Mardi	20,5
Mercredi	20,2
Jeudi	19,5
Vendredi	15,7
Samedi	2
Dimanche	0,6
Total	100

Ainsi, on évite le problème des semaines « à cheval » sur deux trimestres qui posait des difficultés dans l'ancienne méthode. Cependant, on n'est pas assuré d'avoir exactement le même nombre de semaines par trimestre chaque année.

4.3.2.2. Répartition de l'échantillon (pour la partie renouvelée) entre les semaines et les trimestres

Pour un tirage à répartir sur plusieurs trimestres (on prend ici le cas du tirage principal réalisé fin novembre, mais le même principe s'applique aux autres tirages), il est proposé de répartir l'échantillon selon les trimestres et semaines de la façon suivante : selon l'ordre du tirage « par sauts », on affecte d'abord le véhicule tiré à T1-1 (première semaine du premier trimestre), puis le suivant à T3-1, puis T2-1, puis T4-1, puis T1-2, T3-2, T2-2, T4-2, etc. Dans le cas où un trimestre a une semaine de plus que les autres, on peut donc avoir à un moment un décalage (par exemple : T1-1, T3-14, ...). Cette méthode permet de mieux répartir dans l'année les questionnaires des entreprises ayant un nombre de véhicules enquêtés important.

Elle répartit de façon approximativement égale l'échantillon entre les trimestres et semaines, mais on devra tenir compte, au moment des calculs des pondérations, du nombre exact de jours que la partie d'échantillon relative à une période donnée (trimestre par exemple) est censée représenter.

Pour la partie de l'échantillon qui est reconduite, on utilise les numéros de semaine qui ont été affectés au véhicule l'année de tirage. Cependant, deux configurations « exceptionnelles » peuvent se produire :

- un trimestre a une semaine de plus l'année suivante : on réoriente une partie des véhicules tirés la première année sur la « nouvelle » semaine ; ainsi, si on passe d'un trimestre de 13 semaines la première année à 14 semaines l'année suivante, on doit avoir un nombre de véhicules moyen par semaine n_2 la seconde année, tel que $14 \cdot n_2 = 13 \cdot n_1$ où n_1 est le nombre de véhicules moyen par semaine la première année. On doit donc enlever $n_1 - n_2$ soit $1/14^{\text{ième}}$ de n_1 .

- un trimestre a une semaine de moins que l'année précédente : on réoriente les véhicules de la semaine « disparue » sur les autres semaines, selon un raisonnement identique à celui détaillé ci-dessus.

4.4. Calcul de la pondération affectée à un questionnaire donné

A l'issue du tirage de l'échantillon annuel, chaque véhicule dispose d'un coefficient permettant de représenter la population totale des véhicules. Plus précisément, en approximant le plan de sondage par un sondage à probabilités inégales avec remise pour lequel, à chaque tirage, une unité i est tirée avec une probabilité A_i , le total $T(Y)$ de la variable Y est estimé à partir d'un échantillon de taille n par

$$T^Y = \left(\frac{1}{n}\right) \sum_{i=1}^n \frac{y_i}{A_i} \quad \text{où les probabilités } A_i \text{ résultent des valeurs } V_i \text{ fournies dans les tableaux 5 et 6,}$$

soit :
$$A_i = \left(\frac{V_i}{\sum_{\text{base de sondage}} V_i} \right)$$

Le coefficient affecté à un véhicule tiré pour représenter l'ensemble des véhicules vaut donc : $\left(\frac{1}{nA_i}\right)$

Cependant, la production des résultats de l'enquête TRM (par exemple, nombre total de kilomètres parcourus pendant un trimestre ou l'année) nécessite de se placer au niveau des « véhicules-semaines ».

Le coefficient précédent $1/nA_i$ doit tenir compte de la taille d'échantillon « n » affectée au trimestre en question (chaque partie d'échantillon doit être considérée séparément - par exemple celle tirée lors du tirage de fin novembre, ou celle tirée pour un échantillon complémentaire - et conduit donc à un calcul utilisant un « n » différent) ; de même les A_i sont calculés en sommant le dénominateur sur la base de sondage dans laquelle est tiré le véhicule. Ce premier coefficient obtenu permet donc de

représenter les véhicules de la partie de la base de sondage dont est issu le véhicule. Pour chacun des échantillons tirés, on a 7 pondérations possibles selon les caractéristiques du véhicule sélectionné (associées aux 7 valeurs de V_i différentes des tableaux 5 et 6), et prenant en compte la taille de l'échantillon et le dénominateur des A_i (variable selon l'échantillon tiré).

