



UNION DES COMORES
Unité – Solidarité – Développement

Ministère des Finances du Budget et du Secteur Bancaire

*Projet d'Investissement de Soutien aux Capacités Statistiques de la Communauté de
Développement de l'Afrique australe-Comores (SATAT CAP-KM)*

**ENQUETE HARMONISEE SUR LES CONDITIONS DE VIE DES
MENAGES EN UNION DE COMORES**

**NOTE TECHNIQUE
SUR L'APUREMENT, LE TRAITEMENT ET LE MODE DE CALCUL
DES INDICATEUR DU BIEN ÊTRE AINSI QUE LA LIGNE DE
PAUVRETE**

MESURE DE LA PAUVRETE

Janvier 2025

1. CONTEXTE

Les enquêtes auprès des ménages sont la source privilégiée pour produire des indicateurs de suivi des conditions de vie des populations. Les autres sources de données présentent en effet des challenges importants. Les statistiques de sources administratives ont souvent une couverture limitée, les recensements coûtent chers...

Afin de fournir des statistiques de bonne qualité, pertinentes, fiables, à jour et comparables, les enquêtes doivent être conduites régulièrement, utiliser des concepts et des définitions identiques et s'appuyer sur les mêmes méthodes de collecte des données. Ainsi, Dans le cadre du Programme Statistique Régional STAT-CAP 2023-2028, la SADC a sollicité et obtenu l'appui technique et financier de la Banque Mondiale pour la mise en œuvre du projet d'investissement de soutien aux capacités statistiques de la communauté de développement de l'Afrique australe.

L'objectif du Projet est de renforcer les capacités statistiques dans les Etats membres y compris dans la réalisation des enquêtes sur les conditions de vie des ménages qui répondent aux normes harmonisées régionales, et de rendre les micro-données ainsi collectées accessibles au public.

L'Etat Comorien, à travers les Instituts Nationaux de la Statistique des Etudes Economiques et Démographiques (INSEED) constitue le niveau opérationnel pour l'exécution des activités statistiques inscrites aux projets du programme. Un accord de rétrocession a ainsi donc été signé entre les INSEED et le ministère de Finance du Budget et du secteur bancaire pour la l'exécution du projet dont la réalisation d'une enquête panel suite à l'enquête harmonisée sur les conditions de vie des ménages (EHCVM) de 2020.

La première EHCVM, s'est déroulée en deux vagues simultanément dans pays de l'Union. la collecte principale des deux vagues ont été achevées respectivement le mars 2020 pour la première et en juillet 2020 pour la deuxième après trois mois de collecte de données sur le terrain pour chaque vague.

Les chiffres de pauvreté et de bien-être de la première enquête, étant validés et disponibles., l'INSEED et la Banque Mondiale se prépare au lancement officiel de la deuxième EHCVM le en janvier 2024. Cette collecte est prévue pour se dérouler sur l'année 2024

2.Objectifs de l'enquête

L'objectif principal du projet est d'améliorer la fréquence, la pertinence et la qualité des enquêtes sur les conditions de vie des ménages dans les Etats membres.

Les objectifs spécifiques de l'enquête de 2024 sont :

- o Produire des données exhaustives permettant de mener des analyses approfondies pour informer les politiques publiques, comme en 2020 ;
- o Produire des indicateurs de pauvreté comparables avec ceux de 2020 ;

- o Mesurer l'impact des chocs exogènes que le pays a subi dans les cinq dernières années sur le bien-être des ménages (des questions récentes des politiques publiques que, l'enquête doit en tenir compte) ;
- o Analyser la dynamique de la pauvreté .

Section 2 : PREPARATION DES DONNEES

2.1. Principes de base

Les données d'une enquête sont utilisées par l'institution qui en est le maître d'œuvre, mais aussi par d'autres institutions, des chercheurs, des étudiants, etc. Des travaux d'analyses peuvent intervenir plusieurs années après la fin des travaux de collecte. La mémoire humaine n'est pas toujours fiable ; même quand les données sont analysées dans l'institution qui les a produites, les individus ayant travaillé sur la conception et la mise en œuvre de l'enquête peuvent avoir oublié des détails importants ; ces personnes peuvent ne plus être présentes pour cause de mobilité professionnelle, départ à la retraite ; etc. Pour que le potentiel des données produites soient pleinement exploitées, il faut que les bases de données soient bien documentées avant d'être archivées. Cette section traite des règles de base pour nommer les fichiers bruts et les variables.

Nomination des fichiers bruts et variables

Un fichier de données quelconque va porter le nom suivant :

sab_cd_XYZ2020 où :

- ❖ s : est fixe et désigne « section »,
- ❖ ab : le numéro de la section du questionnaire.
(Exemple 01 pour la section 1, 02 pour la section 2, 16 pour la section 16, etc.),
- ❖ cd : est le code du questionnaire.
(Exemple, me pour les fichiers ménages, et co pour les fichiers communautaires),
- ❖ XYZ : code pays à trois caractères (pour Comores c'est : COM).
(Exemple, le fichier ménage de la section 6 (Epargne et crédit) de l'enquête pour les Comores, sera nommé s06_me_COM2020.

Lorsque le nom de la section contient une lettre, le nom du fichier incorporera la lettre.

Exemple, les fichiers de la section 8 (Sécurité alimentaire - Partie A : Echelle d'expérience d'insécurité alimentaire : s08A_me_COM2020.

Partie B : Complément sur la consommation alimentaire des 7 derniers jours) :
s08B_me_COM2020.

Les fichiers issus directement du logiciel *Survey Solutions* ont des noms différents de ceux mentionnés ci-dessus. Le transfert des fichiers de *Survey Solutions* à STATA n'a pas suivi la logique des sections du questionnaire. Il y a certains fichiers issus de *Survey Solutions* qui contiennent les variables de plusieurs sections différentes ; et il y a des cas où des enregistrements d'une même section se retrouvent dans des fichiers distincts. Un programme STATA sera préparé pour convertir les fichiers actuels en fichiers standards tel que décrit ci-dessus. Dans certains cas, cela nécessitera que plusieurs fichiers soient fusionnés (appariés ou additionnés), dans d'autres cas, il faudra qu'un fichier soit scindé en plusieurs.

La duplication d'enregistrement n'est pas possible, les fichiers créés par le logiciel *Survey solution* contiennent une variable d'identification (`interview__id`) unique pour chaque ménage. Bien que lors de la conception de l'enquête, il a été prévu quatre (4) variables pour identifier un ménage :

Le numéro de zone de dénombrement (ZD) (unité primaire de sondage) , le numéro de la concession dans le ZD , le numéro de ménage dans la concession et le numéro de l'unité d'habitation dans la concession , ceci permettra de créer un autre identifiant de ménage si le besoin se présente lors du traitement et les analyses .

Pour les fichiers individus, les variables d'identification sont les variables d'identification du ménage et le numéro de l'individu dans le ménage. Quant aux bases de données communautaires, la variable d'identification est la ZD .

Nomination des variables :

La règle pour nommer les variables est la suivante :

sabqxx où s est fixe :

- ❖ ab : Le numéro de la section du questionnaire (exemple 01 pour la section 1, 02 pour la section 2, 16 pour la section 16, etc.),
- ❖ q : est fixe
- ❖ xx : le numéro de la question.

Exemple, la question 01 de la section 6 du questionnaire ménage : « *Est-ce que [NOM] possède un compte dans les établissements financiers suivants ou une carte prépayée ?* » sera nommée s06q01.

Lorsque le nom de la question contient une lettre, la lettre fait partie du nom de la variable. Exemple la question 8A.07a de la section 8A du questionnaire ménage :

« Au cours des 12 derniers mois, est-il souvent arrivé que vous ou d'autres membres du ménage ayez faim mais n'avez pas mangé parce qu'il n'y avait pas suffisamment d'argent ou de ressources pour la nourriture ? Cela est-il arrivé une ou deux fois, pendant quelques mois mais pas tous les mois, ou presque tous les mois ? » :

La question sera nommée s08Aq07a.

Labélisation des variables et des modalités des variables qualitatives

Les variables des fichiers transférés de *Survey Solutions* n'ont pas de labels, elles doivent être labélisé.

S'agissant des modalités de variables qualitatives, un grand nombre de variables transférées de *Survey Solutions* en STATA n'en ont pas ; les variables qualitatives seront systématiquement labellisées dans le programme STATA qui sera préparé pour convertir les fichiers actuels en fichiers standards.

2.2. Traitement des données manquantes et des valeurs aberrantes

Comme il a été mentionné ci-dessus, l'application de collecte a été écrite avec *Survey Solutions*, un logiciel pour les interviews personnelles assistées par ordinateur. L'application comprend un grand nombre de contrôles de cohérence entre les variables d'un individu pour une même section, les variables d'un individu de sections différentes, et même les variables d'individus différents. En outre, un programme écrit en STATA permet de contrôler les questionnaires et de rejeter ceux ne répondant pas à un certain nombre de critères de qualité. En conséquence, une partie importante de l'apurement des données s'effectue en même temps que la collecte. Néanmoins, il y a des cas où faute d'avoir accès à l'internet, les données n'ont pas été transférées à l'équipe centrale suffisamment à temps pour vérification. Il y a aussi des cas où malgré les rejets, la correction sur le terrain n'a pas été possible. Un complément de travail d'apurement après la collecte s'avère nécessaire. Dans le prolongement de ce travail d'apurement et pour préparer les données à l'analyse, il faut traiter les cas de non-réponse (données manquantes) et procéder à la correction des valeurs aberrantes.

Dans le processus de correction, il est important de souligner que les bases de données brutes doivent refléter les données collectées sur le terrain. Ainsi ces bases conservent les valeurs improbables, les données manquantes et les cas qui n'ont pas été résolus sur le terrain. Le traitement des données manquantes et les corrections de valeurs aberrantes sont seulement répercutés sur les bases de données devant servir aux travaux d'analyses, sauf dans des cas exceptionnels où une correction évidente s'impose d'elle-même.

On considère qu'une donnée est manquante s'il manque une information là où elle aurait dû exister. Pour prendre un exemple, la section emploi est renseignée pour les individus de 5 ans et plus, pour les personnes de moins de 5 ans, cette section n'a aucune information, il ne s'agit pas d'une donnée manquante. Par contre, il y aurait une donnée manquante pour une personne de 5 ans qui n'aurait pas répondu à la première question de cette section emploi. Il existe plusieurs manières de traiter les données manquantes. Premièrement, on peut adopter une solution radicale et supprimer toute l'information d'un individu avec au moins une variable non-remplie. Ce n'est évidemment pas la meilleure des solutions. Les réponses manquantes ne sont pas toujours aléatoires et l'élimination de tout l'enregistrement avec une ou deux réponses manquantes conduit à perdre de l'information utile pour toutes les autres questions bien renseignées. Par exemple dans la section emploi, certains salariés peuvent ne pas déclarer leur salaire et fournir de bonnes réponses par ailleurs, il n'est pas indiqué de supprimer tout l'enregistrement. Deuxièmement, on peut laisser les réponses manquantes telles quelles et laisser le soin à l'analyste d'opérer ses propres choix. Troisièmement, on peut remplacer les informations manquantes par imputation. On utilise les réponses valides pour imputer une valeur à la réponse manquante. Cette troisième solution sera le plus souvent adoptée.

Un certain nombre de précautions doivent être prises. Premièrement, avant de pouvoir procéder à l'imputation, il faut vérifier le pourcentage de non-réponse (NR) parmi les individus statistiques faisant partie du domaine de définition de la variable ; raisonnablement on tolère jusqu'à 3 pourcent de NR. Quand le taux de NR est inférieur à ce seuil, on peut procéder à l'imputation des valeurs manquantes sans crainte d'altérer les données de base. En matière de procédure d'imputation, les variables qualitatives et les variables numériques sont traitées différemment. Pour les variables qualitatives, l'imputation consiste généralement à remplacer la valeur manquante par la valeur modale de la distribution ; on peut affiner la procédure en introduisant par exemple le milieu de résidence ou d'autres variables explicatives pertinentes selon le cas (région). Quand le pourcentage de NR est faible, par exemple inférieur à 2 pourcent, il n'est pas nécessaire de compliquer la procédure. S'agissant des variables numériques, la valeur manquante est généralement remplacée par la médiane des réponses non manquantes. Selon les cas cette imputation sera aussi affinée en prenant en compte des variables explicatives telles que le sexe, l'âge, la région, le milieu de résidence, l'appartenance ethnique, etc. Un cas particulier d'intérêt est celui des valeurs manquantes des quantités consommées. La consommation est une des variables les plus importantes de l'enquête car c'est le principal ingrédient pour construire l'indicateur de mesure du bien-être. Avant de procéder à l'imputation des quantités consommées, il faut procéder à une double-vérification. D'abord s'assurer que le taux de NR dans l'échantillon pour chaque item de consommation est inférieur à 3 pourcent comme mentionné ci-dessus. Ensuite vérifier que pour chaque ménage donné, le pourcentage de NR en termes d'items de consommation dans le ménage (en faisant bien la différence entre zéro consommation et NR, ce qui est faisable dans le questionnaire) est aussi inférieur à 3 pourcent. En effet si le taux de NR, en termes d'items de consommation, est élevé dans le ménage, des imputations créeraient simplement un questionnaire artificiel. Un taux de NR élevé sur les items de consommation va conduire à invalider le questionnaire pour les travaux d'analyse. Si les deux conditions ci-dessus sont réunies, l'imputation est légitime, on impute la NR par la médiane de la quantité consommée dans la région et le milieu de résidence.

