

Mesure du secteur informel : Sensibilité aux non-réponses et validation d'une imputation probabiliste¹

par

Aude Vescovo, Philippe Bocquier, Constance Torelli

Version révisée, octobre 2007

Résumé

La connaissance du secteur informel est un enjeu primordial pour la politique économique. La mesure de sa part dans l'emploi urbain doit donc être réalisée avec le plus grand soin. Les enquêtes 1-2-3 du projet PARSTAT permettent de distinguer les unités de production informelles selon la définition établie par le 15^{ème} Congrès International des Statisticiens du Travail de 1993. Cependant, la non-réponse partielle peut biaiser la mesure des taux d'emplois dans le secteur informel. Le travail présenté ici a pour but de tester la sensibilité de ces taux au traitement de la non-réponse. Il est basé sur des méthodes d'imputation économétrique qui nécessitent une modélisation de la non-réponse et de la réponse renseignée. La répétition de 500 procédures d'imputation permet d'attester la robustesse des résultats obtenus. Les atouts majeurs de la méthode économétrique sont le gain de temps, la moins grande dépendance des résultats de l'imputation au statisticien qui l'opère, et la prise en compte d'effets toutes choses égales par ailleurs.

Ce document fait suite à une longue série de travaux sur les marchés de l'emploi menés par AFRISTAT et DIAL dans le cadre du projet PARSTAT. Il n'aurait pas pu être réalisé sans les contributions techniques des équipes tout au long du processus qui mène de la collecte à l'analyse. En ce qui concerne ce document, nous tenons à remercier Siriki Coulibaly, Sandrine Mesplé-Somps, Philippe de Vreyer, Sébastien Merceron, Eloi Ouédraogo, Christophe Nordman, Javier Herrera et François Roubaud pour leurs remarques et relectures.

I. Introduction

Les données d'enquêtes sont, de façon inévitable, soumises à la non-réponse. La non-réponse peut être complète dans le cas où un individu échantillonné n'est pas trouvé lors du passage de l'enquêteur, ou s'il refuse de répondre. La non-réponse partielle ne concerne que quelques questions pour lesquels l'information n'est pas, au final, disponible. La non-réponse partielle peut être due à l'enquêteur ou au personnel de terrain en général, comme à l'enquêté. Lorsqu'elle est due à l'enquêté, elle est motivée soit par la perception de la question comme étant indiscrete (revenus, impôts, opinion...), soit par la méconnaissance de la réponse (typiquement, choix de la modalité "Ne sait pas").

La non-réponse, qu'elle soit partielle ou totale, peut biaiser fortement les analyses et les estimateurs calculés. L'échantillon n'étant plus représentatif, les statistiques reflèteront le phénomène étudié pour la seule catégorie de la population pour laquelle une réponse est disponible. Si la non-réponse touche un grand nombre de questions pour un même individu, il est préférable d'écartier les individus concernés de l'analyse et de redresser l'échantillon en re-pondérant les individus ayant répondu. En revanche, lorsque la non-réponse est partielle, les méthodes d'imputations sont privilégiées. Dans cet article, nous appliquons et comparons certaines d'entre elles, dans le cadre de la mesure de la part du secteur informel dans le tissu productif urbain en Afrique de l'Ouest.

Le secteur privé formel rassemble toutes les entreprises ayant une relation formelle avec l'Etat (en lui rendant des comptes et en s'enregistrant auprès de lui) : par négation, le secteur informel renvoie à toutes les unités qui échappent à ce contrôle étatique. Il se compose d'un grand nombre de micro-entreprises et est souvent, comme en Afrique de l'Ouest, le secteur pourvoyant le plus d'emplois, mais à des travailleurs ne bénéficiant d'aucune protection sociale. La connaissance du secteur informel est un enjeu primordial pour les décideurs de politique économique que ce soit dans un souci de soutien public à ces entreprises et travailleurs, ou bien dans un souhait d'élargissement de la base de la taxation directe. Pour une présentation plus détaillée des concepts et de leurs implications statistiques et politiques, voir Henley, Reza Arabsheibani, Carneiro (2006).

Le 15ème Conférence Internationale des Statisticiens du Travail (1993) a défini une unité de production informelle (UPI) comme une entreprise non identifiée par un numéro statistique ou fiscal ou dans laquelle aucune comptabilité formelle n'est tenue (Backini-Yetna *et al.* 1999). Nous étudierons donc l'impact des non-réponses aux questions portant sur l'enregistrement statistique et fiscal (ESF) et sur la tenue d'une comptabilité formelle (CF). La mise en pratique de la définition du secteur informel (SI) pose problème dès lors que les données des questions d'ESF et de CF sont soumises au phénomène de non-réponse.

Les données utilisées (Tableau 1) proviennent des enquêtes en phases 1-2-3, menées dans le cadre du Programme d'Appui Statistique à la Surveillance Multilatérale (PARSTAT), entre 2001 et 2003 dans les capitales économiques de sept des huit pays de l'UEMOA (Bénin, Burkina Faso, Côte d'Ivoire, Mali, Niger, Sénégal et Togo), et dont un des objectifs principaux était de mesurer la part du secteur informel dans l'économie urbaine (Amegashie *et al.* 2005).

Dans cet article, nous identifions d'abord le type de mécanisme de non-réponse, proposons une méthode d'imputation robuste, et calculons ensuite des estimateurs non biaisés de la part du secteur informel dans sept capitales économiques d'Afrique de l'Ouest.

II. Mécanismes de non-réponse et modèle d'imputation

II.1. Définitions des mécanismes de non-réponses

La distribution des variables indicatrices de réponse est appelée mécanisme de non-réponse. Cette distribution n'est généralement pas connue, aussi sa forme fait l'objet d'hypothèses. Une description des difficultés engendrées par le fait de ne pas tenir compte des données manquantes, figure dans Little et Rubin (2002). Little et Rubin (2002) proposent une typologie des mécanismes de non-réponse. On parle

- de données manquant complètement aléatoirement ou complètement au hasard (MCAR pour "*missing completely at random*") lorsque la probabilité qu'une observation soit incomplète est une constante,
- de données manquant simplement aléatoirement (MAR pour "*missing at random*") lorsque cette probabilité ne dépend pas de la variable d'intérêt mais d'autres valeurs observées
- et de données manquant non aléatoirement (MNAR pour "*missing not at random*") lorsque cette probabilité dépend de la variable d'intérêt.

Soient S l'échantillon entier, Y_i la variable d'intérêt, a_i l'indicatrice de réponse :

$$a_i = \begin{cases} 1 & \text{si } Y_i \text{ est observée} \\ 0 & \text{sinon} \end{cases}$$

Soit $p_i = P(a_i = 1 | i \in S)$ la probabilité de réponse pour l'individu i . Les $(p_i)_{i \in S}$ ne sont pas connues. On suppose que les individus répondent indépendamment les uns des autres.

Le premier mécanisme de non réponse (MCAR), appelé aussi mécanisme uniforme, correspond au cas où :

$$\forall i \in S, p_i = p \tag{MCAR}$$

La probabilité de réponse ne dépend d'aucunes variables propres à l'individu. Ce type de données manquantes correspond à un cas particulier de sondage aléatoire simple. Il s'observe relativement rarement et les non-réponses sont essentiellement attribuées au travail de terrain. Dans les cas de (MCAR), aucun redressement n'est nécessaire et le biais des estimateurs est nul. Cette hypothèse est très forte et très rarement vérifiée dans les enquêtes.