Exemple : on veut calculer le coefficient d'un véhicule tiré dans l'échantillon complémentaire 1 (tiré fin février), de Siren pair et enquêté au 3ème trimestre. Le coefficient d'extrapolation va être égal à $1/nA_i$, où n est la taille de l'échantillon complémentaire de Siren pairs, tiré fin février et enquêté au 3ème trimestre : approximativement, n vaudra $0,0225N/3$ (mais il est préférable de le calculer exactement avec le nombre de véhicules réellement affectés à ce 3ème trimestre). Quant au A_i , il est calculé à partir de la formule de la page précédente, où, au dénominateur, la somme sur la base de sondage est réalisée sur l'ensemble des véhicules pris en compte dans la base complémentaire de février pour les Siren pairs.

On doit ensuite obtenir un coefficient d'extrapolation permettant de publier des résultats globaux, alors que les données collectées sur les questionnaires relativement aux trajets des véhicules enquêtés ne concernent qu'une semaine donnée. On multiplie donc le coefficient obtenu précédemment par un coefficient égal au nombre de jours du trimestre divisé par 7 (par exemple, $91/7$ pour le deuxième trimestre), selon le trimestre pendant lequel le véhicule a été enquêté, pour avoir le coefficient « final ».

Concernant les estimations annuelles, elles sont obtenues par somme des quatre estimations trimestrielles.

4.5. Simulation de tirages d'échantillon

Plusieurs simulations de tirages d'échantillon ont été réalisées, celles-ci comprennent l'échantillon principal initial de l'année de basculement ainsi que les trois échantillons complémentaires qui « complètent » cet échantillon au cours de l'année de réalisation de l'enquête. La répartition de l'échantillon global des 65 000 véhicules par semaine de collecte indique que la charge d'enquête est plus importante en fin d'année, en passant de 1 175 questionnaires envoyés dans la première semaine de janvier à 1 325 questionnaires envoyés la dernière semaine de décembre.

Par ailleurs, lors de la mise en place de la nouvelle méthodologie, il a été décidé de basculer en une seule fois, autrement dit le premier échantillon avec la nouvelle méthodologie couvrira à la fois les Siren pairs et les Siren impairs. Par conséquent, on ne tiendra pas compte de la partie renouvelée de l'échantillon de l'année précédente.

5. Éléments sur la correction de la non-réponse totale

Au moment de la sortie des résultats de l'enquête (résultats trimestriels ou annuels), deux types d'éléments doivent être pris en compte :

- non-réponses « constatées » dans l'échantillon ;
- informations disponibles dans certaines sources externes à l'enquête : sur la « sortie du champ de l'enquête » de certains véhicules de l'échantillon (vente de véhicules à l'étranger, « mise à la retraite » de certains véhicules), ou sur la présomption d'activité d'autres véhicules, au travers de l'existence d'un contrôle technique dans l'année.

Concernant le premier point (non-réponse), on se trouve, comme pour la plupart des enquêtes auprès des entreprises réalisées par courrier ou par internet, confronté à des « non-retours » (plutôt que des non-réponses) qui mélangent deux causes différentes :

- certaines entreprises refusent de répondre à l'enquête alors que leur véhicule a roulé pendant la période concernée par l'enquête (cas « classique » de non-réponse) ;

- d'autres ne le font pas parce que le véhicule est « hors du champ » (il ne roule plus ou il a été vendu à l'étranger).

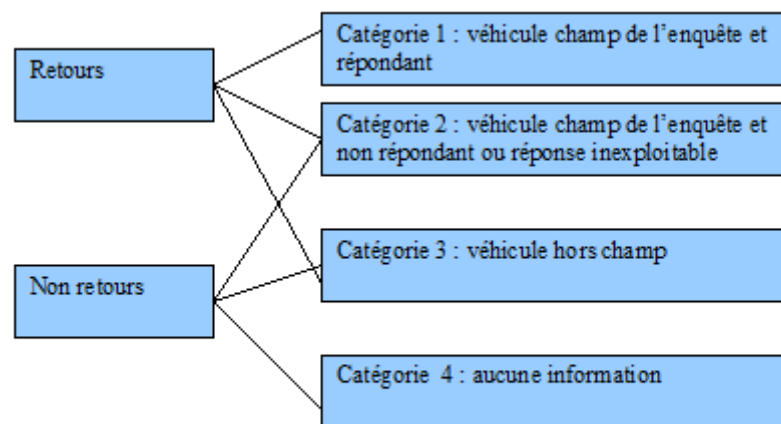
La disponibilité d'informations externes va permettre de rendre les estimations produites plus efficaces.