Le traitement des valeurs aberrantes ne concerne que les variables numériques. Dans ce document on ne traite pas de l'ensemble des variables, on ne considère que les variables de consommation (quantité et valeurs) qui entrent dans la construction de l'agrégat de bien-être. D'autres variables, par exemple celles de revenus, superficies cultivées, récoltes, etc. ne sont pas traitées ici. Leur traitement interviendra plus tard, par exemple quand il faudra construire un agrégat de revenu.

Si la définition d'une valeur manquante est simple, celle d'une valeur aberrante l'est moins. Un bon exemple d'une valeur exceptionnelle qui n'est pas aberrante est la superficie de la Russie en Europe ; elle est trop grande comparée à la moyenne de celles des autres pays européens, pour autant elle est valide. Il faut donc opérer une distinction entre une valeur exceptionnelle et une valeur aberrante. On peut dire que les valeurs aberrantes sont des valeurs irréalistes, en général des valeurs trop faibles ou trop élevées par rapport aux normes connues. Pour les traiter, il est nécessaire de les identifier.

On considère d'abord la consommation alimentaire. Elle est renseignée en quantité (à l'exception des repas pris hors ménage qui le sont en valeurs comme le reste de la consommation). Deux approches sont utilisées pour tenter d'identifier les valeurs aberrantes de consommation.

Pour la première approche, l'on commence par convertir les quantités renseignées en unités non-standard lors de la collecte en unités standards, et ensuite des quantités en unités standards on passe à la consommation en calories. On considère comme valeurs aberrantes les consommations par tête inférieures à 800 kilocalories par personne et par jour et celles supérieures à 4000 kilocalories par personne et par jour ; (il faut noter que ce travail a été déjà fait lors de la collecte, mais comme il a été signalé ci-dessus, compte tenu de l'absence de l'internet dans certaines parties du pays, les corrections n'ont pas toujours été possibles sur le terrain). Les corrections ne seront pas automatiques, les valeurs a priori identifiées comme aberrantes seront scrutées à l'aune d'autres variables. De plus avant toute correction, on établit aussi la distribution de calories en fonction des sous-groupes de produits (céréales, légumes, légumineuses et graines, viandes, poissons, huiles et graisses, etc.).

Considérons d'abord les valeurs de consommation trop faibles, celles inférieures à 800 kilocalories. Deux pistes seront explorées pour les corrections. D'abord on fait appel à la consommation hors ménage. A l'aide de la consommation du ménage, on calcule le prix médian ou le prix moyen d'une calorie, par exemple pour les ménages du 5^{ème} au 95^{ème} décile (il faut exclure les valeurs aberrantes que l'on essaie de corriger pour une moyenne robuste). On majore cette valeur médiane (ou moyenne) par un pourcentage, par exemple 30 pourcent, en faisant l'hypothèse qu'un repas acheté à l'extérieur coûte plus cher qu'un repas préparé à la maison. A l'aide cette valeur médiane majorée, on procède à une estimation grossière de l'apport calorique des repas pris à l'extérieur en divisant la valeur de la consommation des repas pris à l'extérieur par le prix médian d'une calorie consommée. Ensuite à l'aide de la distribution de la consommation en sous-groupes, on identifie les sous-groupes ayant un apport trop faible alors que certains produits de ces sous-groupes figurent dans les ménages considérés. On cherche alors s'il n'y a pas eu de mauvais report évident d'unités.

S'agissant des consommations de calories trop élevées, l'on procède aussi en deux temps. D'abord à l'aide de la section 8B, l'on essaie de détecter s'il n'y a pas eu un grand nombre de personnes

extérieures au ménage ayant pris des repas dans le ménage ; si c'est le cas, des valeurs apparemment anormales deviennent plausibles. Ensuite comme ci-dessus à l'aide de la distribution de la consommation en sous-groupes, on identifie les sous-groupes ayant un apport trop élevé en calories ; on cherche alors s'il n'y a pas eu de mauvais report évident d'unités.

Une fois de plus les corrections ci-dessus seront faites au cas par cas ; lorsque l'examen minutieux des données conduit à voir qu'il y a des erreurs évidentes. En effet une correction aveugle des valeurs dites aberrantes peut conduire à créer des données artificielles et à gommer les inégalités réelles qui existent dans la distribution.

La deuxième approche de correction des valeurs aberrantes est la plus classique. Au niveau de chaque item de consommation, on identifie comme valeurs aberrantes celles qui ont une quantité par tête supérieure à la moyenne plus 2 écart-types, la moyenne plus 3 écart-types ou les 1 pourcent des valeurs les plus élevées (on ne s'intéresse pas à la consommation trop faible au niveau de chaque produit car certains ménages peuvent ne pas en consommer). Une fois les valeurs anormales identifiées à l'aide de ces trois critères, il faut prendre une décision. Dans la majeure partie des cas, la valeur aberrante sera remplacée par la médiane dans la région et le milieu de résidence. Mais comme on l'a mentionné ci-dessus, une correction brutale des valeurs aberrantes peut conduire à gommer l'inégalité. Des alternatives sont prévues, après examen des indicateurs d'inégalité, par exemple remplacer plutôt les valeurs supérieures au centile 99 par la valeur correspondante à ce centile, ce qui suppose implicitement une correction de 1 pourcent de l'échantillon.

Jusqu'à présent la correction des valeurs aberrantes a porté sur la consommation alimentaire. S'agissant de la consommation non-alimentaire et des repas pris à l'extérieur, le traitement est similaire à celui de la deuxième approche de la consommation alimentaire. La détection des valeurs aberrantes se fait par item de consommation. La consommation non-alimentaire ayant une distribution plus dispersée, on retient comme valeurs aberrantes celles supérieures à la moyenne plus 3 écart-types ou les 1 pourcent des valeurs les plus élevées ; on procède pareillement à l'imputation quand l'examen attentif des données pousse à procéder aux corrections.

2.3. Fichiers pour les travaux d'analyse

A partir des données de base, trois bases de données sont créées pour les travaux d'analyse, particulièrement ceux portant sur les questions de pauvreté. Les trois bases de données sont les suivantes : i) un fichier individu ; ii) un fichier ménage-produit ; iii) un fichier ménage avec les principales variables pour l'analyse du bien-être.

Au préalable, il faut mentionner qu'un identifiant unique du ménage sera créé et introduit dans toutes les bases de données d'analyse. Cet identifiant sera composé du numéro de la grappe et du numéro du ménage dans la grappe. Il est donné par $hhid = grappe * 100 + menage$.

La première base de données est un fichier individu qui comprend un ensemble de variables des membres des ménages, pertinentes pour les travaux d'analyse sur les individus. Certaines de ces variables existent dans les fichiers de base (caractéristiques sociodémographiques, éducation, emploi), d'autres doivent être créées à l'aide de variables existants. Certaines des variables qui

existent déjà et qui vont figurer dans ce fichier sont le sexe, l'âge, le lien de parenté avec le chef de ménage, la fréquentation scolaire, etc. Par contre, les variables pertinentes du marché travail doivent être créées ; il en est ainsi par exemple de la situation par rapport à l'activité (actif occupé, chômeur, inactif, pas en âge de travailler). D'autres variables doivent être harmonisées, par exemple les variables d'éducation. Il est important de souligner que cette base de données ne va pas comprendre l'ensemble des informations des individus des ménages, sinon il suffirait de simplement apparier ces fichiers individus ; elle inclura entre 30 et 60 variables le plus souvent utilisées pour les analyses. Il est aussi important de rappeler qu'à la différence des fichiers bruts, les variables figurant dans bases de données à créer auront subi les corrections et imputations comme mentionné à la section 2.2.

La deuxième base de données est un fichier ménage-produit. Ce fichier va contenir l'essentiel de l'information pour construire l'indicateur de bien-être. Il va avoir comme principale variable la dépense annuelle par ménage et par produit de l'ensemble des près de 300 produits ayant fait l'objet du relevé de dépenses. Dans bien des travaux préparatoires à l'analyse de la pauvreté, l'agrégat de bien-être est construit et les outputs intermédiaires ne sont pas conservés. En conséquence lorsque vient le moment de faire des travaux portant par exemple sur l'incidence des exonérations fiscales sur certains produits de consommation alimentaires (riz, blé, huiles végétales, etc.) et non-alimentaires (produits pétroliers, électricité, eau courante, etc.), l'analyste reprend l'exercice de refaire l'agrégat avec moins d'information pour procéder par exemple à l'imputation des valeurs manquantes et à la correction des valeurs aberrantes. Il est donc important de créer ce fichier intermédiaire et le conserver, l'effort marginal est quasi-nul car c'est une étape obligatoire dans la construction de l'indicateur de bien-être.

Il faut noter que dans ce fichier, l'on devrait exclure les produits ne faisant pas l'objet de la construction de l'indicateur de bien-être, notamment les dépenses de construction ou de grosses réparations de logement et les dépenses en acquisition de biens durables. On peut néanmoins prendre l'option de les y inclure (cela peut se justifier par les autres utilisations possibles de la base de données, par exemple répondre rapidement à une requête de la comptabilité nationale). Dans ce dernier cas, on créera une variable indiquant que des produits sont inclus ou exclus de l'agrégat de bien-être. Ce fichier sera très volumineux, il peut avoir jusqu'à 300 000 enregistrements, mais le nombre de variables est limité. Les variables de cette base de données sont les suivantes :

- Identifiant unique du ménage
- Numéro de grappe
- Numéro du ménage
- Région
- Milieu de résidence
- Strate
- Code produit
- Mode d'acquisition
- Dépense annuelle

Le code produit est le code attribué à l'item de consommation dans le questionnaire. Les items qui figurent dans les sections individuelles du questionnaire (dépenses d'éducation, dépenses de santé, achat de cartes prépayées pour téléphone mobile), dans la section logement (loyer, eau courante,

eau achetée à des revendeurs, électricité, frais d'abonnement au téléphone fixe, frais d'abonnement à l'internet, frais d'abonnement au câble) et dans la section portant sur les repas pris à l'extérieur n'ont pas de code. Un code est attribué dans le programme qui crée ce fichier (les codes sont proposés à la section 3). La variable mode d'acquisition permet de distinguer la consommation du ménage selon la provenance. En effet d'un point de vue analytique, il est important de distinguer les achats de l'autoconsommation et des autres sources ; non seulement pour connaître l'origine de l'offre des produits consommés, mais aussi pour d'autres besoins, par exemple une analyse portant seulement sur la consommation monétaire (l'analyse des exemptions fiscales a déjà été mentionné ci-dessus). La variable mode d'acquisition prend les modalités suivantes : achat (code 1), autoconsommation (code 2), cadeau (code 3), valeur d'usage (code 4), loyer imputé (code 5), autoconsommation alternative (code 6).