Le mécanisme non confondu (MAR) est plus plausible dans la pratique que MCAR :

$$\forall i \in S, p_i = P(a_i=1|y,z) = P(a_i=1|z) \quad (\text{MAR1})$$

$$\forall i \in S, \log\left(\frac{p_i}{1-p_i}\right) = \lambda_0 + \lambda_1 z_i \quad (\text{MAR2})$$

Il implique que les données observées représentent un échantillon aléatoire des données complètes, mais seulement au sein de catégories définies par les variables observées. La dépendance entre non-réponse et variable d'intérêt doit être entièrement expliquée par des variables observées. La probabilité de réponse dépend du vecteur de variables auxiliaires X mais pas directement de la variable d'intérêt Y . Cette forme de non-réponse (tout comme le mécanisme MCAR) est considérée comme "négligeable" et est communément désignée par l'anglicisme *ignorable* selon l'expression de Rubin (1976) et Little et Rubin (2002). Comme nous le verrons, elle permet de simplifier considérablement les analyses.

Le mécanisme MNAR, ou mécanisme confondu, correspond au cas où la non-réponse est une conséquence directe de la variable d'intérêt :

$$\forall i \in S, p_i = P(a_i=1|y,z) \neq P(a_i=1|z) \quad (\text{MNAR1})$$

$$\forall i \in S, \log\left(\frac{p_i}{1-p_i}\right) = \lambda_0 + \lambda_1 z_i + \lambda_2 y_i \quad (\text{MNAR2})$$

La probabilité de réponse peut dépendre des valeurs observées et inobservées de la variable d'intérêt. Même conditionnellement à des caractéristiques observées, les réponses observées ne sont pas représentatives de tous les individus échantillonnés. Par exemple, les individus ne renseignant pas l'enregistrement statistique et fiscal (ESF) de leur entreprise peuvent être principalement dans le secteur informel, même conditionnellement à l'éducation, au genre, etc.

II.2. Validation des hypothèses de mécanismes de non-réponses

Dans des cas de non-réponse partielle, un moyen simple permet de rejeter ou non l'hypothèse MCAR. Il s'agit de regarder les déterminants de la réponse à une question de l'enquête. Si le taux de non-réponse à la variable d'intérêt dépend significativement d'une variable explicative, cela remet automatiquement en cause l'hypothèse MCAR. Le choix entre les hypothèses MAR et MNAR est plus délicat.

L'estimation d'un modèle de sélection d'Heckman permet de tester le mécanisme de non-réponse (Gautier 2006). Si le test de sélection rejette l'hypothèse nulle d'indépendance entre équation d'intérêt et équation de sélection, alors la sélection ne peut être ignorée. En effet, dans ce cas, l'hypothèse (MAR) est rejetée, puisque, même conditionnellement aux variables explicatives incluses dans le modèle, la probabilité de répondre est liée à la variable d'intérêt. Enfin, il vaut mieux effectuer un redressement partiel en les supposant, à tort, MAR plutôt qu'ignorer le risque de biais (Chavance 2004). La méthode que nous proposons permet, en présence de l'hypothèse MAR, de calculer la part du secteur informel urbain de façon non biaisée et robuste.

II.3. Méthode d'imputation

Imputation sans correction des biais de sélection

L'imputation aléatoire simple consiste à estimer un modèle dont la variable dépendante est la variable d'intérêt, sur l'ensemble des observations complètes. Puis, sur la base des coefficients estimés et des caractéristiques des non-répondants, nous imputons des réponses à ces non-répondants. Cette méthode nécessite que la distribution conditionnelle de la variable d'intérêt (conditionnelle à des caractéristiques observées) soit la même chez les répondants que chez les non-répondants. C'est à dire que si les répondants et les non-répondants diffèrent, ce n'est pas sur la base de la variable d'intérêt, mais sur la base d'autres caractéristiques observables. Autrement dit, cette méthode n'est valide qu'en présence d'un mécanisme de réponse non confondu (MAR).

Imputation avec correction des biais de sélection

Supposons que l'hypothèse MAR soit valide, mais qu'elle ne puisse pas être vérifiée parce que toutes les variables expliquant la réponse n'ont pas été collectées lors de l'enquête. Dans ce cas, nous proposons une imputation aléatoire avec correction de sélection, à la Heckman (1976). Elle est proche de la méthode précédente mais permet également de conditionner la distribution de la variable d'intérêt par des variables inobservables qui pourrait affecter à la fois la probabilité de répondre et la variable d'intérêt déclarée. Cet effet peut être dû à des inobservables liées à un certain type de capital humain (par exemple la connaissance de certaines démarches administratives, ou l'appartenance à un ménage dans lequel un membre dispose de ces connaissances).

Soit U_{ij} le niveau d'utilité apportée par l'alternative j , $j=0,1$ (l'entreprise n'est pas enregistrée $Y_i=0$ ou l'entreprise est enregistrée $Y_i=1$) à l'individu i . Supposons que l'on soit capable de modéliser ce niveau d'utilité à partir de variables explicatives que l'on note x_i :

$$U_{ij} = \beta_j x_i + \varepsilon_{ij}$$

On définit alors le différentiel d'utilité y_i^* , la variable latente du modèle dichotomique, comme :

$$y_i^* = U_{i1} - U_{i0} = \beta \cdot x_i + u_i \quad \text{où} \quad \beta = \beta_1 - \beta_0 \quad \text{et} \quad u_i = \varepsilon_{i1} - \varepsilon_{i0}$$

Le choix de l'alternative dépend donc du gain d'utilité apporté par une alternative par rapport à l'autre :

$$y_i = \begin{cases} 1 & \text{si } y_i^* > 0 \\ 0 & \text{sinon} \end{cases}$$

Spécifions maintenant l'équation de sélection. La sélection est, elle aussi, dichotomique. On note V_{ik} le niveau d'utilité apportée par l'alternative k , $k=0,1$ ($A_i=0$ pour une valeur censurée ou $A_i=1$ pour une valeur observée) à l'individu i , déterminé par le vecteur de variables explicatives g_{ij} , et a_i^* le différentiel d'utilité associé :

$$V_{ik} = \gamma_k \cdot g_i + \eta_{ik}$$

$$a_i^* = V_{i1} - V_{i0} = \gamma \cdot g_i + v_i \quad \text{où} \quad \gamma = \gamma_1 - \gamma_0 \quad \text{et} \quad v_i = \eta_{i1} - \eta_{i0}$$

$$a_i = \begin{cases} 1 & \text{si } a_i^* > 0 \\ 0 & \text{sinon} \end{cases}$$

Le modèle de sélection est le suivant :

$$\begin{cases} P(y_i = 1 | x_i) = \phi(\beta \cdot x_i) & \text{avec } y_i \text{ observé pour } a_i = 1 \\ P(a_i = 1 | g_i) = \phi(\gamma \cdot g_i) \end{cases} \quad (M)$$

Si on suppose que $(\varepsilon_{i1}, \varepsilon_{i0})$ d'une part, et (η_{i1}, η_{i0}) d'autre part, suivent des lois normales jointes, alors la spécification adoptée est celle d'un modèle probit, avec sélection elle aussi de type probit. Si, pour les deux équations, les résidus sont supposés être distribués selon une loi logistique, alors, la spécification sera de type logit. La spécification logit présente les inconvénients de devoir choisir une distribution logistique et d'une plus grande complexité dans les calculs. Aussi, le modèle de sélection estimé sera un modèle probit avec censure. La fonction ϕ est la fonction de répartition de la loi normale. La méthode du maximum de vraisemblance permet d'estimer ce modèle². Pour que celui-ci soit identifié, il est nécessaire de disposer d'une variable z_i , contenue dans g mais non dans x .

II.4. Variables identifiantes

² Nous utilisons les commandes `probit` et `heckprob` de Stata 9.0 pour estimer respectivement les modèles probit simples et les modèles probit avec sélection.