5.1. Méthode proposée

La méthodologie proposée consiste à s'appuyer sur un traitement des non-réponses approprié au contexte des statistiques d'entreprises (Brion, Ph., Caron, N., Pietri-Bessy, P., 2005). Plus précisément, elle définit quatre catégories d'unités de l'échantillon (qui sont ici des véhicules, et non des entreprises stricto sensu) destinées à affiner le traitement des non-retours :

- catégorie 1 : les retours (donc les questionnaires jugés utilisables pour être pris en compte dans la sortie des résultats) ;
- catégorie 2 : les véhicules dont on est certain qu'ils sont dans le champ, mais pour lesquels on n'a pas de retour, ou pour lesquels la réponse est jugée inexploitable ;
- catégorie 3 : les véhicules qui ne sont plus dans le champ de l'enquête (détruits, vendus à l'étranger) ;
- catégorie 4 : les véhicules classés en « non-retour » pour lesquels on ne dispose d'aucune information supplémentaire, et qui constituent donc le problème majeur concernant la production d'estimations, puisque mélangeant différentes raisons de non-retour.

Graphique 11 : répartition des véhicules de l'échantillon selon différentes catégories



Premièrement, les informations obtenues sur les questionnaires retournés (au travers d'un code affecté par les gestionnaires de l'enquête, indiquant si le véhicule a roulé pendant la semaine d'enquête, s'il est détruit, vendu à l'étranger, ou si le questionnaire est jugé suffisamment bon ou pas pour être utilisé) permettent de classer un certain nombre de questionnaires dans les quatre catégories précédentes. Ensuite, les informations disponibles dans les fichiers externes viennent en complément pour classer un certain nombre d'autres véhicules de l'échantillon dans ces quatre catégories. L'idée est de réduire le plus possible la taille de la catégorie 4.

Plus précisément, pour ce qui concerne les fichiers externes :

- le fichier du contrôle technique indique que certains véhicules de cette catégorie ont roulé pendant l'année d'enquête ; ils sont placés dans la catégorie 2 ;

- le deuxième fichier utilisé donne des informations sur des véhicules « hors champ » (détruits ou vendus à l'étranger) : ceux-ci sont placés en catégorie 3.

Au sein de l'échantillon, ce sont les véhicules des catégories 1 (véhicules dans le champ et répondants) et 3 (véhicules hors champ) qui sont utilisés pour produire les résultats statistiques : de manière évidente, ce sont ceux de la catégorie 1 qui servent pour les résultats à proprement parler, mais ceux de la catégorie 3 (comprenant les véhicules hors champ) doivent également être pris en compte pour les calculs de précision (puisque l'on a dans cette catégorie des unités sélectionnées dans l'échantillon, dont on sait qu'elles comptent pour des valeurs « zéro » dans les résultats, mais qui génèrent de la variance au niveau des estimateurs).

Selon la méthode préconisée, on procède en deux étapes pour calculer les pondérations affectées aux unités de l'échantillon (en partant, au départ, des poids issus du tirage) :

- l'étape 1 permet aux unités des catégories 1 à 3 de « représenter » celles de la catégorie 4 (catégorie sur laquelle on ne sait rien) au travers d'une modification de leurs poids ;
- l'étape 2 permet aux unités de la catégorie 1 de représenter celles de la catégorie 2 (qui sont les « vraies » non-réponses, ou les questionnaires reçus mais d'une qualité telle qu'ils ne sont pas exploitables), également au travers d'une modification de leurs poids.

In fine, on dispose donc de poids pour les catégories 1 et 3, à utiliser pour la production d'estimations statistiques ainsi que de calculs de précisions.

5.2. Précisions sur les deux étapes du calage

La première étape (modification des poids des catégories 1, 2 et 3 pour représenter la catégorie 4) va être réalisée par une procédure de calage sur marge, en prenant en compte les « marges » donnant la distribution des véhicules connue par le fichier dit du parc roulant (qui intègre, chaque trimestre, les nouveaux véhicules) selon les marges suivantes définies lors du tirage de l'échantillon :

- croisement genre (camion/tracteur) avec tranche d'âge en 5 modalités (0 à 2 ans ; 3 à 5 ; 6 à 10 ; 11 à 15 ; 16 à 25) ;
- activité de l'entreprise (variable APEN) selon la nomenclature utilisée pour le tirage (9 modalités) ;
- région en 13 modalités (selon la nouvelle nomenclature).

Cette procédure, appliquée chaque trimestre, va nous assurer que l'échantillon « retrouve » bien les distributions connues selon ces critères, ce qui n'était assuré que de manière approchée en fonction du mode de tirage d'une part (à probabilités inégales par « sauts » dans le fichier), et en raison de la procédure de répartition entre les différents trimestres d'autre part.

La deuxième étape de calage (modification des poids de la catégorie 1 pour représenter la catégorie 2) est réalisée sur les mêmes marges (au sens : utilisant les mêmes critères), mais cette fois ces marges sont produites à partir de l'échantillon extrapolé avec les poids issus de la première étape (pour sa partie limitée aux catégories 1 et 2).