La troisième et dernière base de données est un fichier ménage comprenant les principales variables nécessaires pour les analyses de base de la pauvreté. Ce fichier comprend principalement les variables suivantes :

- Identifiant unique du ménage
- Numéro de grappe
- Numéro du ménage
- Région
- Milieu de résidence
- Strate
- Coefficient d'extrapolation
- Taille du ménage
- Nombre d'adultes équivalents
- Sexe du chef de ménage
- Age du chef de ménage
- Situation matrimoniale du chef de ménage
- Nationalité du chef de ménage
- Niveau d'éducation du chef de ménage
- Situation par rapport à l'activité du chef de ménage (7 derniers jours)
- Situation par rapport à l'activité du chef de ménage (12 derniers mois)
- Profession du chef de ménage (12 derniers mois)
- Branche d'activité du chef de ménage (12 derniers mois)
- Secteur institutionnel de chef de ménage (12 derniers mois)
- Catégorie socioprofessionnelle du chef de ménage (12 derniers mois)
- Existence d'un emploi secondaire pour le chef de ménage (12 derniers mois)
- Dépenses alimentaires
- Dépenses non alimentaires
- Dépenses totales
- Déflateur
- Indicateur de bien-être
- Seuil national de pauvreté
- Quintile de bien-être

2.4. Gestion des bases de données

Un aspect important de la préparation des données est la gestion des bases de données. En effet il arrive fréquemment qu'une base de données existe en plusieurs versions. Si la gestion des fichiers n'est pas faite avec rigueur et méthode, un utilisateur peut se retrouver avec des bases qui ne lui conviennent pas, soit avec moins d'informations qu'il faudrait (par exemple un fichier semi-apuré alors qu'il y a eu des travaux postérieurs), soit avec plus d'informations que nécessaire (par exemple la mise à la disposition d'utilisateurs extérieurs à l'INS de fichiers avec des variables devant préserver la confidentialité des répondants). Pour une gestion rigoureuse des fichiers de données, il est suggéré de créer trois répertoires ; ils sont intitulés *dataraw*, *datain* et *dataout*.

Le premier répertoire, *dataraw*, contient les fichiers bruts. Les bases de données brutes obéissent aux caractéristiques suivantes : i) ce sont les fichiers directement sortis de *Survey Solutions*¹ et convertis en fichiers STATA (ou SPSS) ; ii) ces fichiers contiennent toutes les variables, sans restriction, notamment les variables relatives à la confidentialité des répondants ; iii) ces fichiers contiennent tous les ménages, y compris les ménages n'ayant pas répondu à l'enquête (conserver ces ménages est important pour les travaux méthodologiques) et les ménages qui pourraient être supprimés plus tard des fichiers de travail, du fait de la mauvaise qualité des réponses ; iv) ces fichiers ont été apurés dans la logique du questionnaire, en tenant compte des erreurs de cohérence, mais il n'y a pas d'imputation de valeurs manquantes, ni de traitement des valeurs aberrantes.

Le deuxième répertoire, *datain*, va contenir les fichiers bruts ayant subi quelques modifications et qui ont vocation à être mises à la disposition des utilisateurs. Il y a trois différences avec les bases de données brutes de *dataraw* : i) contrairement aux fichiers bruts de départ, les variables d'anonymisation ont été ôtées de ces fichiers ; ii) les ménages n'ayant pas répondu aux questionnaires sont également retirés ; iii) des corrections mineures peuvent être faites sur ces bases de données, au-delà des corrections portant sur les erreurs de cohérence. Si ce genre de correction est faite, elle doit être documentée. Il est pratique de distinguer les bases de données de ce deuxième groupe de celles du premier, puisqu'elles portent le même nom. Les bases de données de ce groupe vont conserver les noms proposés à la section 2.1, cela rend les choses faciles car elles ont pour vocation à être mises à la disposition des utilisateurs. Les bases de données du premier groupe (*dataraw*) auront un ajout dans leur nom, « brut ». Ainsi pour reprendre un des exemples déjà pris, le fichier ménage de la section 6 (Epargne et crédit) de l'enquête du Bénin, version brute, sera nommé *s06_me_BEN2018_brut*. Dans le fichier à mettre à la disposition des utilisateurs, il faudra ôter « brut ».

Le troisième répertoire, *dataout*, est dédié aux bases de données créées pour les travaux d'analyse. Ces bases de données sont également mises à la disposition des utilisateurs. Il est important que les bases de données mises à la disposition des utilisateurs soient accompagnées de métadonnées, par exemple un document comme celui-ci qui expliquent les procédures de corrections et

¹ Comme il a été noté ci-dessus, les fichiers directement exportés de *Survey Solutions* ne sont pas standardisés en ce sens qu'ils ne correspondent pas aux sections des questionnaires. Une application STATA va permettre de transformer ces fichiers en standards ; mais ils auront le même contenu que les bases de données directement exportées de *Survey Solutions*, seule leur organisation change.

d'imputation des données, de création de variables, etc. les do-files ayant servis à construire l'indicateur de bien-être, etc.

3. CONSTRUCTION DE L'INDICATEUR DE BIEN-ETRE

3.1. Fondements théoriques et choix méthodologiques

3.1. 1. Fondements théoriques

La pauvreté est un phénomène complexe et multidimensionnel. On peut l'examiner sous l'angle des libertés individuelles, d'accès à l'éducation, d'accès à la santé, etc. Mais de toutes ces dimensions, la plus courante est la pauvreté monétaire. Pour cette dimension, une personne est pauvre si elle n'a pas le minimum de ressources nécessaires pour satisfaire ses besoins fondamentaux (se loger, se nourrir, se vêtir, etc.). Mesurer la pauvreté passe par deux étapes, d'abord identifier les ménages ou individus pauvres et ensuite calculer les indicateurs permettant de cerner l'ampleur et la profondeur du phénomène. La première étape, l'identification des pauvres nécessite elle aussi deux choses : i) un indicateur de mesure du bien-être ; ii) un seuil de pauvreté, c'est-à-dire un niveau de l'indicateur de bien-être en deçà duquel un ménage sera considéré comme pauvre.

Les fondements théoriques² de la mesure du bien-être relèvent de la théorie du comportement du consommateur ; notamment du concept de "*métrique monétaire de l'utilité*"³. On suppose que l'économie comprend n biens, $i = 1, 2, \dots, n$. Un ménage dispose d'un revenu x , et fait face à un système de prix p ($p_i, i = 1, \dots, n$); on appelle u l'utilité du consommateur. On définit la fonction de dépense $e(u, p)$ comme la dépense minimale pour atteindre le niveau d'utilité u . Puisque le ménage minimise sa dépense, on a $e(u, p) = x$.

Si on considère un ménage h en particulier, on note u^h son utilité et on suppose que le ménage fait face à un système de prix p^h et consomme les quantités q^h . On aussi suppose un système de prix de référence p^0 . Pour le ménage h , la dépense minimale pour atteindre le niveau d'utilité u^h avec le système de prix de référence est noté :

$$u_m^h = e(u^h, p^0)$$

Pour mesurer le bien-être, il faut évaluer cette dépense. La difficulté que cela pose est que l'utilité n'est pas observable. Si on fait appel à l'identité de Roy qui montre que la dérivée de la fonction de dépense par rapport au prix est la quantité, et en faisant un développement limité du premier ordre, on a :

$$u_m^h = e(u^h, p^0) \approx e(u^h, p^h) + (p^0 - p^h)q^h = p^h q^h + p^0 q^h - p^h q^h = p^0 q^h$$

² Les développements théoriques de cette section sont largement inspirés de Deaton et Zaidi (2002).

³ Le concept de "*métrique monétaire de l'utilité*" a été proposée par Paul Samuelson, 1974. "A Complementary - An Essay on the 40th anniversary of the Hicks-Allen revolution in the demand theory", *Journal of Economic Literature*, 15, 24-55.

Ainsi, u_m^h qui est approximativement le produit des quantités consommées dans le ménage, multipliées par les prix de référence, peut être utilisée comme une mesure du bien-être du ménage. En fait le problème n'est toujours pas résolu car les quantités ne sont pas systématiquement renseignées dans les enquêtes. S'il peut être possible de mesurer les quantités consommées de produits alimentaires, l'exercice est plutôt compliqué pour les produits non-alimentaires. En revanche, les valeurs de la consommation sont moins difficiles à renseigner. Pour cette raison, on essaye d'écrire u_m^h comme une fonction des dépenses de consommation et d'un indice des prix. On va considérer un indice spatial des prix de Paasche, pondéré par les quantités dont la formule est la suivante :

$$P_p^h = \frac{p^h \cdot q^h}{p^0 \cdot q^h}$$

On a alors :

$$u_m^h \approx p^0 q^h = p^0 q^h \frac{p^h q^h}{p^h q^h} = p^h q^h \frac{p^0 q^h}{p^h q^h} = \frac{x^h}{P_p^h}$$

Ainsi un premier indicateur de mesure du bien-être du ménage est la dépense de consommation du ménage ajustée par un indice spatial des prix de Paasche. Il est important de relever que dans cette formule, chaque ménage a son propre indice des prix, qui dépend des prix du marché et de la structure de consommation du ménage.

Cette approche pose un certain nombre de problèmes. D'abord du point de vue du dialogue social, il n'est pas aisé d'expliquer aux non-spécialistes que chaque ménage a son propre indice des prix ; en fait certains pensent que cette approche n'est pas transparente. Sur le plan pratique ensuite, calculer un indice des prix pour chaque ménage rend l'exercice difficile, d'autant plus difficile qu'une enquête est sujette à des erreurs de collecte et plus on va dans le détail, plus les erreurs de collecte affectent les résultats.

L'alternative à cette approche est celle du "*ratio de mesure du bien-être*"⁴. On définit un ratio de bien-être comme le rapport entre la dépense minimale pour atteindre le niveau d'utilité u^h avec le système de prix p^h , et la dépense minimale pour atteindre l'utilité quand on se situe au niveau d'un seuil de pauvreté z , avec le même système de prix. Formellement, on a :

$$wr^h = \frac{e(u^h, p^h)}{e(u^z, p^h)}$$

La quantité ci-dessus est un nombre sans dimension dans un contexte où un indicateur monétaire est plus approprié. On définit une quantité monétaire, le "*ratio de mesure du bien-être*" qui est la grandeur précédente multipliée par la dépense minimale pour atteindre le niveau d'utilité au niveau du seuil de pauvreté avec le système des prix de référence :

⁴ Le concept de "*ratio de mesure du bien-être*" a été proposée par Blackorby et Donaldson (1988). "Welfare ratios and distributionally sensitive cost-benefit analysis", *Journal of Public Economics*, 34, 265-90.

$$u_r^h = \frac{e(u^h, p^h)}{e(u^z, p^h)} \times e(u^z, p^0) = e(u^h, p^h) \div \frac{e(u^z, p^h)}{e(u^z, p^0)}$$

De cette formule, et compte tenu des résultats précédents, il apparaît que le ratio de mesure du bien-être est le produit de la dépense du ménage divisée un deuxième terme exprimé sous forme de rapport. Ce deuxième terme est le rapport de la dépense minimale pour atteindre le niveau d'utilité u^z (au niveau du seuil de pauvreté) quand le ménage fait face au système de prix courant, par la dépense minimale pour atteindre le niveau d'utilité u^z quand le ménage fait face au système de prix de référence. Ainsi présenté, ce deuxième terme est le vrai indice du coût de la vie pour les ménages se situant au niveau du seuil de pauvreté. En procédant comme précédemment au développement limité de la fonction de dépense au voisinage de p^h , on montre que ce vrai indice du coût de la vie peut être approximée par un indice de Laspeyres.

$$P_{Lz}^h = \frac{p^h q^z}{p^0 q^z} = \sum_{i=1}^n \frac{p_i^0 q_i^z}{p^0 q^z} \times \frac{p_i^h}{p_i^0} = \sum_{i=1}^n w_i^{z0} \times \frac{p_i^h}{p_i^0}$$

Ainsi on peut écrire :

$$u_r^h = \frac{e(u^h, p^h)}{P_{Lz}^h} = \frac{p^h q^h}{P_{Lz}^h} = \frac{x^h}{P_{Lz}^h}$$

Ainsi un autre indicateur de mesure du bien-être du ménage est la dépense de consommation du ménage ajustée par un indice spatial de prix de Laspeyres. Il faut avoir à l'esprit que le premier indicateur de mesure du bien-être présente l'avantage d'être celui qui dérive directement de la théorie du consommateur ; le second est plus facile à calculer et à expliquer.