La modélisation de la sélection nécessite l'utilisation de variables dites identifiantes dans l'équation de sélection. Elles sont supposées influencer sur la probabilité d'être sélectionné, ici sur la probabilité de répondre, mais pas sur la variable d'intérêt (ESF ou CF). Nous avons choisi de créer 10 variables indicatrices correspondant au taux de réussite de l'enquêteur (moins de 50% de réponse, de 51% à 55%, de 56% à 60%..., de 96% à 100% de réponse) et d'y adjoindre une variable indicatrice valant 1 si c'est l'individu étudié qui a répondu au questionnaire, 0 si c'est un autre membre du ménage. Ces variables peuvent influencer sur la probabilité de répondre mais pas sur l'appartenance à un secteur institutionnel. Elles sont donc de bonnes variables identifiantes, nécessaires à l'estimation d'un modèle de sélection. Il n'est pas souhaitable d'introduire directement dans les modèles une indicatrice par enquêteur, car cela pose des problèmes de convergence de l'estimation par le maximum de vraisemblance du fait du grand nombre de variables explicatives alors utilisées. En revanche, les résultats ne sont pas sensibles aux autres façons de prendre en compte les performances des enquêteurs (de la façon décrite ci-dessus, en introduisant le taux de réponse de l'enquêteur en variable continue, ou en créant une variable indicatrice par décile de la distribution de ce taux de réponse). L'estimation des modèles est faite sur les données non pondérées. En revanche, les taux d'emploi dans le secteur informel tiennent compte des coefficients d'extrapolation.

II.5. Imputations multiples

Procéder à une seule imputation et calculer les taux d'emploi par secteur à partir de celle-ci est risqué. En effet, cette imputation résulte d'un tirage aléatoire (selon une loi uniforme entre 0 et 1) qui peut, par malchance, se situer en marge de la distribution aléatoire. D'autres imputations peuvent donner des taux différents. Aussi, pour que nos résultats soient robustes, 500 procédures d'imputations de couples de réponses à l'enregistrement statistique et à la tenue de comptabilité formelle ont été répétées. Nous calculons le taux d'emploi par secteur pour chaque procédure d'imputation, puis la moyenne (et l'écart-type) des 500 taux d'emploi par secteur ainsi obtenus. L'imputation donnant le taux d'emploi dans une entreprise informelle le plus proche du taux moyen obtenu à la suite des 500 réplications est finalement choisie et sera conservée pour les travaux futurs. Nous avons effectué 500 réplications pour obtenir un taux final égal au taux moyen, au moins jusqu'au deuxième chiffre décimal, pour toutes les villes.

Les modèles sont estimés sur les données, parfaitement comparables, des enquêtes 1-2-3 du projet PARSTAT, récoltées dans 7 capitales économiques de l'Union Économique et Monétaire d'Afrique de l'Ouest.

III. Caractérisation des mécanismes de non-réponse et validation de la méthode

III.1. La question de la non-réponse sur l'informalité des entreprises dans les enquêtes 1-2-3 PARSTAT

L'enquête 1-2-3 a été développée pour répondre à l'échec de la technique classique qui consiste à recenser les établissements et à en extraire ceux appartenant au secteur informel. Cette technique ne permet pas d'obtenir un échantillon représentatif d'unités de production informelle (UPI). Une stratégie alternative a été mise au point qui est celle des enquêtes mixtes ou en phases (ménages/établissements), dont l'enquête 1-2-3 est une pionnière.

Elle consiste à identifier un échantillon représentatif d'UPI à partir d'une enquête sur l'activité des ménages pour ensuite réaliser une seconde enquête spécifique sur ces dernières. Pour procéder à cette identification, sont posées à tous les actifs occupés des ménages de phase 1, des questions relatives à l'entreprise dans laquelle ils travaillent. Ces questions sont : "L'entreprise dans laquelle vous travaillez, ou que vous dirigez, possède-t-elle un identifiant fiscal ?" et pour les travailleurs indépendants (patrons et travailleurs à leur propre compte) seulement "Tenez-vous une comptabilité ?".

Dans chaque ville, 2500 ménages sont enquêtés. La population d'intérêt est celle des 34715 actifs occupés, de 10 ans et plus, sur laquelle le taux d'emploi dans le secteur informel est calculé (Tableau 1). L'imputation concerne les actifs occupés de 10 ans et plus déclarant travailler dans une entreprise privée. En effet, il est d'usage de classer les fonctionnaires dans le secteur formel et les employés des ménages dans le secteur informel.

Les données des enquêtes 1-2-3 sont soumises au phénomène de non-réponse, notamment aux questions d'enregistrement statistique ou fiscal de l'entreprise, et de tenue de comptabilité formelle par les travailleurs indépendants. Le taux de non-réponse à la question d'ESF va de 4,9% à Abidjan à 14,6% à Niamey, celui à la question de tenue de comptabilité de 0,5% à Dakar à 12,6% à Bamako (Tableau 2).

Tableau 2 : Importance de la non-réponse aux critères de formalité

	Niamey	Ouaga-dougou	Dakar	Bamako	Cotonou	Lomé	Abidjan
% de non-réponse à l'ESF	14,55	13,90	9,20	6,59	7,10	10,34	4,86
% de non-réponse à la CF	4,74	2,15	0,47	12,60	1,20	11,65	0
Effectif des NR à la CF	93	51	13	341	36	317	0

*Champ : Actifs occupés de 10 ans et plus, déclarant travailler dans une entreprise privée
Données non pondérées*

L'hypothèse MCAR, pour l'ensemble des travailleurs, est rejetée pour toutes les villes. En effet, la non-réponse est principalement due aux travailleurs dépendants qui sont très logiquement moins au fait du fonctionnement administratif de l'entreprise dans laquelle ils travaillent (Tableau 3).

Tableau 3 : CSP et non-réponse à la question d'ESF

	Niamey	Ouaga-dougou	Dakar	Bamako	Cotonou	Lomé	Abidjan
--	--------	--------------	-------	--------	---------	------	---------

% de NR à l'ESF	14,55	13,90	9,20	6,59	7,10	10,34	4,86
% de NR à l'ESF chez les indépendants	4,59	4,56	4,19	2,55	1,07	2,83	1,11
% de NR à l'ESF chez les dépendants	28,92	28,71	14,91	18,14	18,83	25,48	8,91
% d'indépendants parmi les NR	18,63	20,11	24,26	28,63	9,91	18,29	11,82
% d'indépendants parmi les répondants	65,95	67,97	56,18	77,27	70,33	72,43	53,99
% d'indépendants	59,07	61,32	53,24	74,06	66,04	66,83	51,94

*Champ : Actifs occupés de 10 ans et plus, déclarant travailler dans une entreprise privée
Données non pondérées*

Si les travailleurs indépendants sont majoritaires dans les travailleurs d'entreprises privées, les proportions sont totalement inversées parmi les non-répondants. Cette sur-représentation des travailleurs dépendants parmi les non-répondants se retrouve dans toutes les villes dans des proportions similaires. La non-réponse à la question d'ESF serait donc principalement due à la méconnaissance du fonctionnement administratif de leur entreprise par les travailleurs dépendants.

Cela implique d'étudier séparément les mécanismes de non-réponse des travailleurs dépendants et indépendants (patrons et travailleurs à leur compte propre). Il est possible que les mécanismes de réponse soient différents selon ces deux sous-populations. Les tableaux 4 et 5, présentent respectivement la caractérisation descriptive de la non-réponse des travailleurs dépendants et des travailleurs indépendants à la question de l'ESF³. Ils permettent de rejeter l'hypothèse (MCAR), par, entre autres, les constats suivants. Les travailleurs dépendants possédant un contrat de travail écrit répondent davantage. La non réponse est masculine chez les travailleurs indépendants.