- 5.2.1. Le travail de calage doit être réalisé chaque trimestre, en fonction des éléments les plus récents connus (pour les fichiers externes, ou prenant en compte des véhicules récents pour les marges) ; de plus, l'échantillon à caler intègre l'échantillon complémentaire tiré pour chacun des trimestres (de 2 à 4).

L'ensemble du processus a donné lieu à la mise au point d'un ensemble de programmes qui sont accompagnés de sorties destinées à s'assurer de son bon fonctionnement, à l'issue de chacune des deux étapes, sur la distribution des poids ainsi que sur les nombres de véhicules extrapolés selon le genre et l'âge.

5.3. Exploitation trimestrielle de l'enquête et correction des variations saisonnières

Il est important de préciser que des traitements statistiques supplémentaires sont réalisés pour l'exploitation infra-annuelle de l'enquête permanente sur l'utilisation des véhicules routiers de marchandises. En effet, les séries concernant l'activité du transport routier de marchandises sont sensibles aux saisons : le chiffre du TRM national au deuxième trimestre est par exemple en moyenne sur la période 2001-2016 de l'ordre de 3 % supérieur à la moyenne annuelle (et même 6 % pour l'année 2008). Les résultats infra-annuels de l'enquête TRM sont par conséquent présentés en CVS-CJO (corrigés des variations saisonnières et des jours ouvrables). La série corrigée des variations saisonnières (CVS), construite à partir de la série initiale dite « série brute », permet de neutraliser les évolutions habituelles au cours de l'année pour mieux faire ressortir à la fois les tendances de fond et les évolutions exceptionnelles. Celle corrigée des jours ouvrables (CJO) permet de neutraliser l'impact des nombres inégaux de jours ouvrables d'un trimestre à l'autre, de la même façon que la série CVS neutralise l'impact des différentes saisons. La combinaison des CVS et des CJO fournit une information sur l'évolution instantanée des phénomènes économiques, abstraction faite des phénomènes calendaires explicables habituels. Les modèles de désaisonnalisation sont revus chaque année lorsque les résultats de l'année précédente deviennent définitifs, peu avant la publication des données concernant le premier trimestre. Chaque trimestre, les résultats prennent en compte les nouveaux retours de questionnaires sur le dernier trimestre mais aussi ceux sur le trimestre précédent. Ces nouveaux questionnaires sont donc intégrés dans le calcul des coefficients saisonniers, ceux-ci sont donc réestimés chaque trimestre, ce qui peut faire réviser la série.

6. Conclusion et perspectives

Les premières études méthodologiques sur le sujet réalisées dans le cadre de ce travail l'ont été avec la base de sondage concernant la totalité du parc qui n'est pas la base optimisée (avec l'apport des contrôles techniques) finalement retenue. En effet, la constitution de cette base optimisée n'était pas disponible lors du démarrage de ces travaux. Ce changement d'optique a un impact important sur la base de sondage, avec une diminution du nombre de poids lourds anciens, ce qui nécessite de s'assurer que certaines des options choisies sont encore valables. En particulier, il est nécessaire de :

- réexaminer le découpage en tranches de deux variables servant à définir les probabilités inégales : puissance fiscale pour les tracteurs routiers et âge ;
- recalculer les allocations de Neyman ;
- il peut également être intéressant de vérifier si les valeurs utilisées pour les échantillons trimestriels complémentaires (poids de l'ordre de 3 % en matière de trafic réalisé) sont toujours valables ou doivent être ajustées ;
- examiner les réponses des entreprises de la division 45 de la NAF (commerce automobile) à l'enquête TRM. 14 % des véhicules d'entreprises de cette division ont roulé selon l'enquête, ce qui a conduit dans un premier temps à ne pas préconiser d'exclure les entreprises de ce code NAF du champ de l'enquête. Or, de nombreuses entreprises de ce secteur se plaignent de ne pas être concernées par l'enquête. Ce sujet devra donc être instruit de manière plus fine.

Plusieurs simulations de tirages d'échantillon ont pu être réalisées, ce qui a permis d'améliorer les programmes de tirage. Il est important de poursuivre avec des simulations sur la partie des traitements post-collecte (partie calage) avec les programmes qui ont été développés. L'idée serait d'utiliser la même méthode que pour estimer la précision issue du nouveau plan de sondage, à savoir de considérer le fichier des résultats de TRM en affectant à chaque véhicule une pondération comme s'il avait été sélectionné avec le nouveau plan de sondage.

Étant donné les changements prévus avec en particulier l'utilisation d'une base de sondage optimisée et la prise en compte de la dimension entreprise, ce qui constitue les deux grandes avancées dans la rénovation de la méthodologie de cette enquête, il est préconisé de mettre en œuvre la nouvelle méthodologie en une seule fois à l'occasion d'un nouveau millésime d'enquête.