Dans l'un ou l'autre cas, on peut remarquer que cette approche théorique semble ne pas prendre en compte des différences existant dans la composition des ménages. Or il est clair qu'un ménage de quatre personnes a plus de besoin qu'un ménage d'une personne, ne serait-ce que sur le plan alimentaire. De même un ménage composé de personnes âgées aura des besoins différents de ceux d'un ménage composé de jeunes adultes, par exemple en matière de santé. La simplification dans la présentation ci-dessus ne fait pas ressortir ce facteur. Mais il faut se rappeler que pour un ménage donné, la fonction de dépense est la dépense minimale pour atteindre un niveau d'utilité donné. Or dans la théorie du consommateur, l'utilité du ménage ne dépend pas seulement des quantités consommées, mais aussi des caractéristiques du ménage. Ainsi la fonction de dépense devrait s'écrire $e(u^h, p^h, a^h)$, où a désigne les caractéristiques spécifiques du ménage. Supposons que ce ménage h soit le ménage de référence ou dans une situation de référence ; supposons que les caractéristiques de ce ménage changent (par exemple une naissance dans le ménage) et que le ménage ait maintenant les caractéristiques du ménage j . On peut définir la quantité :

$$cv = e(u^h, p^h, a^j) - e(u^h, p^h, a^h)$$

Cette quantité (variation compensée) est la mesure de ce dont le ménage h a besoin après le changement de composition, pour retrouver son niveau initial de bien-être. On peut aussi définir le ratio :

$$eq = \frac{e(u^h, p^h, a^j)}{e(u^h, p^h, a^h)}$$

Ce ratio est une échelle d'équivalence. Si on a deux ménages faisant face au même système de prix et ayant le niveau de bien-être, le ratio ci-dessus indique de combien la dépense du ménage de caractéristiques a^j devrait être divisée pour atteindre le même niveau de bien-être que le ménage ayant les caractéristiques a^h .

Il s'agit de partir de cette approche théorique de construire un indicateur de mesure du bien-être. Dans cet exercice on utilise les données issues de l'enquête auprès des ménages, notamment les modules comportant les dépenses de consommation.

3.1. 2. Agrégation des dépenses de consommation

L'exercice de construire un indicateur de bien-être passe par trois grandes étapes. Dans un premier temps, l'on agrège les différentes rubriques de la valeur de la consommation du ménage et l'on obtient un agrégat de consommation. Ensuite, cet agrégat est normalisé pour prendre en compte les différences de composition des ménages. Enfin sont pris en compte les différences de coût de la vie entre les ménages. Il est important de signaler que la période de référence pour les travaux analytiques est l'année, ainsi l'agrégat de consommation doit être annualisé. Généralement, la valeur annuelle de la consommation est le produit de la valeur sur une période donnée multiplié par la fréquence. Par exemple si la consommation alimentaire est renseignée sur 7 jours, la dépense déclarée est multipliée par 365 et divisée par 7. Pour les composantes non-matérielles de la consommation (valeur d'usage des biens durables, loyer imputé des ménages propriétaires), les valeurs calculées sont automatiquement annuelles dès lors que les variables utilisées dans le calcul sont elles-mêmes annualisées. Par exemple si loyer des ménages locataires est annualisé, le loyer imputé des ménages propriétaires estimé à partir d'un modèle du loyer des locataires sera aussi annualisé.

La consommation inclut la consommation alimentaire et non-alimentaire. Pour la consommation alimentaire, l'agrégat va inclure la consommation monétaire des produits achetés ; la consommation alimentaire non-matérielles résultant de l'autoconsommation et d'autres sources (cadeau, prélèvements d'un commerce du ménage, troc) et la valeur monétaire des repas consommés hors du domicile. La consommation alimentaire (sauf repas pris hors du domicile) est principalement renseignée en quantité, et il faut trouver un système de prix pour la valoriser. Cette question pratique est traitée dans la section 3.2. Un repas pris hors du ménage est soit un plat cuit acheté dans un restaurant (formel ou informel) et consommé soit dans le lieu où il a été acheté, soit à la maison, soit dans un autre lieu comme au bureau. Les repas pris à l'extérieur (achat ou cadeau) sont directement renseignés en valeur ; ces valeurs sont annualisées et intégrées dans l'agrégat de consommation alimentaire.

L'agrégation de la consommation non-alimentaire commence une distinction entre les biens non-durables et les biens durables. Les biens durables sont des biens de consommation qui peuvent rendre des services au ménage sur une longue période de temps, généralement plus d'un an. Les biens durables les plus courants sont les moyens de locomotion (voiture, motocyclette, vélo, etc.),

les appareils électro-ménagers (téléviseur, réfrigérateur, congélateur, four, lave-linge, lave-vaisselle, climatiseur, chaîne de musique, radio, ventilateurs, etc.), les gros meubles (salon, table à manger et chaises, bibliothèque, autres armoires, etc.), et les appareils électroniques et d'autres biens (ordinateur, téléphone, téléphone portable, appareils photos, instruments de musique comme la guitare ou le piano, appareils motorisés de jardinage, bijoux et montres de valeur, tapis, etc.). Les biens qui n'entrent pas dans ces différentes catégories sont les biens non-durables.

L'agrégation des biens non-durables (par exemple les allumettes, le pétrole lampant, les vêtements, le carburant) et des services consommés par le ménage (par exemple le loyer, l'électricité, le nettoyage des vêtements) est directe. La valeur d'acquisition du bien ou du service est annualisée et incluse dans l'agrégat. Avant de traiter des biens durables et du logement, il convient de discuter des questions des dépenses d'éducation, des dépenses de santé et des dépenses des fêtes et cérémonies.

L'éducation est un investissement en capital humain et pour cette raison, certains analystes pensent que les dépenses d'éducation ne devraient pas faire partie de l'agrégat de consommation. De plus les dépenses d'éducation n'interviennent qu'à un moment donné du cycle de vie, et par conséquent elles ne sont pas bien adaptées pour une comparaison équitable du bien-être des ménages. Si on prend l'exemple de deux ménages ayant des revenus similaires, l'un avec des enfants d'âge scolaire et l'autre sans enfants d'âge scolaire, peut-être des retraités. Le ménage ayant des enfants d'âge scolaire est amené à supporter des dépenses d'éducation, celui composé de retraités épargne une partie importante de son revenu. On attribue un niveau de vie supérieur au premier ménage alors qu'il est contraint d'assurer l'éducation des enfants. Cela étant dans la pratique les dépenses d'éducation font presque toujours partie de l'agrégat de consommation et l'enquête harmonisée va souscrire à cette habitude.

Quant aux dépenses de santé, la principale raison avancée pour ne pas les inclure dans l'agrégat de consommation est qu'elles n'améliorent pas le niveau de vie du ménage, elles permettent juste de rattraper le niveau de vie que la personne malade a perdu du fait de cette situation. L'autre difficulté évoquée pour l'intégration des dépenses de santé dans l'agrégat de consommation est qu'elles sont sujettes à des erreurs de mesure importantes, notamment la prise en charge par les assurances qui n'est pas toujours bien cernée. En effet si l'on intègre les dépenses de santé, on devrait intégrer la totalité du coût des prestations de santé, pas seulement la part des dépenses supportées directement par le ménage assuré. Malgré ces difficultés, la pratique est de les inclure dans l'agrégat de consommation. De plus l'effort a été fait dans l'enquête harmonisée pour renseigner ces dépenses avec exactitude. Cela étant, du fait qu'ils sont proches des biens durables, les dépenses supportées pour les appareils thérapeutiques de toutes sortes, y compris les lunettes à usage médical, sont exclues de l'agrégat.

Le troisième cas particulier à examiner est celui des dépenses relatives aux fêtes et cérémonies. Lors de ces événements, il y a les dépenses alimentaires (alimentation et boissons) et des dépenses non alimentaires (habillement, chaussures et coiffure, location de salles et de chaises, etc.). On considère deux groupes : i) celui des cérémonies où les ménages organisent une réception (mariages, funérailles et deuils, baptêmes, communion et confirmation, etc.) ; ii) les autres fêtes qui sont pour l'essentiel des fêtes religieuses (Noël, Nouvel an, Pâques, fin de Ramadan, Tabaski, autres fêtes religieuses). Les dépenses des ménages des événements classés dans le premier groupe

sont d'emblée exclues de l'agrégat de consommation. La principale raison est qu'il s'agit de dépenses exceptionnelles ; en plus ces dépenses bénéficient plutôt à d'autres ménages et pas au ménage les ayant supportées. Du reste, ces dépenses sont plus proches d'un investissement social que d'une consommation. Quant aux événements du second groupe, il y a des raisons qui militent pour que cette consommation fasse partie de l'agrégat de bien-être et d'autres qui militent pour le contraire. La raison évidente pour intégrer les dépenses de consommation de Noël ou de la Tabaski dans l'agrégat de consommation est qu'au moins une partie de ces dépenses fait partie de la consommation du ménage. Mais on peut y opposer des arguments contraires. Premièrement il arrive assez souvent que le ménage consomme la nourriture préparée pendant plusieurs jours et dans ce cas, en prenant en compte la consommation des fêtes il y a double-compte avec la consommation alimentaire annualisée des sept derniers jours qui a déjà pris en compte les 365 jours de l'année. En fait même si on considère que le ménage consomme cette nourriture ce seul jour de fête, il ne faut plus annualiser la consommation des sept derniers jours par 365, il faudrait exclure les jours de fête. Deuxièmement, une partie de la nourriture des fêtes est donnée en cadeau à d'autres ménages et l'on peut constater des échanges de cadeaux en nourriture pendant certaines fêtes telles que la Tabaski. Mais on peut aussi argumenter que le ménage reçoit des cadeaux d'autres ménages et que l'un compense l'autre. Rien que pour la raison de double-emploi, il n'est pas indiqué d'introduire les dépenses d'alimentation de ces fêtes dans l'agrégat de consommation. De même les dépenses de location de salle et de chaises supportées lors de ces événements sont exclues de l'agrégat de consommation non-alimentaire. En fait les seules dépenses qu'on intègre dans l'agrégat de consommation non-alimentaire sont les dépenses d'habillement et chaussures de ce second groupe. La raison de ce choix réside dans le fait que les vêtements acquis pendant ces fêtes sont une vraie consommation du ménage.

S'agissant maintenant des biens durables⁵, ils sont consommés sur une période longue, plusieurs années. La période de référence pour la mesure de la pauvreté est d'un an ; par conséquent, la valeur d'acquisition qui reflète plutôt la valeur du bien durant toute sa durée de vie, n'est pas la variable appropriée pour la mesure de la pauvreté. D'un point de vue théorique et pratique, un bien n'est durable n'est pas consommé quand il est acquis, le ménage en consomme une fraction chaque année. Cette fraction consommée est *la valeur d'usage* et n'étant pas directement observable, elle doit être estimée à partir des données de l'enquête. Pour ce faire l'on procède en deux temps. D'abord sont exclus de l'agrégat de consommation les biens durables achetés dans l'année. En effet la valeur d'acquisition de ces biens figure généralement dans les enquêtes, et particulièrement dans l'enquête harmonisée. La raison en est que l'enquête a d'autres objectifs que l'évaluation de la pauvreté, par exemple fournir les données nécessaires pour l'élaboration des comptes nationaux pour laquelle ces données sont pertinentes. Ensuite le module portant sur le stock de biens durables sert au calcul de la valeur d'usage.

Pour le calcul les notations qui suivent sont utilisées. Pour une date (année) t , on appelle p_t^k la valeur actuelle d'un bien durable d'un type k ($k=1, \dots, m$) acheté à la date 0. On suppose l'existence d'un marché des capitaux où l'on peut prêter et emprunter de l'argent à un taux d'intérêt r . La valeur d'usage du bien à la fin de la période t est la différence entre la valeur actualisée du bien en début de période et sa valeur en fin de période (ou début de la prochaine période). Si vu_t^k est la valeur d'usage du bien k à la date t , on peut donc écrire :

⁵ Voir par exemple Amendola et Vecchi (2014).

$$vu_t^k = (1 + r)p_t^k - p_{t+1}^k$$

La valeur du bien en début de période traduit le coût d'opportunité du fait que le ménage pouvait vendre le bien à sa valeur en début de période et placer cet argent sur le marché des capitaux. Il y a aussi une relation entre la valeur du bien entre les dates 0 et 1 (début et fin de la période 1), elle est donnée par la formule suivante :

$$p_1^k = p_0^k(1 + \pi)(1 - \partial^k)$$

Où π est le taux nominal d'inflation et ∂^k le taux de dépréciation du bien k. Cette formule exprime le fait qu'à la date 1, le ménage a consommé une fraction du bien qui est mesurée par le taux de dépréciation. Mais le bien neuf ne vaut plus son prix nominal à la fin de la période, il vaut ce prix majoré de l'inflation. De manière générale, on peut écrire :

$$p_t^k = p_0^k(1 + r)^t(1 - \partial^k)^t$$

On peut estimer la valeur d'usage vu du bien k à la date t par :

$$vu_t^k = (1 + r)p_t^k - (1 + \pi)(1 - \partial^k)p_t^k \approx p_t^k(r - \pi + \partial^k)$$

Pour obtenir la valeur d'usage, il faut avoir la valeur actuelle du bien à la date t ainsi que le taux de dépréciation. Si on reprend la formule ci-dessus de la valeur du bien à la date t en fonction de sa valeur à la date zéro d'acquisition, on peut calculer la valeur d'usage par la formule :

$$\partial^k - \pi = 1 - \left(\frac{p_t^k}{p_0^k}\right)^{\frac{1}{t}}$$

Ainsi disposant de la valeur actuelle du bien, de la valeur d'acquisition et de l'âge, on a les éléments pour calculer le taux de dépréciation et en déduire la valeur d'usage. Dans une enquête, la valeur v_t^k de tous les biens de type k est difficile à obtenir car les biens sont d'un âge différent. Par exemple si un ménage a deux téléviseurs, deux motos, on demanderait la valeur de chacun de ces biens, cela rendrait encore plus difficile la collecte des données. En général, on dispose de la valeur d'acquisition et de la valeur actuelle du dernier bien, et du nombre de biens. Si on appelle s_t^k le nombre de biens durables d'un type k, p_0^k , le prix d'acquisition du dernier bien, on estime la valeur par le produit $s_t^k p_0^k$.