III.2. Estimation des modèles

Les déterminants de la tenue d'une comptabilité formelle ne peuvent être estimés par un modèle dichotomique avec sélection que pour deux villes, celles dont l'effectif de non-réponse parmi les travailleurs indépendants est le plus élevé, Bamako et Lomé (Tableau 2). En effet, les effectifs de non-réponse sont très faibles dans les autres capitales économiques, variant de 13 observations à Dakar (0,5%) à 93 à Niamey (4,7%), ce qui rend leur impact sur l'estimation du taux d'informalité négligeable. Dans les capitales malienne et togolaise, l'effectif de non-réponse est suffisant avec respectivement 341 et 317 travailleurs indépendants (11,7% et 12,6%) n'ayant pas répondu à la question de tenue de comptabilité. On note Y_i^1 la variable d'ESF et A_i^1 l'indicatrice de réponse correspondante, et Y_i^2 la variable de comptabilité et A_i^2 l'indicatrice de réponse correspondante. Les modèles suivants sont estimés :

³ disponibles sur demande auprès des auteurs

- E1_D* : Estimation Probit de a_i^1 pour les travailleurs dépendants.
- E2_D* : Estimation Probit de y_i^1 pour les travailleurs dépendants.
- E1_I* : Estimation Probit de a_i^1 pour les travailleurs indépendants.
- E2_I* : Estimation Probit de y_i^1 pour les travailleurs indépendants.
- C1_I* : Estimation Probit de a_i^2 pour les travailleurs indépendants.
- C2_I* : Estimation Probit de y_i^2 pour les travailleurs indépendants.
- E3_D* : Estimation Probit de y_i^1 , avec sélection, pour les travailleurs dépendants.
- E3_I* : Estimation Probit de y_i^1 , avec sélection, pour les travailleurs indépendants.
- C3_I* : Estimation Probit de y_i^2 , avec sélection, pour les travailleurs indépendants (seulement à Bamako et Lomé).

Les probabilités prédites par les modèles de sélection nous permettent d'imputer des réponses aux non-répondants. La valeur imputée est égale à l'arrondi du produit entre la probabilité prédite par le modèle et un nombre tiré selon une loi uniforme sur l'intervalle [0 ; 1]. Rappelons que la non-réponse à la comptabilité des travailleurs indépendants est marginale. De plus, pour qu'un travailleur dépendant soit considéré comme travaillant dans une entreprise formelle, il suffit qu'il déclare son entreprise enregistrée. Dans ce cas, la tenue d'une comptabilité par son dirigeant est donc totalement ignorée⁴. Les faibles effectifs de non-réponse pour la comptabilité ainsi que la plus grande sévérité des critères pour les indépendants légitiment d'utiliser la même procédure d'imputation que pour les travailleurs dépendants. Tout travailleur indépendant dirigeant une entreprise enregistrée (resp. non enregistrée) et ne répondant pas à la question de comptabilité se verra affecté au secteur formel (resp. informel). Les travailleurs indépendants n'ayant pas non plus répondu à la question sur l'ESF se verront imputer une réponse à la question de comptabilité.

III.3. L'ESF des travailleurs dépendants

Trois principales raisons peuvent expliquer la non-réponse à la question d'ESF de la part d'un travailleur dépendant.

⁴ La solution serait de distinguer, parmi les travailleurs dépendants déclarant leur entreprise enregistrée, ceux dont le patron tient probablement une comptabilité et ceux dont le patron n'en tient probablement pas. Cependant, ceci ne pourrait se faire que grâce à un plus grand nombre de variables de l'entreprise, renseignées pour les deux sous-populations, ce qui n'est pas le cas.

La première est la crainte de "retombées", si l'enquêté n'est pas convaincu de l'anonymat de l'enquête. Soit l'informalité pousse directement à la non-réponse, par crainte de pénalité en cas de non-enregistrement (amende ou *bakchich*), soit la formalité pousse à la non-réponse par crainte de questions plus précises et de pression fiscale. Cette raison correspondrait à un mécanisme de réponse confondu (MNAR), sans détermination possible du secteur institutionnel impliquant la non-réponse.

La deuxième raison est la qualité du travail de l'enquêteur. Elle correspond à un mécanisme aléatoire, le fait de ne pas répondre ne dépend d'aucune caractéristique, observée ou non de l'enquêté. Cela correspondrait à l'hypothèse MCAR, situation dans laquelle seul l'effet du personnel de terrain influe sur le fait de répondre ou non.

La troisième raison est la méconnaissance du statut juridique et fiscal de l'entreprise dans laquelle l'enquêté travaille. Les déterminants de la non-réponse sont alors l'éducation, la taille de l'entreprise (plus l'entreprise est grande, moins le travailleur dépendant est au fait de son caractère administratif), le type d'emploi (les emplois administratifs seront caractérisés par une plus grande probabilité de répondre que les emplois techniques). Or l'éducation joue, d'une part, sur l'appartenance à un secteur institutionnel (de façon très schématique, l'éducation formelle pousse vers le secteur formel, l'apprentissage et la faible éducation vers le secteur informel), d'autre part sur la connaissance et l'intérêt porté au fonctionnement administratif de son entreprise.

L'hypothèse MAR revient ici à considérer que des variables explicatives observées influent sur la probabilité de répondre à la question d'enregistrement statistique (entre autres le capital humain). Au sein de chaque catégorie formée par ces variables, la probabilité de répondre serait aléatoirement distribuée. Par exemple, toutes les femmes de 25 à 50 ans, ayant un niveau d'éducation primaire, devraient avoir la même probabilité de répondre, qu'elles travaillent dans une entreprise informelle ou non.

Ces deux dernières raisons, la qualité du travail de l'enquêteur et la méconnaissance du statut juridique et fiscal de l'entreprise, nous semblent réalistes. Une partie de la non-réponse est due au fait que les travailleurs dépendants ignorent très souvent le fonctionnement administratif de leur entreprise, quel que soit le secteur institutionnel (hypothèse MAR). Les statistiques descriptives montrent en effet que les travailleurs dépendants qui répondent sont plus éduqués, occupent des emplois plus qualifiés, dans des entreprises de plus grande taille, et possèdent plus souvent des contrats de travail. A la méconnaissance éventuelle du fonctionnement de l'entreprise s'ajoute l'effet aléatoire de la performance de l'enquêteur, responsable en partie de la non-réponse des enquêtés qu'il doit interroger.

Cependant, le mécanisme de non-réponse peut sembler confondu (MNAR), puisque les travailleurs dépendants ayant des caractéristiques de travailleurs du secteur informel répondent moins (Tableau 4). Nous faisons l'hypothèse que des variables de capital humain, observées ou non, influent à la fois sur le fait de répondre, et à la fois sur le type d'entreprise dans lequel l'individu travaille ; et nous tenons compte de l'effet de sélection.

L'analyse *toutes choses égales par ailleurs* des déterminants de la non-réponse conforte le choix de l'hypothèse MAR, due la qualité du travail de l'enquêteur et la méconnaissance du statut juridique et fiscal de l'entreprise. Sexe, âge et éducation n'influent pas sur la probabilité de répondre des travailleurs dépendants (Tableau 6, colonnes de droite). Les branches d'activité jouent très peu, seulement dans certaines villes, et dans des sens opposés. La possession de bulletin de paie et celle de contrat écrit sont positivement corrélée à la réponse et semblent déterminantes. Ainsi, les travailleurs ne répondant pas sont ceux qui, sans contrat ni bulletin de paie, ignorent le statut de l'entreprise dans laquelle ils travaillent. Les tests de Fisher de l'effet de collecte montrent que la qualité du terrain est le plus grand déterminant de la réponse. Malgré ce fort pouvoir explicatif, les pseudo R^2 n'atteignent que 26% à Abidjan et descendent jusqu'à 15% à Ouagadougou. L'hypothèse MAR est choisie ici, correspondant à la méconnaissance des travailleurs dépendants du fonctionnement administratif de l'entreprise dans laquelle ils travaillent et aux contre-performances de certains enquêteurs.

III.4. L'ESF des travailleurs indépendants

Certains facteurs identifiant bien la non-réponse, l'hypothèse MCAR est également rejetée pour les seuls indépendants.