Cependant, il faut s'attendre à des ruptures sur les niveaux pour certaines séries, notamment du fait que des tables de référence anciennes ne seront plus utilisées.

Le travail de raccord des niveaux des différentes séries, avant et après le changement de méthodologie, est à réaliser. Les estimations de la marche d'escalier sont délicates à réaliser et la difficulté réside dans le fait que la vraie valeur ne sera connue que lors de la production des premiers résultats avec le dispositif rénové. Les évolutions devraient être moins touchées. En effet, on peut penser que, s'il existe un biais, celui-ci est constant et ne joue pas sur les évolutions, le système précédent ayant fonctionné à méthodologie constante. A noter que le changement de méthodologie nécessitera de revoir et d'actualiser tous les modèles de désaisonnalisation pour les utilisations infra-annuelles de l'enquête, ce qui aura très probablement des conséquences en révisant les séries plus fortement que les autres années.

Bibliographie

[1] Brion, Ph., « Esane, le dispositif rénové de production des statistiques structurelles d'entreprises », *Courrier des Statistiques*, n°130, Insee, mai 2011.

[2] Brion, Ph., Caron, N., Pietri-Bessy P., « Redresser la non-réponse totale dans les enquêtes entreprises : les pièges à éviter. Illustration avec l'enquête Innovation », *JMS 2005*, Insee.

[3] Gallais, A. « Tirage à probabilités inégales et calage : les nouveaux principes de l'enquête transports routiers de marchandises », *lettre du SSE*, n°49, 2^{ème} trimestre 2002, disponible à l'adresse suivante :

http://www.statistiques.developpement-durable.gouv.fr/fileadmin/documents/Sources_et_Methodes/Operations_statistiques_et_production_d_indices/Transports/Fiches_qualite/Bilan_qualite_TRM%202011%20C.pdf

[4] Lutinier, B, « Le transport routier de marchandises français entre 1996 et 2016 : impacts de la crise économique et de la concurrence », *Datalab essentiel*, n° 128, décembre 2017.

Annexe 1 : Extension ou non du champ de l'enquête aux DOM et aux véhicules de 16 à 25 ans : calcul des erreurs quadratiques moyennes

Pour juger de la pertinence d'une extension de champ de l'enquête dans le cadre de la production de résultats agrégés, on compare l'efficacité de l'enquête limitée à l'ancien champ et celle de la même enquête réalisée sur un champ étendu à taille d'échantillon constante.

Pour cela, il est nécessaire d'observer comment se comportent, dans les deux cas (champ limité ou étendu), les deux composantes de l'erreur quadratique moyenne⁴ qui se définit par $EQM = (\text{Biais})^2 + \text{Variance}$:

- carré du biais, qui dans le cas « champ étendu » sera nul (on suppose ici que la population des véhicules de plus de 25 ans a un poids infime dans des agrégats comme le nombre de tonnes-kilomètres), alors que le biais peut être non négligeable dans le cas où on travaille sur un champ « limité » ;

- variance de l'estimation produite, qui dans le cas où on tire l'échantillon sur une population plus large devrait être plus grande que celle résultant de l'échantillon tiré sur le champ « limité ».

Ainsi, pour les deux cas (inclusion ou non des DOM, et extension aux véhicules de 16 à 25 ans), l'apport des extensions a été apprécié en comparant l'erreur quadratique moyenne obtenue pour l'estimation du nombre de tonnes-kilomètres en sélectionnant dans les deux bases de sondage (actuelle et étendue) un échantillon de même taille. Concernant la variance, elle a été estimée par la formule donnant la variance d'un plan de sondage à probabilités inégales avec remise dans une population de taille N (on fait donc « comme si » le tirage était réalisé à probabilités inégales, sans prendre en compte l'apport potentiel du tirage systématique⁵ :

$$V(T^Y) = \left(\frac{1}{n}\right) \sum_{i=1}^N A_i \left(\frac{Y_i}{A_i} - T(Y)\right)^2$$

où n est la taille de l'échantillon, T^Y l'estimation du total de la variable Y, A_i la probabilité pour le véhicule i d'être sélectionné à chaque tirage et $T(Y)$ le total de la variable Y dans la base de sondage.

A_i est calculée à partir des valeurs V_i fournies dans le tableau 5, soit

$$A_i = \left(\frac{V_i}{\sum_{\text{base de sondage}} V_i} \right) .$$

Pour estimer le total d'une variable Y à partir d'un échantillon de taille n, on utilise l'estimateur :

$$T^Y = \left(\frac{1}{n}\right) \sum_{i=1}^n \frac{Y_i}{A_i}$$

Pour le calcul avec la base de sondage étendue, on a estimé que le passage d'une base de sondage de taille N_1 à une autre de taille N_2 multiplie la variance par un facteur compris entre N_2/N_1 et $(N_2/N_1)^2$: ce calcul est approximatif, mais permet d'avoir des ordres de grandeur ; il repose sur le fait qu'on passe d'une sommation sur N_1 termes à une sommation sur N_2 termes, et que les A_i ainsi que $T(Y)$ sont revus, les premiers à la baisse, le second à la hausse.