La dernière composante de l'agrégat de consommation est le loyer imputé des ménages propriétaires et ceux logés gratuitement. Le bien durable par excellence dont dispose les ménages est le logement. Une maison construite va durer plusieurs dizaines d'années et à l'instar des biens durables, le ménage n'en consomme que l'usage qu'il en fait. Cependant il est difficile de mettre en œuvre la même approche que celle des biens durables ; elle se heurte à plusieurs difficultés, notamment celle de la valeur du logement. Pour cette raison, une méthode alternative est utilisée. Etant donné qu'il y a des personnes qui louent leur appartement, l'on estime une fonction hédonique de logement pour les locataires, et cette fonction sert à imputer un loyer fictif aux ménages propriétaires et ceux qui sont logés gratuitement (accessoirement la même procédure est

utilisée pour imputer le loyer des ménages locataires pour lesquels le loyer est non-déclaré). Si L est le loyer supporté par les ménages en locations, X les caractéristiques du logement et de l'environnement du ménage, U le terme d'erreur, le modèle estimé est le suivant :

$$\ln L_h = X_h \beta + U_h$$

Les variables pouvant être incluses dans X sont les suivantes : nombre de pièces, nature des murs, nature du toit, nature du sol, nature des toilettes, électricité, eau courante, évacuation des ordures ménagères, évacuation des eaux usées. On pourrait estimer séparément ce modèle pour le milieu urbain et le milieu rural et techniquement cela devrait donner de meilleurs résultats. La difficulté que pose cette approche est l'étroitesse du marché de location en milieu rural, beaucoup de ménages ruraux sont propriétaires et le nombre de ménages locataires est généralement faible dans les échantillons d'enquête. On préfère donc estimer un modèle au niveau national en y incorporant des variables telles que le milieu de résidence, la région, etc. qui permettent de prendre en compte les variables non observables. On va aussi mobiliser des variables de l'environnement du ménage du questionnaire communautaire telles que les distances au réseau électrique, au réseau d'eau courante, etc.

3.1.3. Normalisation par la composition des ménages

Comme décrit ci-dessus, la première étape pour construire un indicateur de mesure du bien-être a consisté à calculer l'agrégat de consommation qui est obtenu comme la somme de la consommation alimentaire, la consommation non-alimentaire de biens non-durables et services, la valeur d'usage des biens durables et le loyer imputé des ménages propriétaires et ceux logés gratuitement. La deuxième étape consiste à tenir compte des différences dans la taille et de la composition démographique des ménages.

D'après la théorie, pour chaque ménage il faut calculer une échelle d'équivalence qui tienne compte de la taille et de la composition des ménages. L'échelle d'équivalence se justifie par une réalité simple : i) les besoins des personnes sont différents selon l'âge, le genre et même le type d'activité dans laquelle ces personnes sont engagées (par exemple les besoins en nourriture d'un jeune enfant sont différents de ceux d'un adulte) ; ii) il y a des économies d'échelle dans un ménage (par exemple un ménage de trois personnes n'a pas nécessairement besoin de trois fois plus de téléviseurs qu'un ménage d'une personne). L'échelle d'équivalence est donc une norme permettant de rendre les ménages comparables. Pour déterminer une échelle d'équivalence d'un ménage h , on considère d'abord un ménage de référence, par exemple celui d'un homme adulte vivant seul. Pour le ménage h donné, l'échelle d'équivalence est la somme du nombre d'équivalent-adultes par rapport au ménage de référence, ajusté par un facteur d'économies d'échelle. Quand l'échelle d'équivalence est calculée pour chaque ménage, on divise l'agrégat de consommation du ménage (calculée ci-dessus) par cette échelle. L'alternative à une échelle d'équivalence est la taille du ménage. Chaque approche présente des avantages et des inconvénients.

L'avantage de l'échelle d'équivalence est son assise théorique. Cependant évaluer les échelles d'équivalence pose des difficultés en économétrie appliquée. En fait il n'y a pas d'approche économétrique robuste permettant d'estimer les échelles d'équivalence de manière fiable. Face à cette difficulté, l'habitude est d'utiliser une échelle d'équivalence qui existe. L'une des plus

utilisée est l'échelle de la FAO, adaptée au cas des pays en développement, elle est donnée par le tableau 2 ci-dessous.

Tableau 2. Coefficient par sexe et âge pour évaluer l'échelle d'équivalence

| | Homme | Femme |
|----------------|-------|-------|
| Moins d'un an | 0.255 | 0.255 |
| 1 à 3 ans | 0.450 | 0.450 |
| 4 à 6 ans | 0.620 | 0.620 |
| 7 à 10 ans | 0.690 | 0.690 |
| 11 à 14 ans | 0.860 | 0.760 |
| 15 à 18 ans | 1.030 | 0.760 |
| 19 à 50 ans | 1.000 | 0.760 |
| 51 ans et plus | 0.790 | 0.660 |

On peut aussi utiliser une approche plus générale. Si on appelle A le nombre d'adultes dans un ménage (individus de 18 ans au moins), et E le nombre d'enfants (individus de moins de 18 ans), une formule plus générale pour calculer une échelle d'équivalence adaptée est donné par :

$$AE = (1 + \alpha(A - 1) + \beta E)^\theta$$

Pour avoir l'échelle d'équivalence, il faut estimer les trois paramètres α , β et θ . Les deux premiers expriment respectivement les besoins d'un adulte supplémentaire dans le ménage et ceux d'un enfant, par rapport à ceux du premier adulte. Quant au troisième paramètre, il prend en compte les économies d'échelle. Tous les trois paramètres ont des valeurs comprises entre 0 et 1. Si on prend le cas des pays à revenus faibles et intermédiaires, le paramètre α devrait être proche de 1, par contre la paramètre β devrait être plus proche de zéro ; en effet les besoins des enfants sont moins grands dans des pays en développement, du fait par exemple du coût de l'éducation qui s'étale sur une durée plus courte (faible espérance de vie scolaire dans ces pays). Quant au paramètre θ , il devrait aussi être proche de 1 ; la raison en est que l'alimentation a une grande part du budget et est sujette à de faibles économies d'échelle, les *biens publics* des ménages (appareils électroménagers, voiture, instruments de musique, etc.) qui sont sujets à plus d'économies d'échelle ont une part relativement peu élevée dans le budget des ménages.

Si à défaut d'une échelle d'équivalence on utilise plutôt la taille du ménage, l'avantage est clairement dans la facilité de la mise en œuvre. L'autre avantage est l'interprétation facile de l'indicateur de bien-être, la consommation par tête du ménage est compréhensible pour toute personne. Le principal inconvénient est qu'on gomme les différences dans la composition des ménages. En le faisant, on surestime le niveau de pauvreté des ménages de grande taille, cela affecte le ciblage des programmes de lutte contre la pauvreté.

Face à la difficulté d'estimer une échelle d'équivalence, on retient la taille du ménage malgré les inconvénients dus à son utilisation. Ainsi l'agrégat de consommation calculé au niveau du ménage est normalisé en le divisant par la taille du ménage. Cependant, il sera procédé à tests de sensibilité au profil de pauvreté ; notamment en remplaçant la taille par une échelle d'équivalence.

3.1. 4. Normalisation par les différences du coût de la vie

Les données ont été collectées pendant 6 mois sur une période de 9 mois, trois mois pour la première vague, trois mois de repos et trois mois pour la seconde vague. Les travaux de collecte couvrent donc 9 mois. Les différences de coût de la vie entre les ménages se situent à deux niveaux : i) les ménages d'une même localité interviewés à des périodes différentes font face à des prix différents ; ii) les ménages vivant dans des localités différentes éloignées les unes des autres supportent des prix différents du fait des coûts de transaction des lieux de production (ou d'importation) aux lieux de consommation.

Dans les travaux d'analyse de la pauvreté, ces deux facteurs sont pris en compte en calculant deux indices, un indice temporel et un indice spatial des prix. Le calcul de ces indices à partir des données d'enquête n'est pas sans difficulté, en particulier pour les biens non-alimentaires pour lesquels il est très difficile de renseigner les prix de manière adéquate du fait du grand nombre de produits, du grand nombre de variétés d'un même produit, des questions de différences de qualité, etc. On peut surmonter cette difficulté en utilisant l'indice national des prix à la consommation finale (IPC) des ménages comme déflateur temporel. Mais le panier de l'IPC est souvent fort éloigné du panier d'une enquête pauvreté, le premier donne souvent un gros poids à la capitale. On peut aussi, si les prix des produits non-alimentaires sont quasi-stables pendant la période couverte par l'enquête, procéder à la seule déflation temporelle du sous-agrégat de consommation alimentaire. Pour l'aspect spatial, on peut trouver une solution idoine, on revient sur cette question.

L'approche la plus solide consiste à calculer un seul indice, qui prend en compte aussi bien les aspects temporel et spatial ; cette approche est utilisée. Mais au lieu de l'approche classique avec des indices des prix, la normalisation sera faite avec les seuils de pauvreté et cela demande une explication. Un seuil de pauvreté représente le coût minimum de biens et services nécessaires pour qu'un ménage puisse disposer du minimum vital. Si on considère deux localités A et B, la différence entre les seuils de pauvreté doit représenter la différence du coût de la vie entre ces deux localités (tout au moins pour les ménages se situant autour du seuil de pauvreté), ce qui est requis d'un déflateur spatial ; le même résultat est valable entre une date t_1 , et une date t_2 . L'on combine les deux aspects en calculant un seuil de pauvreté référence et des seuils de pauvreté spatio-temporels ; le rapport entre chaque seuil de pauvreté spatio-temporel et le seuil de pauvreté de référence fournit le déflateur approprié, pour les ménages d'une localité donnée qui ont été enquêtés à un moment donné. L'on précise aussi qu'entre l'indice de Paasche et de Laspeyres, le dernier est retenu.

Sur le plan temporel on va considérer trois cas : i) l'évolution des prix est proche de zéro pendant les neuf mois entre le début et la fin de la collecte ; ii) l'évolution des prix est faible pendant les trois mois de chaque vague, mais il y a une certaine inflation entre la vague 1 et la vague 2 ; iii) pour chaque vague il y a une certaine inflation pendant les trois mois de collecte. Si un pays est dans le premier cas, tous les ménages sont dans un groupe temporel unique ; pour le deuxième cas on distingue deux groupes, un pour chaque vague ; pour le dernier cas on crée six groupes, un pour chaque mois de collecte. Les évolutions récentes des prix à la consommation montrent que la majorité des Etats membres de l'UEMOA devraient être dans la première ou la deuxième catégorie. Sur le plan spatial, les prix varient généralement d'une région à l'autre, entre les zones

agroécologiques, entre les milieux de résidence. La solution la plus simple est de travailler par région. Mais il y a un grand nombre de régions, et si l'on croise la région et le milieu de résidence, cela ferait un trop grand nombre de domaines, et l'échantillon de chaque domaine/temps serait trop faible pour une estimation robuste des seuils de pauvreté. On va plutôt travailler par zone agroécologique ; chaque pays sera divisé en un maximum de 5 zones agroécologiques. Donc le nombre maximum de domaines spatio-temporels est de 5 zones agroécologiques fois 2 milieux de résidence fois 2 groupes temporels soit 20 domaines spatio-temporels. Si i désigne la zone agroécologique, (i variant par exemple de 1 à 5), j désigne le milieu de résidence, (j variant de 1 à 2), k désigne le groupe temporel, (k variant de 1 à 2), on construit un seuil de pauvreté pour chacun de ces domaines et un seuil de pauvreté national qui est le seuil de pauvreté de référence. Formellement :

$Z_{i,j,k}$ et Z_{ref} désignent respectivement les seuils de pauvreté du domaine i, j, k et de référence, le déflateur D pour le domaine i, j, k est :

$$D_{i,j,k} = Z_{i,j,k}/Z_{ref}$$

En définitive, l'indicateur de mesure de bien-être retenu dans le cadre de cette étude est un agrégat de consommation finale par tête, après avoir procédé à une normalisation par l'indice du coût de la vie.