Ces travailleurs peuvent plus difficilement ignorer le fonctionnement administratif de l'entreprise qu'il dirige (cela peut toutefois arriver cependant, en cas de reprise de l'activité d'un parent par exemple). Les indépendants sont les travailleurs qui devraient avoir le plus de crainte de révéler le caractère administratif et fiscal de leur entreprise (MNAR). Ils déclarent cependant relativement bien leur ESF (de 1,1% de NR à 4,6% de NR seulement).

Les statistiques descriptives du Tableau 5 montrent que les indépendants ne répondant pas à la question d'ESF sont en moyenne plus éduqués, à Cotonou, et dans une plus faible mesure à Lomé. De même les cadres et assimilés répondent moins que les autres dans ces deux villes. Les autres branches d'activité, ainsi que la possession d'un contrat (du reste très rare pour un indépendant) et la régularité de l'emploi ne semblent pas corrélées à la non-réponse. Ces variables sont pourtant de bons critères du caractère formel de l'entreprise. Cette absence de corrélation jouerait davantage en faveur de l'hypothèse MAR. En revanche, la taille de l'entreprise (supérieure ou égale à 6 personnes) est corrélée à la non-réponse, alors qu'elle est parfois supposée être un indicateur de formalité.

L'analyse multivariée permet d'affiner cette description (Tableau 7, colonnes de droite). Chez les travailleurs indépendants, être une femme influe là encore sur la probabilité de répondre. L'âge et l'éducation ne jouent pas sur la probabilité de répondre, hormis à Cotonou où le fait d'avoir au moins le niveau primaire favorise la non-réponse. La probabilité de répondre n'est jamais dépendante du niveau de rémunération, excepté à Dakar. Dans certaines capitales, ouvriers ou artisans répondent davantage. Cependant, il n'existe pas de profil d'emploi particulier lié à la non-réponse à la question d'ESF.

Enfin, les variables de collecte (performance de l'enquêteur et indicatrice valant 1 si l'intéressé a répondu lui-même) ont un rôle très important. Sans ces variables de terrain, les modèles de réponse ont un faible pouvoir explicatif (pseudo-R² variant de 7% à 20%), avec elles, les pseudo R² varient de 13% à 38%. Les F-tests de nullité jointe des coefficients de ces variables montrent eux aussi l'importance de l'effet "collecte".

Ainsi les travailleurs indépendants peuvent avoir des réticences à répondre (MNAR) et la capacité de l'enquêteur à vaincre ces réticences peut avoir un rôle.

III.5. La tenue d'une comptabilité formelle par les travailleurs indépendants

Les travailleurs indépendants peuvent craindre que l'enquêteur leur demande de mettre le nez dans leurs comptes, même si l'enquêteur doit les rassurer sur l'anonymat de l'enquête. Cependant, les faibles taux de refus de répondre à la phase 2 de l'enquête (qui porte entre autres sur les comptes détaillés des unités de production informelle) laissent à penser que les travailleurs indépendants n'ont pas de réticences particulières à parler de leur comptabilité.

Par ailleurs, une fois une entreprise informelle identifiée, l'enquêteur doit relever son adresse exacte, pour ensuite procéder, quelques semaines après, à la seconde phase de l'enquête. L'enquêteur peut être tenté de ne pas relever toutes les UPI, pour alléger son travail. La non-réponse, d'ailleurs plus faible pour cette question, a donc plus de chance d'être due à l'enquêteur.

Il n'y a aucune non-réponse à cette question dans la ville d'Abidjan. Cela ne provient pas de la bonne volonté des enquêtés ou de la persévérance des enquêteurs, mais d'une imputation des réponses lors de l'apurement des fichiers après la saisie. Or nous ne disposons d'aucune information sur le mode d'imputation et il n'est plus possible de retrouver les valeurs originales. Pour les six autres capitales économiques, le taux de non-réponse va de 0,5% à Dakar à 12,6% à Bamako. Les effectifs de travailleurs indépendants ne répondant pas à cette question sont donc très faibles, ce qui est préférable du point de vue de l'analyse mais qui posera des problèmes de spécification des estimations de modèles de sélection utilisés pour l'imputation.

La rémunération est positivement corrélée à la probabilité de répondre (Tableau 8, colonnes de droite). La taille de l'entreprise joue là encore négativement sur la réponse. C'était déjà le cas pour la question relative à l'ESF, à croire que les données manquantes sont le fait seulement des grandes entreprises (6 personnes et plus)⁵. Il est possible que les enquêteurs ne posent pas la question de CF si la réponse leur semble évidente du fait de la taille de l'entreprise. Cependant, cette hypothèse n'est pas vérifiée dans deux villes, Dakar et Cotonou, où la non-réponse à la question de CF (rare) n'est pas dépendante de la qualité du terrain, puisque les statistiques des tests de Fisher de nullité jointe des coefficients des variables de collecte ne sont plus significatives.

IV. Application du modèle de sélection et résultats de l'imputation probabiliste

IV.1. Déterminants de la non-réponse

Les analyses précédentes montrent que le pouvoir explicatif des caractéristiques individuelles et de l'entreprise sur la non-réponse est faible. La probabilité de répondre semble en grande partie dépendre de la qualité du terrain, en particulier pour l'ESF. Quel que soit le secteur institutionnel auquel ils appartiennent, les travailleurs dépendants peuvent ignorer l'ESF de leur entreprise et/ou être interrogés par un mauvais enquêteur.

Si les travailleurs indépendants peuvent plus difficilement ignorer l'ESF de leur entreprise, rien ne permet d'affirmer que la formalité ou l'informalité de l'entreprise influe sur la non-réponse, d'autant plus que les variables de terrain jouent là aussi de façon très significative.

Il en est de même pour la tenue d'une comptabilité formelle, pour laquelle la qualité de la phase 2 des enquêtes 1-2-3 permet d'affirmer qu'un travailleur indépendant n'est pas tenté de dissimuler ses comptes à l'enquêteur. L'hypothèse MAR est donc probable et une imputation aléatoire simple sur la base des caractéristiques observées X aboutirait à des estimateurs non biaisés.

Cependant l'effet du capital humain (éducation, profession intellectuelle) favoriseraient la réponse des dépendants et l'effet taille de l'entreprise inciterait dépendants et indépendants à ne pas répondre. Ces effets peuvent être contrôlés par un modèle de sélection à la Heckman. Avec un tel modèle, les effets de l'éducation et de la taille de l'entreprise sur l'ESF seront nets, corrigés du biais de sélection dû à la non-réponse. En effet, si l'éducation et la taille d'entreprise sont liées à l'informalité et à la non-réponse, cette procédure corrige le biais que ces corrélations induiraient.

Nous faisons l'hypothèse qu'il existe une (des) variable(s) influant à la fois sur l'ESF et sur la probabilité de répondre, ce qui est exactement le cas d'application d'un modèle de sélection

⁵ Les possibles interactions entre taille et revenu ont été testées et ne biaisent pas les estimations.

à la Heckman sur variable dépendante dichotomique, qui suffit à redresser le biais de non-réponse (voir Sales *et al.* 2004 dans le cas d'une variable dépendante linéaire)

IV.2. Rareté des effets de sélection : validation de l'hypothèse MAR.

L'estimation du modèle E3D montre que pour six capitales sur sept, il n'y a aucun effet de sélection dans l'estimation de la probabilité des travailleurs dépendants de diriger une entreprise enregistrée (Tableau 9). A Ouagadougou seulement, un effet de sélection est observé. Le test de sélection montre que les variables inobservées influent sur l'ESF et sur la probabilité de répondre mais dans des sens opposés. Ce serait donc la seule ville pour laquelle l'hypothèse (MAR) n'est pas validée et donc pour laquelle la prise en compte de la sélection est nécessaire à l'imputation.

La conclusion est similaire pour les travailleurs indépendants (Tableau 10), pour lesquels un effet de sélection est observé pour la seule ville de Bamako. Il n'existe donc pas de variables inobservées corrélées à l'informalité et influant sur la probabilité de répondre, à l'exception de Bamako (test d'existence de sélection significatif au seuil de 5%). La sélection y est cependant très faible, puisque les coefficients estimés des analyses avec et sans prise en compte de la sélection sont quasiment égaux.