⁴ Ici, on propose une comparaison des deux valeurs de l'EQM obtenues sans qu'on procède à une modélisation destinée à produire une estimation de la partie manquante, dans le cas où on travaille sur un champ limité (par exemple aux moins de 16 ans). Une telle modélisation permettrait sans doute de réduire l'EQM de l'estimateur utilisant le champ étendu, via une réduction notable du biais.

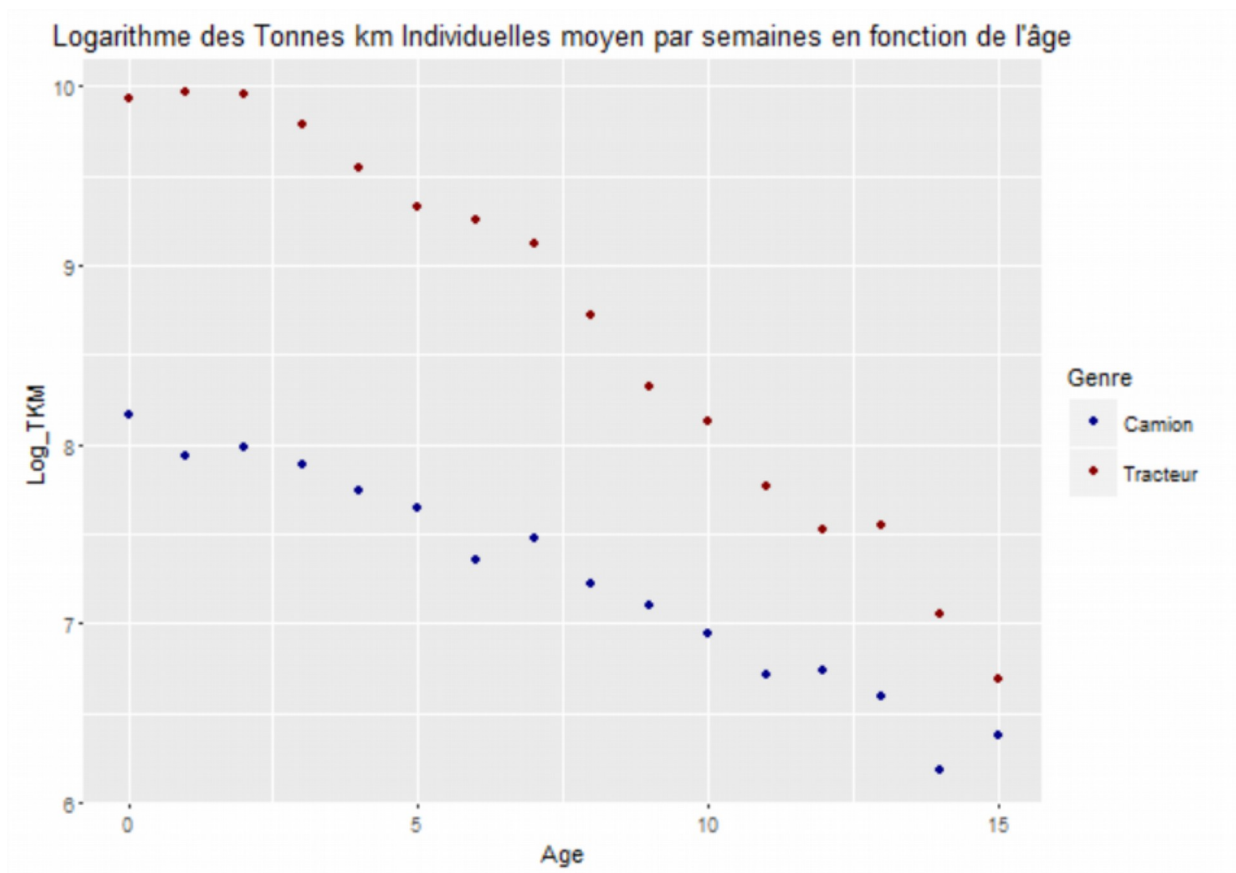
⁵ la phase de calage n'est pas non plus prise en considération dans ces calculs

Les calculs réalisés ont été menés sans tenir compte des contrôles techniques, c'est-à-dire avec une base de sondage non optimisée.

Calcul de l'EQM pour l'inclusion des poids lourds de 16 à 25 ans

Le biais dû aux véhicules de 16 à 25 ans sur la métropole a été estimé à partir de travaux de modélisation de la variable « tonnes-kilomètres » en fonction de l'âge des véhicules. Ainsi, à partir des données de l'enquête TRM, on voit que le logarithme du nombre de tonnes-kilomètres effectués par an et par véhicule diminue à peu près de façon linéaire (en enlevant les deux premiers points) en fonction de l'âge et dépend du genre du véhicule.