3.2. Application dans le cadre de l'enquête harmonisée

3.2.1. La consommation alimentaire

La consommation alimentaire est principalement renseignée dans les sections 7A et 7B du questionnaire ménage. La section 7A renferme les informations sur les repas pris hors du domicile et la section 7B traite de la consommation alimentaire dans le ménage. Dans ces deux sections, la consommation est mesurée pour les 7 derniers jours précédents le passage de l'agent enquêteur dans le ménage.

Il y a d'autres informations portant sur la consommation alimentaire du ménage dans deux autres sections. La section 9A traite des dépenses relatives aux fêtes et cérémonies au cours des 12 derniers mois, dépenses qui incluent l'alimentation. La section 9C est relative aux dépenses des 30 derniers jours, principalement les dépenses non-alimentaires. Mais la consommation de whisky et liqueurs, et des vins modernes a été mise dans cette section ; en effet il a été estimé lors de la conception du questionnaire qu'il était plus indiqué d'avoir une période plus longue pour la consommation de ces produits.

La construction de l'agrégat de consommation commence par la section 7B. La consommation alimentaire est renseignée en quantité, quantité elle-même renseignée en une unité non-standard (UNS). Il y a trois variables de quantité, la quantité totale consommée dans le ménage, la quantité qui provient de la production propre du ménage et celle qui provient d'autres sources que les deux précédentes (cadeau, prélèvement de son propre commerce, troc, etc.). La consommation alimentaire est calculée séparément en isolant celle provenant des achats, celle provenant de la production propre du ménage (autoconsommation) et celle provenant d'autres sources. Cette

approche présente un double avantage : i) il y a possibilité de valider les résultats en comparant à d'autres sources (par exemple sur l'autoconsommation) ; ii) l'information détaillée peut être utilisée à des fins d'analyse sans avoir besoin de recourir de nouveau à construire l'agrégat (il y a des cas par exemple où l'on s'intéresse seulement à la consommation monétaire).

Le fait que la consommation soit renseignée en quantité nécessite une information sur les prix pour la valoriser. Les prix ont été renseignés à deux sources. D'abord dans la section 7B du questionnaire ménage, les prix sont collectés à deux conditions : i) que le ménage ait acheté au moins une partie du produit consommé au cours des 7 derniers jours ; ii) que cet achat ait eu lieu dans les 30 derniers jours précédant le passage de l'agent enquêteur dans le ménage. Autrement dit on ne dispose pas d'information sur les prix dans le questionnaire pour les ménages dont l'achat a eu lieu au-delà d'un mois, on ne dispose pas non plus d'information sur les prix dans le questionnaire pour les ménages dont la totalité de la consommation provient de sources non-monétaires. Ensuite une seconde information sur les prix a été collectée sur les marchés des localités d'enquête.

Les notations suivantes sont utilisées :

- La variable $s07bq03a$ est la quantité totale consommée mesurée en UNS ;
- La variable $s07bq04$ est la quantité consommée provenant de la production propre du ménage, elle est mesurée dans la même UNS que $s07bq03a$;
- La variable $s07bq05$ est la quantité consommée provenant d'autres sources, elle est mesurée dans la même UNS que $s07bq03a$;
- La variable $s07bq07a$ est la quantité du produit acheté, si cette variable est renseignée (voir les deux conditions ci-dessus) ;
- La variable $s07bq08$ est la valeur du produit acheté, si cette variable est renseignée (voir les deux conditions ci-dessus) ;
- La variable $fconv03a$ est le facteur de conversion des UNS en unité standard (UST) pour l'unité utilisée pour mesurer les variables $s07bq03a$, $s07bq04$ et $s07bq05$ ⁶ ;
- La variable $fconv07a$ est le facteur de conversion des UNS en unité standard (UST) pour l'unité utilisée pour mesurer les variables $s07bq07$.

Avec ces notations, pour un ménage le prix unitaire d'acquisition pa en UST est donné par la formule suivante :

$$pa = s07bq08 / (s07bq07a \times fconv07a)$$

Pour les ménages d'une grappe, on calcule aussi pag , le prix médian d'acquisition en UST calculé à partir des ménages ayant acheté le produit.

Il peut arriver que dans une grappe, la consommation d'un produit provienne exclusivement de la production propre des ménages ; dans ce cas on ne peut avoir aucune des variables pa et pag . On fait alors recours aux prix relevés sur les marchés. On appelle pm le prix en UST du produit relevé sur le marché. A partir de la variable pm , on calcule pms le prix médian du produit dans la strate, pmr le prix médian du produit dans la région et pmn le prix médian du produit au niveau national.

⁶ Les facteurs de conversions proviennent de l'enquête sur les UNS menée avant l'enquête principale.

Avec ces données, la valeur de la consommation alimentaire sur les 7 derniers jours par les formules suivantes :

- Achat : $dep = (s07bq03a - s07bq04 - s07bq05) \times fconv03a \times pa$
- Autoconsommation : $dep = s07bq04 \times fconv03a \times pa$
- Cadeau et autres : $dep = s07bq05 \times fconv03a \times pa$

Ce calcul n'est possible que si la variable pa est disponible pour le ménage. Si elle ne l'est pas on utilise pag , si pag n'est pas disponible, on utilise pm , si pm ne l'est pas, on utilise pms , si pms ne l'est pas, on utilise pmr et si pmr ne l'est pas, on utilise pmn .

La consommation doit être annualisée, on obtient la valeur annuelle de la consommation par la formule $depan = dep \times 365/7$. Cette variable est calculée par ménage, produit et mode d'acquisition ; ces trois variables sont conservées dans le fichier. Ainsi le fichier créé a la structure décrite à la section 2.3.

Le traitement des repas pris hors ménage est moins compliqué, puisque les valeurs sont directement renseignées dans le questionnaire. Les repas pris hors ménage et acheté par le ménage lui-même sont renseignées dans les variables $s07aq02$, $s07aq05$, $s07aq08$, $s07aq11$, $s07aq14$, $s07aq17$ et $s07aq20$. Les repas pris hors ménage et reçu en cadeau sont renseignées dans les variables $s07aq03$, $s07aq06$, $s07aq09$, $s07aq12$, $s07aq15$, $s07aq18$ et $s07aq21$. Dans chaque cas, ces sept variables renseignent respectivement sur le petit-déjeuner, le déjeuner, le dîner, une collation, une boisson chaude, une boisson non-alcoolisée et une boisson alcoolisée. On annualise indépendamment la consommation des repas pris hors ménage et acheté par le ménage et les repas pris hors ménage et reçu en cadeau selon les formules suivantes :

- Achat : $depan = (s07aq02 + s07aq05 + s07aq08 + s07aq11 + s07aq14 + s07aq17 + s07aq20) \times 365/7$
- Cadeau : $depan = (s07aq03 + s07aq06 + s07aq09 + s07aq12 + s07aq15 + s07aq18 + s07aq21) \times 365/7$

Après avoir créé cette variable, pour rester dans la logique du fichier ménage/produit, on génère un code produit (cette section du questionnaire n'en a pas). On attribuera le code 151 aux repas pris hors ménage (ce code est à la suite de ceux des produits alimentaires).

Comme il a été signalé ci-dessus, deux items de consommation alimentaire (whisky et vins modernes) sont renseignés à la section 9C qui est relative aux dépenses des 30 derniers jours. Pour ces deux items la consommation est renseignée en valeur. Il s'agit d'acquisition sous forme d'achat et la formule suivante est utilisée :

- Achat : $depan = s09cq03 * 12$

A ce niveau on a l'ensemble de la consommation alimentaire devant impérativement faire partie de l'agrégat de consommation. Néanmoins il y a les dépenses alimentaires occasionnées par les fêtes et cérémonies et qui sont renseignées à la section 9A. Comme il a été discuté dans la section précédente, ces dépenses ne sont a priori pas intégrées dans l'agrégat de consommation. On les calcule néanmoins pour des tests de sensibilité. La formule suivante est utilisée :

- Achat : $depan = (s09aq03 + s09aq04)$ si $s09aq02 = 1, \dots, 8$.

Comme cette rubrique de consommation n'a pas de code attribué, on attribue le code 152.

Il est important de noter qu'une information auxiliaire sur l'autoconsommation a été renseignée dans les sections 16C (agriculture), 17 (élevage) et 18 (pêche). L'objectif premier de ces modules du questionnaire ménage est l'analyse des sources de revenu. Il se pose néanmoins la question de savoir si cette information peut être utilisée pour la construction de l'agrégat de consommation alimentaire. La réponse simple est non, pour un certain nombre de raisons : i) la période de référence de l'autoconsommation des produits de ces trois sections n'est pas systématiquement bien spécifiée de manière à produire une annualisation uniforme de cette consommation (par exemple pour les produits agricoles, la période de référence est la campagne agricole qui peut débiter à des périodes différentes dans différentes régions du pays) ; ii) les items de consommation de ces trois sections n'embrassent pas systématiquement l'ensemble des produits autoconsommés (pour l'agriculture, il manque les produits de l'arboriculture, pour l'élevage, il manque le lait et les produits laitiers et les questionnaires ne traitent pas de l'autoconsommation de boissons locales) ; iii) les UNS de ces modules du questionnaire pourraient dans certains cas ne pas avoir de facteurs de conversion en UST. Cependant compte tenu des erreurs de collecte qu'il pourrait y avoir dans le module consommation, on ne peut pas a priori ignorer totalement cette information. On calcule la valeur annuelle de l'autoconsommation de ces produits et leur attribuer le code 6 (autoconsommation alternative) pour la variable mode d'acquisition. Ainsi il est possible de comparer les deux distributions (autoconsommation d'origine et alternative) et comprendre laquelle est la mieux renseignée.

L'autoconsommation alternative est annualisée selon les formules ci-dessus :

- Produits agricoles : $depan = s16cq13a \times fconv \times pm \times nmois$ où ces variables représentent respectivement la quantité consommée, le facteur de conversion à tirer de la base de données des UNS, le prix relevé sur le marché et le nombre de mois depuis le début de la campagne (normalement 12 puisque la campagne 2017/18 qui est celle de l'enquête est achevée).
- Viande : $depan = s17q18 \times fconv \times pm$ où ces variables représentent respectivement la quantité consommée, le facteur de conversion à tirer de la base de données des UNS, le prix relevé sur le marché. Pour rester dans la logique du traitement des événements des 12 derniers mois, la consommation de viande provenant de baptêmes et autres cérémonies est exclue de ce calcul.
- Œufs : $depan = s17q48 \times s17q43 \times fconv \times pm$ où ces variables représentent respectivement la quantité consommée en moyenne par mois, le nombre de mois de consommation, le facteur de conversion à tirer de la base de données des UNS, le prix relevé sur le marché.
- Poissons $depan = s18q14a \times fconv \times pm \times nmois1$ pour la saison haute, et pour la saison basse $depan = s18q20a \times fconv \times pm \times nmois2$. $nmois1$ et $nmois2$ représentent respectivement le nombre mois de saison haute et le nombre de mois de saison basse ; les autres variables ont une signification évidente.

Comme dans le cas de la consommation alimentaire, l'agrégat est construit pour chaque item de consommation. Mais une fois de plus, cette consommation n'est pas introduite directement dans l'agrégat de consommation.