Le modèle de sélection concernant la tenue de comptabilité n'a pu être estimé que pour les villes de Bamako et Lomé (Tableau 11). Dans ces deux villes, l'effet de sélection est significatif, mais dans des sens différents. Pour les autres villes, les non-réponses sont en nombre tellement faible qu'elles ne peuvent très certainement pas biaiser les estimations. Les coefficients estimés par les modèles probit simple et probit avec sélection sont similaires pour les capitales malienne et togolaise. Aussi la sélection y est faible.

La rareté des effets de sélection se retrouve dans la faible variation des probabilités prédites (Tableau 12). Pour les travailleurs dépendants comme indépendants, la probabilité moyenne d'être enregistré et son écart-type sont égaux ou très similaires entre les deux spécifications (probit simple ou modèle de sélection). Les statistiques descriptives des probabilités prédites diffèrent à Ouagadougou, la ville pour laquelle a été observé un effet de sélection par la non-réponse des travailleurs dépendants. Un effet de sélection significatif a été observé à Bamako, pour les travailleurs indépendants. Il ne se répercute pas sur la probabilité moyenne de diriger une entreprise enregistrée

Tableau 12 : Sensibilité des probabilités prédites à la spécification des modèles d'imputation

	Probabilité moyenne de travailler dans une ent. enregistrée (travailleurs dépendants)		Probabilité moyenne de diriger une ent. enregistrée (travailleurs indépendants)		Probabilité moyenne de tenir une comptabilité (travailleurs indépendants)	
	Sans sélection	avec sélection	Sans sélection	avec sélection	avec sélection	Sans sélection
Niamey	0,42 (0,32)	0,40 (0,32)	0,10 (0,14)	0,10 (0,15)	0,06 (0,09)	
Ouagadougou	0,49 (0,33)	0,58 (0,30)	0,10 (0,13)	0,09 (0,13)	0,04 (0,08)	

Dakar	0,49 (0,40)	0,47 (0,40)	0,06 (0,10)	0,06 (0,10)	0,08 (0,08)	
Bamako	0,45 (0,32)	0,42 (0,32)	0,07 (0,12)	0,07 (0,12)	0,05 (0,10)	0,04 (0,09)
Cotonou	0,49 (0,39)	0,51 (0,39)	0,03 (0,09)	0,03 (0,09)	0,17 (0,11)	
Lomé	0,42 (0,31)	0,41 (0,31)	0,09 (0,15)	0,09 (0,15)	0,03 (0,07)	0,06 (0,12)
Abidjan	0,56 (0,37)	0,54 (0,37)	0,04 (0,09)	0,04 (0,09)		

Données non pondérées

Les différences de distribution de probabilités prédites avec ou sans sélection, à Ouagadougou pour l'ESF (et dans une moindre mesure à Lomé pour la tenue de comptabilité) montrent l'importance de vérifier l'existence d'une sélection sur inobservables. En effet le modèle probit simple sous-estime, dans ce cas, la probabilité de travailler dans une entreprise enregistrée.

Le tableau 13 ci-dessous résume les choix d'imputations opérés pour chaque capitale :

Tableau 13 : Modèles servant à l'imputation

		Niamey	Ouaga- dougou	Dakar	Bamako	Cotonou	Lomé	Abidjan
ESF		$E3_D$ et $E3_I$ (sélection)						
comptabilité	ESF non renseigné	$C2_I$			$C3_I$	$C2_I$	$C3_I$	$C2_I$
	ESF renseigné	Le travailleur indépendant est considéré tenir une comptabilité si et seulement si l'entreprise est enregistrée.						

$E3_D$ (resp. $E3_I$) : estimation de la probabilité de travailler dans (resp. de diriger) une entreprise possédant un identifiant statistique ou fiscal, pour un travailleur dépendant (resp. indépendant), avec prise en compte de la sélection due à la NR.

$C2_I$: estimation de la probabilité de tenir une comptabilité, pour un travailleur indépendant, sans prise en compte de la sélection due à la NR.

$C3_I$: estimation de la probabilité de tenir une comptabilité, pour un travailleur indépendant, avec prise en compte de la sélection due à la NR.

IV.3. Résultats des imputations et comparaison avec des méthodes alternatives

Les taux d'emploi dans le secteur informel obtenus après imputation sont robustes. En effet, l'écart entre les taux minimum et maximum obtenus par des répliquions d'imputation est compris entre 0,3 et 0,8 points de pourcentage (Tableau 14).

L'imputation a permis de réduire de façon satisfaisante les taux de données manquantes à la question de l'ESF (Tableau 15). Les valeurs manquantes résiduelles aux critères de formalités sont dues à des valeurs manquantes des variables explicatives du modèle de sélection. Le taux maximal de non-réponse à la question d'ESF est maintenant de 2,5% (observé à Cotonou), contre 14,6% avant imputation (observé à Niamey).

L'imputation a été moins efficace pour la comptabilité formelle. En effet, les individus n'ayant pas répondu à cette question ont très souvent des variables explicatives non renseignées.

L'impact est faible car la CF ne concerne qu'une partie de l'échantillon des travailleurs (les indépendants) et la non-réponse y était initialement faible.

Les taux obtenus par cette procédure d'imputation (notée M0) seront comparés aux taux issus de traitements différents de la non-réponse, correspondant aux méthodes ci-dessous.

M1 : Calcul du taux en ignorant complètement la non-réponse.

M2 : Calcul du taux en considérant toute unité dont le travailleur ne répond pas à la question d'ESF non enregistrée statistiquement ou fiscalement, et tout chef d'unité ne répondant pas à la question de comptabilité comme ne tenant pas de comptabilité. Cette méthode correspond, avec les notations de la section II.4, à :

$$a_i^1 = 0 \rightarrow y_i^1 = 0 \text{ et } a_i^2 = 0 \rightarrow y_i^2 = 0.$$

M3 : Taux figurant dans Statéco 99. Normalement, ces taux ont été calculés sur la base de calcul de la variable « SI » dont le programme de construction est résumé par le Tableau 16. La variable a par la suite été corrigée « manuellement » à l'aide de critères tels que le type de contrat, la taille de l'entreprise, le type d'activité, les conditions d'emploi ou la possession d'un bulletin de paie (des variables intervenant dans les modèles estimés ici), pour tenir compte au cas par cas des valeurs manquantes non considérées dans le programme initial.

Tableau 17 : Sensibilité du taux d'emploi dans le secteur informel au traitement de la non-réponse

	M0 : imputation [IC 95%]	M1 : en ignorant la NR	M2 : la NR équivaut à l'informalité	M3 : Statéco 99
Niamey	70,89 [70,68 ; 71,10]	68,13	72,40	71,1
Ouagadougou	72,44 [72,14 ; 72,75]	70,95	74,41	73,4
Dakar	75,46 [75,35 ; 75,57]	73,94	76,04	76,4
Bamako	77,81 [77,66 ; 77,95]	76,79	79,76	77,5
Cotonou	80,24 [80,05 ; 80,42]	79,88	81,18	80,3
Lomé	80,44 [80,24 ; 80,64]	79,18	82,26	81,0
Abidjan	74,35 [74,24 ; 74,47]	73,92	74,94	74,7

*Données pondérées, Ensemble des actifs occupés de 10 ans et plus
Les taux compris dans l'intervalle de confiance à 95% de M0 sont marqués en gras.*

Les taux obtenus en faisant varier le traitement des données manquantes sont présentés dans le tableau 17 ci-dessus. Ils attestent de la sensibilité de la répartition institutionnelle à la prise en compte des non-réponses, puisque le traitement des non-réponses peut faire varier le taux d'emploi dans le secteur informel de plus de 4 points de pourcentage.