Graphique 12 :



La modélisation par régression linéaire conduit à des tonnes-kilomètres moyennes estimées pour les poids lourds de 16 à 25 ans par genre. Ensuite, en multipliant ce nombre moyen de tonnes-kilomètres par le nombre de véhicules de chaque âge dans la base de sondage de 2016, on obtient une estimation du total de tonnes-kilomètres parcourues, ce qui in fine nous permet d'estimer le biais à 2,7 %. Ce biais est moindre si on tient compte des véhicules roulants, c'est-à-dire avec la base de sondage optimisée.

Nous obtenons :

avec le champ initial $EQM = (0,027)^2 T(Y)^2 + \text{Variance}$

après extension $EQM = (809/588)^k \text{ Variance initiale}$, avec k coefficient compris entre 1 et 2

Les calculs qui ont été menés sur la précision attendue conduisent à un coefficient de variation de l'ordre de 0,6 %.

On a donc :

$$\text{EQM avant} = \{(0,027)^2 + (0,006)^2\} T(Y)^2 = (0,000729+0,000036) T(Y)^2 = 7,65 \cdot 10^{-4} T(Y)^2$$

$$\text{EQM après} = (1,38)^k (0,006)^2 T(Y)^2 = 4,97 \cdot 10^{-5} \text{ ou } 6,86 \cdot 10^{-5} T(Y)^2$$

Pour la variable « nombre de tonnes-kilomètres », l'inclusion de la population des véhicules de 16 à 25 ans dans la base de sondage fournirait, à taille d'échantillon constante et selon les hypothèses faites, une EQM plus faible : le rapport entre l'EQM après et l'EQM avant est compris entre 0,11 et 0,15.

En revanche, pour d'autres variables pour lesquelles le coefficient de variation est sans doute plus élevé (par exemple les estimations des coefficients de variation relatives aux différentes catégories de marchandises, sont comprises entre 2 % et 10 %), on peut penser que le biais va rester du même ordre (environ 2,7 %), et que la part de l'EQM due à la variance va être plus forte, et conduire à une EQM après supérieure à la valeur de l'EQM avant. Ceci devrait être particulièrement le cas pour les variables dont le coefficient de variation est estimé comme supérieur à 3 %.

Calcul de l'EQM pour l'inclusion des DOM

Le biais dû aux véhicules des DOM a été estimé différemment, car nous ne disposons pas de données correspondantes dans les données actuelles de l'enquête. Nous avons utilisé les données d'Esane (Élaboration des Statistiques ANnuelles d'Entreprises – voir Brion, Ph., 2011) dans le secteur transport : l'activité des entreprises d'outre-mer a été comparée avec celle des entreprises de métropole. L'unité de comparaison utilisée est le chiffre d'affaires des entreprises du domaine du transport (code APE entre 49 et 53, ainsi que le code 7712Z). On estime ainsi que les entreprises de métropole ont un chiffre d'affaires moyen par véhicule deux fois plus élevé que leurs homologues des DOM. Nous considérerons alors que les véhicules roulant dans les DOM comptabiliseront deux fois moins de tonnes-kilomètres par an, ce qui permet d'approximer les tonnes-kilomètres par véhicule dans les DOM. En les multipliant par le nombre de véhicules roulants, nous obtenons une estimation du nombre de tonnes-kilomètres réalisées dans les DOM. Le biais est estimé à 0,36 %.

Afin d'évaluer l'EQM, étant donné que le biais a été évalué, on s'intéresse au coefficient multiplicateur N_2/N_1 qui devient $823/809 = 1,017$.

On a dans ce cas :

$$\text{EQM avec le champ initial} = \{(0,0036)^2 + (0,006)^2\} T(Y)^2$$

$$\text{EQM après extension} = (1,017)^k (0,006)^2 T(Y)^2$$

Comme dans le cas de l'inclusion des poids lourds de 16 à 25 ans, l'EQM est plus faible pour l'estimation du total de tonnes-kilomètres à partir de la base étendue, mais la situation s'inverse pour des variables pour lesquelles le coefficient de variation est plus élevé.

Annexe 2 : Estimation de la variance statistique d'une variable dans une strate à partir d'un plan de sondage à probabilités inégales

Il faut estimer sur une strate (appelée U) la quantité $(1/N) \sum_U (Y_i - \bar{Y})^2$ où $\bar{Y} = (1/N) \sum_U Y_i$,

soit estimer $(1/N) \sum_U Y_i^2 - (\bar{Y})^2$

Le sondage « actuel » a été réalisé dans la base de sondage à partir d'un sondage à probabilités inégales.