3.2.2. La consommation non-alimentaire de biens non-durables et des services

Pour les produits non-alimentaires, on considère que le produit est consommé quand il est acheté, sauf dans le cas des biens durables. Ainsi la consommation est égale à la dépense et cette dernière est enregistrée. Les dépenses en biens non-alimentaires sont enregistrées dans les sections 9B pour ceux renseignés au cours des 7 derniers jours, 9C (sauf les produits portant le code 301 et 302 correspondant respectivement au whisky et vins modernes) pour ceux renseignés au cours des 30 derniers jours, 9D pour ceux renseignés au cours des 3 derniers mois, 9E pour ceux renseignés au cours des 6 derniers mois et 9F pour ceux renseignés au cours des 12 derniers mois. Les dépenses de loyer, eau, électricité, etc. sont renseignées à la section 11 sur le logement. Les dépenses de service de téléphonie mobile sont consignées à la section 1. Enfin, les dépenses d'éducation et de santé renseignées au niveau individuel respectivement dans les sections 2 et 3.

Pour chacune des sections 9x, où x représente la section (B, C, D, E, F), la consommation annuelle du produit est la suivante :

- Achat : $depan = s09xq03 \times freq$

Pour x égal à B, C, D, E, F, *freq* prend respectivement les valeurs 365, 12, 4, 2 et 1, correspondant à la fréquence à laquelle la dépense est mesurée.

Il y a deux situations à prendre en compte, les double-emplois et les dépenses en biens durables et d'investissement en logement.

- A la section 9C, le whisky et les vins modernes ont été comptabilisés avec les dépenses alimentaires, il faut donc exclure ces produits (codes 301 et 302).
- A la section 9F, en plus des biens de consommation courante il y a des dépenses d'investissement en logement et des dépenses en acquisition de biens durables. Les dépenses de petites réparations de logement sont considérées comme de la consommation, mais les dépenses de grosses réparations et les dépenses de construction sont de l'investissement. De même les dépenses en biens durables sont à exclure, puisque c'est la valeur d'usage qui est assimilée à de la consommation. Ainsi à la section 9F, on exclut les produits de codes 603 à 607, ainsi que 610 et 611, ces dépenses correspondent à un investissement en logement (acquisition ou grosses réparations). On exclut aussi les produits de codes 612 et 613, 616 et 617, 623 et 624, 634 et 635, les dépenses correspondantes sont relatives aux biens durables.

Les dépenses de logement ne sont pas renseignées à la section 9, mais plutôt à la section 11. On y trouve les dépenses suivantes : loyer mensuel des ménages locataires (s11q05) ; facture d'eau (s11q24) selon une périodicité hebdomadaire, mensuelle, tous les 2 mois, trimestrielle ; achat d'eau mensuel auprès des revendeurs (s11q26) ; facture d'électricité (s11q37) selon une périodicité hebdomadaire, mensuelle, tous les 2 mois, trimestrielle ; facture de téléphone fixe ou carte prépayée (s11q45) selon une périodicité hebdomadaire, mensuelle, tous les 2 mois, trimestrielle ; facture d'abonnement à l'internet ou recharge de la clé internet (s11q48) selon une périodicité hebdomadaire, mensuelle, tous les 2 mois, trimestrielle ; facture d'abonnement au câble ou télévision par satellite (s11q52) selon une périodicité hebdomadaire, mensuelle, tous les 2 mois, trimestrielle. L'annualisation de ces différentes dépenses est faite de la manière suivante :

- Loyer des locataires : $depan = s11q05 \times 12$, on attribue le code 331
- Facture d'eau courante : $depan = s11q24 \times freq$, on attribue le code 332
- Achat d'eau auprès des revendeurs : $depan = s11q26 \times 12$, on attribue le code 333
- Facture d'électricité : $depan = s11q37 \times freq$, on attribue le code 334
- Facture de téléphone : $depan = s11q45 \times freq$, on attribue le code 335
- Facture d'abonnement à l'internet : $depan = s11q48 \times freq$, on attribue le code 336
- Facture d'abonnement au câble : $depan = s11q52 \times freq$, on attribue le code 337

Dans les formules ci-dessus, $freq$ prend les valeurs 52, 12, 6 et 4 respectivement pour une périodicité hebdomadaire, mensuelle, tous les deux mois et trimestrielle.

Les dépenses en cartes prépayées, y compris les transferts de crédit au cours des 7 derniers jours, sont renseignées à la section 1, variable s01q38. La dépense annuelle est calculée ainsi :

- Crédit de téléphonie mobile : $depan = s01q38 \times 365/7$, on attribue le code 338

Etant donné que les dépenses en carte de crédit de téléphonie mobile sont renseignées au niveau individuel, il faut faire la somme de cette dépense (s01q38) pour tous les individus qui en ont une avant de multiplier par 365 et diviser par 7.

Les dépenses d'éducation sont consignées à la section 2, les variables s02q20 à s02q27. Pour des raisons analytiques, l'agrégation des dépenses d'éducation doit se faire par type dépense et par niveau d'éducation.

Les trois types de dépenses à considérer sont les suivants :

- Frais de scolarité et cotisations obligatoires : $depan = s02q20 + s02q21$
- Livres, fournitures, transport, cantine : $depan = \sum_{i=22}^{26} s02qi$
- Cours de soutien : $depan = s02q27$

On considère ensuite quatre niveaux en utilisant la variable s02q14 : i) primaire et maternelle ; ii) secondaire 1 ; iii) secondaire 2 ; iv) post-secondaire et université.

On attribue les codes suivants aux dépenses d'éducation :

- Primaire : les codes 661 pour les frais de scolarité ; 662 pour les livres et fournitures ; et 663 pour les cours de soutien.
- Secondaire 1 : les codes 664 pour les frais de scolarité ; 665 pour les livres et fournitures ; et 666 pour les cours de soutien.
- Secondaire 2 : les codes 667 pour les frais de scolarité ; 668 pour les livres et fournitures ; et 669 pour les cours de soutien.
- Post-secondaire et supérieur : les codes 670 pour les frais de scolarité ; 671 pour les livres et fournitures ; et 672 pour les cours de soutien.

Les dépenses de santé sont consignées à la section 3, les variables s03q13 à s03q18 ; s03q24, s03q26, s03q27, s03q29 à s03q31. Comme pour l'éducation, l'agrégation des dépenses de santé se fait par type dépense. Les dépenses supportées pour les appareils thérapeutiques et les verres correcteur (s03q26 et s03q27), sont exclues de l'agrégat.

Les types de dépenses à considérer sont les suivants :

- Frais de consultation auprès d'un généraliste : $depan = s03q13 \times 4$, code 681
- Frais de consultation auprès d'un spécialiste : $depan = s03q14 \times 4$, code 682
- Frais de consultation auprès d'un dentiste : $depan = s03q15 \times 4$, code 683
- Frais de consultation auprès d'un guérisseur : $depan = s03q16 \times 4$, code 684
- Frais des examens médicaux : $depan = s03q17 \times 4$, code 685
- Frais de médicaments : $depan = s03q18 \times 4$, code 686
- Frais d'hospitalisation : $depan = s03q24 \times s03q20$, code 691
- Vaccinations et autres dépenses préventives : $depan = s03q29 + s03q30 + s03q31$, code 692

L'agrégation des dépenses de santé ne prend pas encore en compte l'éventuel problème de la prise en charge des dépenses de santé par les assurances et les mutuelles ; la question sera approfondie après discussion de la manière dont le module a été renseigné. Il convient de savoir précisément comment l'information a été collectée, notamment est-ce que la consigne a été donnée d'inscrire exactement ce que le ménage sort de sa poche ou s'il fallait renseigner le coût effectif du service reçu.

3.2.3. La valeur d'usage des biens durables

On estime la valeur d'usage des biens durables à l'aide de la méthodologie décrite à la section 3.1.2. Le stock des biens durables figure à la section 10. Il est important de souligner que tous les biens figurant dans cette section ne sont pas nécessairement des biens durables. Les biens portant les codes 44 et 45, qui sont les terrains et les immeubles sont exclus du calcul de l'agrégat de consommation. D'autres biens peuvent être avoir un double usage, tout en étant des biens pour l'usage du ménage, ils peuvent aussi être des biens de production, c'est-à-dire des biens pour un usage commercial. C'est par exemple le cas du fusil de chasse, mais cela peut être aussi la pirogue (pêche), le réfrigérateur ou le congélateur pour la congélation des glaces artisanales, etc. il faudra vérifier dans chaque cas qu'il s'agit de biens durables.

Les variables renseignées dans cette section sont le type du bien, le nombre de biens de chaque type, l'âge, le prix d'acquisition et la valeur actuelle du dernier bien acquis. La valeur d'usage se calcule en deux temps, d'abord la formule appropriée de la section 3.1.2 sert au calcul du taux de dépréciation. Pour chaque type de bien, on retient le taux de dépréciation médian, taux qui est alors appliqué à tous les ménages possédant ce bien. Le calcul du taux de dépréciation implique les seules observations pour lesquelles l'information est complète, âge, prix d'acquisition, prix au coût de remplacement. Pour le calcul, certains biens ont moins d'un an, c'est-à-dire zéro en année révolu. Etant donné que la formule n'admet pas d'âge zéro, on attribue 0.5 an à tous les biens dont l'âge est inférieur à un an. Quand le taux de dépréciation est calculé, on estime la valeur d'usage en utilisant la formule appropriée de la même section 3.1.2. Pour ce calcul, il est nécessaire d'avoir les variables suivantes : la valeur actuelle, le taux d'intérêt et le taux de dépréciation. Le taux d'intérêt est une variable exogène connue, le taux de dépréciation a été calculé. Quant à la valeur actuelle, on l'estime par le nombre de biens multiplié par la valeur au coût du remplacement du dernier. Pour les ménages n'ayant pas déclaré la valeur au coût du remplacement, on impute la valeur médiane des biens du type considéré à partir des ménages ayant une valeur valide.

3.2.4. Le loyer imputé des ménages propriétaires et ceux logés gratuitement

Le modèle pour estimer le loyer imputé des ménages propriétaires et ceux logés gratuitement a été exposé à la section 3.1.2. La variable expliquée du modèle est le logarithme du loyer des ménages locataires. Les variables explicatives sont le nombre de pièces, la nature des murs, la nature du toit, la nature du sol, la nature des toilettes, l'électricité, l'eau courante, le mode d'évacuation des ordures ménagères, le mode d'évacuation des eaux usées, le milieu de résidence et la région. A ces variables on ajoute cinq variables de l'environnement du ménage tiré du questionnaire communautaire. Il s'agit des variables s01q06 (principale voie pour accéder à ce village/quartier), s01q08 (moyens de transport en commun disponibles pour amener les gens dans ce village/quartier), s01q11 (existence d'un réseau de distribution électrique dans ce village/quartier), s01q12 (existence d'un réseau de distribution d'eau courante dans ce village/quartier), s01q13 (existence d'un réseau de téléphonie mobile dans ce village/quartier). A l'exception du nombre de pièces, les autres variables sont qualitatives ; chacune d'elles sera convertie en plusieurs variables dichotomiques, en tenant compte du poids des différentes modalités.

Le modèle est un modèle semi-logarithmique, la procédure la mieux adaptée en STATA pour la prédiction est non pas la commande *regress* classique, mais plutôt la commande *predlog*. La prédiction est alors obtenue avec *YHTSMPEAR*, qui a de meilleures propriétés.

Il est important de souligner que ce modèle est aussi utilisé pour imputer le loyer effectif des ménages locataires qui n'en auraient pas déclaré un.

3.2.5. Normalisation et indicateur de bien-être

A ce niveau on dispose d'un agrégat de consommation qu'on appelle Y . Pour obtenir l'indicateur de bien-être, on prend en compte les différences dans la composition des ménages. Comme expliqué ci-dessus, on considère la taille, qu'on appelle $hsize$, c'est le nombre d'individus qui vivent régulièrement dans le ménages (ceux qui y vivent depuis au moins 6 mois, ou depuis moins de 6 mois avec l'intention d'y passer au moins 6 mois).

On calcule aussi un déflateur pour chaque domaine spatio-temporel. Comme on l'a expliqué à la section 3.1.4, les déflateurs sont calculés à partir des seuils de pauvreté (on revient sur la construction des seuils de pauvreté à la section 4). On appelle $D_{i,j,k}$ le déflateur du domaine i, j, k (i désigne la zone agroécologique j le milieu de résidence et k désigne le groupe temporel). L'indicateur de bien-être d'un ménage appartenant au domaine spatio-temporel i, j, k , est alors :

$$pcexp = Y / (hsize \times D_{i,j,k})$$

4. CONSTRUCTION DU SEUIL DE PAUVRETE

Le seuil de pauvreté est conçu de manière à permettre aux personnes qui sont classées comme non-pauvres de satisfaire au minimum leurs besoins vitaux. La difficulté est de définir ces besoins minima (alimentaires et non alimentaires). L'approche utilisée pour construire le seuil de pauvreté

est celle du coût des besoins de base⁷. La construction d'un seuil de pauvreté passe par deux étapes, construire un seuil de pauvreté alimentaire, puis un seuil non-alimentaire et de faire la somme des deux.