Le taux d'emploi informel calculé après l'imputation probabiliste M0 est systématiquement plus élevé que celui calculé en ignorant complètement les données manquantes, M1 (faisant implicitement l'hypothèse MCAR). Les non-répondants seraient donc plus probablement des

travailleurs de l'informel, ce qui ne justifie pas pour autant d'imputer directement les non-répondants à ce secteur (M2).

La méthode probabiliste identifie une légère surestimation des taux publiés dans 6 villes sur 7, significative pour 4 villes. A Cotonou, le taux publié appartient à l'intervalle de confiance obtenu par les 500 répliques. A Niamey, il est égal à la borne supérieure de l'intervalle. A Bamako seulement, le taux d'emploi dans le secteur informel publié dans Statéco 99 est inférieur à celui obtenu après l'imputation probabiliste, en dehors même de l'intervalle de confiance.

L'imputation probabiliste confirme cependant les différences de taux d'emploi dans le secteur informel entre les capitales, relevées dans Statéco n°99. Les capitales voisines Lomé et Cotonou sont celles où le taux d'emploi dans le secteur informel est le plus élevé (supérieur à 80%). Viennent ensuite Bamako, Dakar, Abidjan, Ouagadougou, et Niamey avec 70,9% d'emplois dans le secteur informel.

Ces résultats sur le secteur informel ne sont pas généralisables à toutes les enquêtes emplois en Afrique. Le mécanisme de non-réponse et la procédure d'imputation doivent être étudiés pour chaque enquête, comme le prouvent les exceptions de Ouagadougou, Bamako et Lomé.

V. Conclusions

La répartition du marché du travail entre les différents secteurs institutionnels est sensible au traitement des données manquantes. Aussi il est important de traiter celles-ci de façon homogène et non biaisée. Le traitement de la non-réponse à la question relative à la tenue d'une comptabilité formelle à Abidjan rappelle que si une imputation est effectuée en amont de l'analyse, à un quelconque stade de la saisie ou du traitement des données, il est nécessaire de la documenter. Attribuer systématiquement une valeur positive ou négative à la non-réponse revient à ignorer les déterminants de la non-réponse et introduit donc un biais dans la mesure des différents secteurs. Ceci peut se répercuter sur la composition et les caractéristiques mêmes des secteurs.

La méthode d'imputation probabiliste proposée dans ce document montre que la non-réponse, en général, est liée positivement à l'appartenance au secteur informel. Cela ne veut pas dire qu'il est souhaitable d'imputer directement les non-répondants à ce secteur. La non-réponse à la seule question de comptabilité n'est pas forcément liée à l'informel dans toutes les villes. C'est davantage la non-réponse à la question de l'enregistrement statistique qui est corrélée avec l'appartenance au secteur informel.

Les taux d'emploi dans le secteur informel obtenus après imputation sont proches de ceux figurant dans Statéco n°99, l'écart étant de moins d'un point de pourcentage, eux-mêmes obtenus à partir d'une variante de la définition SCN93 du secteur informel incluant la taille de

l'entreprise et d'autres variables corrélées avec l'informalité lors des cas indéterminés. Mais l'imputation probabiliste est malgré tout préférable parce qu'elle est basée sur une estimation toutes choses égales par ailleurs. Lorsqu'elle contrôle pour les effets de sélection, l'imputation probabiliste permet d'estimer sans biais les indicateurs. Elle ne nécessite pas de présupposer *a priori* du sens de l'effet des variables explicatives sur le caractère informel de l'entreprise. Enfin, elle affranchit le statisticien de l'étude fastidieuse des multiples croisements de variables explicatives. À l'avenir, il est préférable de procéder à l'imputation probabiliste, en prenant garde à la qualité de la spécification (choix des variables explicatives, contrôle des biais de sélection...), faillible tout comme le statisticien.

Mesurer l'importance du secteur informel défini à partir des deux critères d'enregistrement fiscal ou statistique et de tenue d'une comptabilité formelle n'est donc pas si simple qu'il paraît au premier abord. D'ailleurs, les instances internationales se penchent à nouveau sur les différentes mesures du secteur informel. Leurs nouvelles résolutions font mention de la possession de bulletin de paie ou de régularité de l'emploi comme critère supplémentaire de mesure du secteur informel. Plus qu'une différence méthodologique, il s'agit d'un changement conceptuel puisque les populations concernées diffèrent. L'introduction des critères de bulletin de paie ou de contrat écrit centre l'analyse sur le travailleur, alors que la définition du SCN93 centrait l'analyse sur l'entreprise.

La question de l'unité d'analyse est plus qu'une simple question académique. En effet, le choix de cette unité est influencé par les objectifs politiques et influe à son tour les politiques préconisées. Ne pas se restreindre à une analyse en termes d'unités de production diminue le risque de confusion entre travailleur et entreprise. En effet, il existe une proportion loin d'être négligeable de travailleurs sans protection sociale ou sans contrat écrit dans le secteur dit formel, privé comme public. Aussi, puisque l'unité d'analyse est liée à l'objectif politique, la statistique ne peut s'appuyer exclusivement sur l'entreprise, au risque d'oublier les travailleurs précaires des entreprises formelles.

VI. Tableaux

Tableau 1 : Echantillons de la phase 1 des enquêtes 1-2-3.

	Niamey	Ouagadougou	Dakar	Bamako	Cotonou	Lomé	Abidjan	Ensemble
Individus enquêtés	14538	13746	19060	13002	11572	9907	11344	93169
Population de la ville	684600	870660	1940270	1166820	811880	802410	3069140	9345780
Actifs occupés de 10 ans et plus	4212	4907	6331	4436	5299	4652	4878	34715
Actifs occupés de 10 ans et plus travaillant dans une entreprise privée	3320	3861	5148	3655	4549	4070	4178	28781
...dont travailleurs dépendants	1359	1493	2407	948	1545	1350	2008	11110
...dont travailleurs indépendants	1961	2368	2741	2707	3004	2720	2170	17671

Tableau 2 : Importance de la non-réponse aux critères de formalité (dans le texte)

Tableau 3 : CSP et non-réponse à la question d'ESF (dans le texte)

Tableau 4 : Caractérisation descriptive de la non-réponse des travailleurs dépendants à la question de l'ESF (disponible sur demande auprès des auteurs)

Tableau 5 : Caractérisation descriptive de la non-réponse des travailleurs indépendants à la question de l'ESF (disponible sur demande auprès des auteurs)

Tableau 6 : Modèles $E1_D$ et $E2_D$: Enregistrement statistique ou fiscal (y_i^1) des travailleurs dépendants, sans prise en compte de la sélection, et mécanisme de réponse à la question de l'enregistrement (a_i^1)