Les différents termes de l'expression $(1/N) \sum_U Y_i^2 - (\bar{Y})^2$ sont estimés de la façon suivante :

- La taille N de la sous-population U étant inconnue, elle est estimée par la somme des poids de sondage w (après calage) des unités de l'échantillon appartenant à cette sous-population U, soit par $\sum_s w_i 1(i \in U)$.
- $\sum_U Y_i^2$ est estimé par $\sum_s w_i Y_i^2 1(i \in U)$ à partir des unités de l'échantillon appartenant à U.
- $\sum_U Y_i$ est estimé par $\sum_s w_i Y_i 1(i \in U)$ à partir des unités de l'échantillon appartenant à U.

Annexe 3 : Précision obtenue pour le plan de sondage stratifié avec allocation de Neyman et celle obtenue en arrondissant les taux de sondage obtenus avec l'allocation de Neyman, comparées à celle de l'ancien plan de sondage

Le tableau 13 fournit, pour différentes variables, la précision obtenue pour le plan de sondage stratifié utilisant une répartition optimale de Neyman sur la variable tonnes-kilomètres, et de taille 58 000 - colonnes du milieu -, et pour celui avec des taux de sondage « arrondis » (colonne plan de sondage manuel), toujours avec 58 000 questionnaires.

Ces deux plans de sondage sont comparés à l'ancien plan de sondage, utilisant un tirage à probabilités inégales et toujours avec une taille de 58 000 questionnaires, ceci afin d'avoir une comparaison pertinente.

Tableau 13 : Calcul de coefficients de variation (en %) pour différentes variables dans le cadre de l'ancien plan de sondage et du nouveau

	Précision de l'estimation					
	Ancien plan de sondage		Optimum de Neyman (58000)		Plan de sondage manuel	
	Tonnes km	V km	Tonnes km	V km	Tonnes km	V km
Global	0,73%	0,70%	0,61%	0,74%	0,63%	0,69%
Compte Propre	1,84%	1,67%	1,44%	1,68%	1,45%	1,46%
Compte d'Autrui	0,84%	0,80%	0,67%	0,77%	0,70%	0,75%
Marchandises						
M01 : Produits de l'agriculture, chasse, forêt, pêche	1,30%	2,03%	1,40%	1,92%	1,45%	1,73%
M02 : Houille et Lignite ; Pétrole brut et gaz naturel	9,08%	9,13%	6,48%	9,33%	6,80%	9,26%
M03 : Minerais, tourbe et autres produits d'extraction	2,38%	2,28%	1,90%	2,08%	1,96%	2,03%
M04 : Produits alimentaires, boisson et Tabac	1,92%	2,23%	1,36%	2,10%	1,43%	1,82%
M05 : Textiles, cuir et produits dérivés	9,27%	11,27%	8,86%	14,56%	7,61%	10,43%
M06 : Bois, pâte à papier et produits de l'édition	3,64%	4,09%	2,92%	3,70%	2,89%	3,47%
M07 : Coke et produits pétroliers raffinés	4,09%	4,37%	2,96%	4,14%	3,12%	3,96%
M08 : Produits chimiques, caoutchouc, plastique et combustibles nucléaires	3,11%	3,69%	2,34%	3,12%	2,45%	3,08%
M09 : Autres produits minéraux non métalliques	1,87%	2,11%	1,60%	1,99%	1,62%	1,84%
M10 : Métaux de base, produits métalliques	3,50%	4,29%	2,65%	4,42%	2,76%	3,60%
M11 : Machines et matériel non classables ailleurs, produits des TIC* et instruments de précision	3,61%	3,57%	2,83%	3,21%	2,94%	3,16%
M12 : Matériel de transport	4,99%	4,95%	3,79%	4,45%	3,94%	4,36%
M13 : Meubles; autres produits manufacturés non classables ailleurs	5,77%	6,54%	4,59%	5,55%	4,71%	5,34%
M14 : Matières premières secondaires, déchets	3,87%	4,10%	2,90%	3,89%	3,01%	3,54%
M15 : Courrier, Colis	6,30%	5,83%	4,39%	5,17%	4,67%	4,98%
M16 : Équipement pour le transport de fret	4,93%	4,26%	3,88%	3,72%	4,08%	3,60%
M17 : Déménagement, biens non marchands, véhicules en réparation	6,98%	6,57%	6,22%	7,14%	6,32%	6,25%
M18 : Marchandises groupées	1,88%	1,77%	1,28%	1,54%	1,36%	1,51%
M19 : Marchandises non identifiables	6,82%	7,30%	5,02%	5,38%	5,20%	5,55%
M20 : Autres marchandises non classables ailleurs	13,32%	22,44%	13,59%	13,95%	14,73%	16,15%

Source : Enquête TRM 2014