Pour la première étape, il faut déterminer un panier de biens alimentaires permettant à chaque individu de couvrir K kilocalories par jour, K compris entre 2100 et 2400 kilocalories est une norme internationalement admise en matière de consommation alimentaire (la quantité exacte sera arrêtée plus tard). La valorisation de ce panier fournit le seuil alimentaire. Une norme pareille n'existe malheureusement pas pour les biens non alimentaires. En partant du constat que même les personnes les plus pauvres sont obligées de satisfaire un minimum de besoins non-alimentaires, deux solutions sont proposées. La première consiste à déterminer comme composante non alimentaire du seuil de pauvreté la part de la consommation non-alimentaire des ménages dont la consommation totale est égale au seuil de pauvreté. La seconde solution consiste à déterminer comme composante non alimentaire du seuil de pauvreté la part de la consommation non-alimentaire des ménages dont la consommation alimentaire est égale au seuil de pauvreté. La seconde solution donne évidemment un seuil plus élevé que la première.

Comme il a été expliqué à la section 3, on construit un seuil de pauvreté national, et on construit des seuils de pauvreté par domaine spatio-temporel qui vont servir de déflateurs.

Pour construire le seuil de pauvreté alimentaire national, il faut constituer un panier de biens alimentaires. Avant de constituer le panier, il faut rappeler que le choix a été fait de travailler avec des indices de Laspeyres. La conséquence sur le plan pratique est de retenir un panier de référence au niveau national. De plus ce panier doit être construit pour les ménages se situant au niveau du seuil de pauvreté. Dans la pratique, on travaille avec les ménages compris entre le troisième et le huitième décile, on ignore donc les ménages des deux déciles les plus pauvres et ceux des deux déciles les plus riches pour constituer le panier.

Pour cet exercice, l'agrégat de consommation alimentaire présenté à la section 3 est le principal input. Le panier est constitué des produits alimentaires les plus consommés totalisant de l'ordre de 80 pourcent de la consommation alimentaire totale, hors repas pris à l'extérieur⁸. Dans la pratique avant d'arrêter le panier définitif pour un pays, il faut vérifier que le panier reflète aussi la consommation des différentes zones agroécologiques. Cela se fait en vérifiant par exemple que le panier retenu représente aussi un minimum de 70 pourcent de la consommation de chaque zone agroécologique. Par exemple si un produit largement consommé dans une région est au-delà du panier initialement retenu, il faut élargir le panier.

Supposons que le panier final comprend n biens, $i=1, 2, \dots, n$. On appelle V_i la valeur de la consommation totale du bien i des déciles 3 à 8 (attention, il doit bien s'agir de valeurs extrapolées), POP la population totale des déciles 3 à 8, p_i le prix moyen national du bien i (au kilogramme), et c_i l'apport en calories de 100 grammes de bien i consommé. On calcule

⁷ Voir Ravallion (1998).

⁸ Les repas pris à l'extérieur ne peuvent pas être utilisés dans la construction du seuil de pauvreté. La raison en est que l'exercice requiert une correspondance entre les quantités consommées et leur apport en calories, cette information n'est pas disponible pour les repas pris à l'extérieur.

respectivement la valeur de la consommation moyenne par tête et par jour du bien i et la quantité moyenne par tête et par jour du bien i par :

$$v_i = \frac{V_i}{POP \times 365} \text{ et } q_i = \frac{v_i}{p_i} = \frac{V_i}{pop \times 365 \times p_i}$$

Ainsi l'apport quotidien en calories de la quantité consommée de bien i est donné par :

$$cq_i = q_i \times 10 \times c_i$$

Et l'apport quotidien en calories de la consommation totale du panier est donné par :

$$cq = \sum_{i=1}^n cq_i$$

Ainsi pour bénéficier des K kilocalories, on fait l'hypothèse que les produits non pris en compte ont le même apport en calories. Ainsi les quantités qu'il faudrait à un individu pour couvrir ses besoins alimentaires quotidiens sont données par : $qa_i = q_i \times K/cq$

Le seuil de pauvreté alimentaire est alors donné par : $z_{al} = 365 \times \sum_{i=1}^n p_i qa_i$

Pour pouvoir calculer le seuil non-alimentaire on va introduire trois autres variables. On va appeler $y_h = y_h^{al} + y_h^{na}$, la consommation totale par tête du ménage h , comme somme de la consommation alimentaire par tête et de la consommation non-alimentaire par tête. On considère S_1 le sous-ensemble de ménages dont la consommation totale par tête est égale au seuil alimentaire (dans la pratique, on retient les ménages dont la consommation totale par tête se situe à plus ou moins 5 pourcent du seuil de pauvreté alimentaire). On définit aussi S_2 le sous-ensemble de ménages dont la consommation alimentaire par tête est égale au seuil alimentaire, (de même on retient les ménages dont la consommation alimentaire par tête se situe à plus ou moins 5 pourcent du seuil de pauvreté alimentaire). On peut calculer deux seuils non-alimentaires différents (ω_h est une pondération appropriée pour ce calcul, notamment le poids du ménage multiplié par sa taille).

Le premier seuil est donné par : $Z_{na}^b = \sum_{h \in S_1} \omega_h y_h^{na}$

Le deuxième seuil est donné par : $Z_{na}^h = \sum_{h \in S_2} \omega_h y_h^{na}$

L'un des deux sera retenu comme non-alimentaire. Pour arrêter définitivement lequel, on tiendra compte d'un certain nombre de facteurs, par exemple l'écart avec les seuils précédemment en vigueur dans le pays. Si on appelle simplement le seuil retenu Z_{na} , le seuil de pauvreté au niveau national est donné par $Z_{ref} = Z_{al} + Z_{na}$

Pour la construction des seuils de pauvreté spatio-temporels, dès lors que l'indice de Laspeyres a été retenu, le panier de consommation du niveau national est le même pour tous les domaines spatio-temporels, seuls les prix vont varier d'un domaine à l'autre. Ainsi pour le seuil de pauvreté alimentaire de chaque domaine spatio-temporel, il suffit de prendre l'équation du calcul du seuil

alimentaire et de remplacer le vecteur des prix du niveau national par le vecteur des prix du domaine spatio-temporel considéré. Le seuil de pauvreté non-alimentaire de chaque domaine s'obtient aussi par l'approche décrite ci-dessus pour le seuil non-alimentaire du niveau national. En particulier si le seuil minimum a été retenu au niveau national, on retient aussi le seuil minimum pour chaque domaine ; et si le seuil maximum a été retenu au niveau national, on retient aussi le seuil maximum pour chaque domaine. Après ces calculs le seuil de pauvreté de chaque domaine est bien sûr obtenu comme la somme des seuils de pauvreté alimentaire et non-alimentaire.

Comme il a été expliqué à la section 3, le déflateur pour un ménage appartenant à un domaine donné est le rapport entre le seuil de pauvreté de ce domaine et le seuil de pauvreté national.

5. TESTS DE SENSIBILITE

Aussi bien la construction de l'indicateur de bien-être que le seuil de pauvreté se font avec un certain nombre d'hypothèses, certaines sont robustes, d'autres le sont moins. Il est donc indiqué de voir l'impact d'hypothèses alternatives sur les indicateurs d'inégalité, sur les indicateurs de pauvreté et sur le classement en matière de niveau de vie et de pauvreté.

Autoconsommation alternative

Outre l'autoconsommation renseignée dans la section de la consommation des 7 derniers jours, on trouve des informations sur l'autoconsommation dans les modules agriculture, élevage et pêche. Il se pose naturellement la question de savoir laquelle des deux sources retenir. Pour prendre la bonne décision, les deux distributions de l'autoconsommation seront comparées et en fonction du résultat, une décision sera arrêtée. Même si, comme c'est souvent le cas l'autoconsommation tirée du module consommation fournit des résultats plus robustes, on peut quand même faire appel à l'autoconsommation alternative pour améliorer l'autoconsommation normale dans des cas où il y aurait des erreurs de mesure évidentes sur la seconde.

Corrections de valeurs aberrantes

Comme il a été indiqué ci-dessus, la correction des valeurs aberrantes peut influencer sur l'inégalité, en particulier l'indice de Gini. Il est retenu de remplacer les valeurs supérieures à la moyenne plus trois écart-types par la médiane. On va calculer l'indice de Gini sous plusieurs hypothèses : i) en ignorant les corrections ; ii) avec les corrections retenues ; et iii) avec des corrections alternatives (par exemple une correction consistant plutôt à remplacer les valeurs trop grandes par le maximum des valeurs acceptables - le fractile 99 ou la borne supérieure de la moyenne plus trois écart-types). La variation de l'indice de Gini sous ces différentes hypothèses permettra de voir l'impact des corrections de prendre la bonne décision.

Prise en compte des différences dans la composition des ménages

La taille est retenue pour normaliser les différences dans la composition démographique des ménages. C'est une grande simplification dans la mesure où on ignore les différences de besoins entre les membres du ménage et les économies d'échelle. Si par exemple les économies d'échelle sont importantes, on surestime le niveau de pauvreté des ménages de grande taille. Mais plus grave

il peut arriver que le classement en termes de niveau de vie change. On va utiliser deux échelles d'équivalence pour tester l'hypothèse du classement en termes de niveau de vie, notamment l'échelle de la FAO et une échelle calculée à partir de la formule générale de la section 3.1.3 avec pour α, β et θ les valeurs respectives de 0.7, 0.5 et 0.9. L'on regardera principalement le classement par région, milieu de résidence et taille du ménage.

Prise en compte des différences du coût de la vie

L'indice de Laspeyres a été retenu pour normaliser les différences spatio-temporelles du coût de la vie. Une alternative est l'indice de Paasche. L'on va calculer les déflateurs avec l'indice de Paasche et mener le même exercice que ci-dessus, le classement des ménages sur l'échelle du niveau de vie par région, milieu de résidence et taille du ménage. On va aussi comparer les distributions de l'indicateur de bien-être des deux approches.

Pour le calcul de l'indice de Paasche, le calcul est proche de celui de Laspeyres, mais au lieu du panier de consommation du niveau national, l'on devra construire un panier de consommation pour chaque domaine spatio-temporel et utiliser ce panier pour construire le seuil de pauvreté.

Seuils de pauvreté

Il y a aussi un certain nombre d'hypothèses pour construire les seuils de pauvreté. D'abord le panier de produits alimentaires peut varier en fonction de la sous-population retenue comme référence, il a été retenu de travailler avec les déciles 3 à 8. La population de référence doit être celle des ménages vivant autour du seuil de pauvreté. Ainsi on peut aussi réduire cette population de référence aux déciles 3 à 6 par exemple, arguant que dans aucun des pays l'incidence de la pauvreté sera supérieure à 60 pourcent.

L'on sait aussi que l'approche du coût des besoins de base conduit à deux seuils de pauvreté, un seuil min et un seuil max. Ces hypothèses alternatives vont conduire à construire d'autres seuils de pauvreté et il faudra examiner l'impact sur le profil de pauvreté.

6. ANALYSE DE LA PAUVRETE

A compléter pour l'atelier 2, prévu après la fin de la collecte des données des deux vagues.

REFERENCES BIBLIOGRAPHIQUES

Amendola, N., and Giovanni Vecchi (2014). Durable Goods and Poverty Measurement, Policy Research Working Paper 7105, Washington, D.C., World Bank.

Deaton, A. S., and Salman Zaidi (2002). Guidelines for Constructing Consumption Aggregates for Welfare Analysis, LSMS Working Paper 135, Washington, D.C., World Bank.

Foster, J., Suman Seth, Michael Lokshin, Zurab Sajaia (2013). A Unified Approach to Measuring Poverty and Inequality, Theory and Practice, Washington, D.C., World Bank.

Institut de la Banque Mondiale, Cours d'Introduction à l'Analyse de la Pauvreté (2005). World Bank.

Ravallion, M, (2016). *The Economics of Poverty, History, Measurement and Policy*, Oxford University Press.

Ravallion, M, (1998). Poverty Lines in Theory and Practice, LSMS Working Paper 133, Washington, D.C., World Bank.

Ravallion, M, (1996). Comparaisons de la Pauvreté, Concepts et Méthodes, LSMS Document de Travail 122, Washington, D.C., World Bank.