MD1 et MD2	Niamey		Ouagadougou		Dakar		Bamako		Cotonou		Lomé		Abidjan	
	y_i^1	a_i^1	y_i^1	a_i^1	y_i^1	a_i^1	y_i^1	a_i^1	y_i^1	a_i^1	y_i^1	a_i^1	y_i^1	a_i^1
contrat de travail écrit	0,503*** (0,181)	0,480*** (0,158)	0,403** (0,172)	0,371*** (0,138)	0,512*** (0,120)	0,256** (0,116)	0,890*** (0,156)	0,102 (0,183)	0,233 (0,154)	0,263* (0,137)	0,373*** (0,120)	-0,043 (0,100)	0,457*** (0,114)	0,325* (0,174)
bulletin de paie	0,510** (0,200)	0,343** (0,165)	1,156*** (0,255)	0,258 (0,157)	1,467*** (0,127)	0,672*** (0,124)	0,648*** (0,162)	0,569*** (0,208)	1,319*** (0,172)	0,778*** (0,165)	0,627*** (0,174)	0,455** (0,179)	0,736*** (0,130)	0,205 (0,182)
femme	-0,267** (0,122)	0,107 (0,098)	-0,338** (0,132)	0,133 (0,101)	-0,012 (0,109)	-0,157 (0,096)	-0,564*** (0,147)	0,238 (0,155)	-0,203 (0,156)	0,262** (0,124)	-0,027 (0,117)	0,046 (0,100)	-0,009 (0,121)	0,065 (0,149)
âge	0,077*** (0,025)	0,004 (0,019)	0,038* (0,023)	-0,024 (0,020)	0,023 (0,019)	-0,007 (0,018)	0,092*** (0,022)	0,039** (0,017)	0,167*** (0,029)	0,000 (0,027)	0,011 (0,024)	0,052** (0,021)	0,058** (0,027)	-0,032 (0,031)
âge au carré	-0,001*** (0,000)	-0,000 (0,000)	-0,001 (0,000)	0,000 (0,000)	-0,000 (0,000)	0,000 (0,000)	-0,001*** (0,000)	-0,000** (0,000)	-0,002*** (0,000)	0,000 (0,000)	-0,000 (0,000)	-0,001** (0,000)	-0,001* (0,000)	0,000 (0,000)
Au moins niveau primaire	0,104 (0,120)	0,103 (0,100)	0,393*** (0,111)	0,088 (0,088)	0,296*** (0,096)	0,112 (0,090)	-0,207 (0,128)	0,059 (0,131)	0,140 (0,150)	-0,127 (0,131)	0,365*** (0,110)	-0,015 (0,103)	0,266*** (0,101)	0,110 (0,132)
Ent. de 6 personnes et +	0,967*** (0,118)	-0,406*** (0,094)	1,038*** (0,108)	-0,408*** (0,090)	0,980*** (0,109)	-0,124 (0,087)	0,586*** (0,124)	-0,337** (0,133)	0,862*** (0,134)	-0,485*** (0,126)	0,913*** (0,106)	-0,433*** (0,093)	1,175*** (0,107)	0,264* (0,143)
entre 1 et 2 salaires minimaux	0,333** (0,140)	0,201 (0,125)	0,364** (0,147)	-0,044 (0,112)	0,349*** (0,106)	0,139 (0,100)	0,063 (0,148)	-0,104 (0,149)	0,244 (0,152)	-0,020 (0,124)	0,628*** (0,154)	-0,050 (0,140)	0,646*** (0,134)	-0,166 (0,158)
entre 2 et 4 salaires minimaux	0,194 (0,194)	0,023 (0,146)	0,638*** (0,176)	0,078 (0,141)	0,269 (0,168)	0,345** (0,164)	0,371** (0,162)	0,136 (0,174)	0,687*** (0,198)	0,211 (0,175)	0,811*** (0,166)	0,423*** (0,162)	0,502*** (0,133)	-0,109 (0,166)
plus de 4 salaires minimaux	-0,536 (0,378)	0,180 (0,323)	-1,000** (0,427)	0,040 (0,401)	-0,176 (0,415)	-0,633** (0,305)	0,814** (0,401)		0,324 (0,564)	0,806 (0,752)	0,884** (0,407)	0,808** (0,353)	1,038*** (0,259)	0,507 (0,466)
5 ans et plus d'expérience	-0,117 (0,121)	0,149 (0,098)	0,031 (0,128)	0,315*** (0,101)	-0,201** (0,091)	-0,030 (0,078)	-0,307** (0,126)	0,060 (0,130)	-0,295* (0,153)	0,197 (0,138)	0,058 (0,135)	-0,072 (0,126)	-0,148 (0,112)	0,217 (0,141)
Cadres, ingénieurs et intellectuels	0,215 (0,257)	0,163 (0,211)	0,283 (0,323)	0,263 (0,280)	0,061 (0,202)	0,417* (0,216)	-0,449* (0,263)	0,231 (0,334)	0,493* (0,271)	0,370 (0,271)	0,656** (0,307)	-0,405* (0,216)	0,206 (0,216)	-0,497* (0,259)
Prof. intermédiaires et administratives	0,389** (0,197)	-0,063 (0,156)	0,328* (0,192)	0,105 (0,142)	0,365*** (0,137)	0,212 (0,130)	0,114 (0,174)	0,032 (0,183)	0,467** (0,188)	-0,014 (0,161)	0,330* (0,169)	-0,215 (0,146)	0,481*** (0,134)	0,049 (0,191)
Agriculteurs	-1,981*** (0,277)	-0,307* (0,168)	-1,430*** (0,469)	0,236 (0,438)	-1,108*** (0,290)	0,186 (0,252)	-1,260*** (0,252)		-0,436 (0,730)		-0,928*** (0,267)	0,076 (0,278)	-0,346 (0,234)	0,433 (0,452)
Artisans	-0,252* (0,132)	-0,331*** (0,104)	-0,323*** (0,122)	-0,034 (0,095)	-0,705*** (0,121)	-0,045 (0,100)	-0,533*** (0,159)	0,091 (0,147)	-0,527*** (0,170)	0,092 (0,143)	-0,248* (0,132)	-0,157 (0,114)	-0,084 (0,131)	-0,480*** (0,166)
Ouvriers et manœuvres	-0,562*** (0,188)	-0,176 (0,167)	-0,053 (0,205)	-0,104 (0,145)	0,390** (0,185)	0,091 (0,174)	-0,214 (0,177)	0,050 (0,196)	0,279 (0,213)	-0,177 (0,205)	-0,145 (0,167)	0,147 (0,183)	0,602*** (0,183)	-0,287 (0,190)
Constante	-1,956*** (0,341)	1,781*** (0,320)	-1,680*** (0,348)	2,081*** (0,356)	-1,832*** (0,310)	1,335*** (0,297)	-2,075*** (0,357)	0,903*** (0,324)	-4,247*** (0,530)	1,188*** (0,433)	-1,542*** (0,346)	1,123*** (0,336)	-2,810*** (0,431)	2,503*** (0,524)
Test de nullité jointe des coeff. des variables de collecte	155,97***		137,07***		109,48***		126,64***		132,16***		230,01***		113,56***	
Observations	948	1329	956	1345	1810	2103	751	877	834	1010	982	1317	1323	1435
Pseudo R2	0,385	0,173	0,406	0,146	0,588	0,156	0,369	0,225	0,591	0,251	0,358	0,234	0,503	0,255

Champ : Dépendants de 10 ans et plus, déclarant travailler dans une entreprise privée

Contrôles : Statut migratoire et variables de collecte

Données non pondérées

Les tranches de salaires proposées correspondent exactement à des multiples du salaire minimum en vigueur pour les villes de Niamey, Dakar et Bamako (cf les questionnaires nationaux).

Tableau 7 : Modèles $E1_i$ et $E2_i$: Enregistrement statistique ou fiscal (y_i^1) des travailleurs indépendants, sans prise en compte de la sélection, et mécanisme de réponse à la question de l'enregistrement (a_i^1) :

MI1 & MI2	Niamey		Ouagadougou		Dakar		Bamako		Cotonou		Lomé		Abidjan	
	y_i^1	a_i^1	y_i^1	a_i^1	y_i^1	a_i^1	y_i^1	a_i^1	y_i^1	a_i^1	y_i^1	a_i^1	y_i^1	a_i^1
contrat de travail écrit	0,207 (0,290)	-0,464 (0,301)	0,192 (0,267)	-0,166 (0,383)	-0,211 (0,227)	0,013 (0,237)	0,440** (0,181)	-0,338 (0,324)	1,404*** (0,402)		0,615*** (0,185)	-0,173 (0,272)	0,460 (0,281)	0,183 (0,447)
femme	-1,253*** (0,138)	0,432*** (0,149)	-0,639*** (0,104)	0,645*** (0,123)	-0,503*** (0,123)	0,255** (0,118)	-1,074*** (0,135)	0,601*** (0,143)	-0,140 (0,150)	0,426* (0,228)	-0,319*** (0,109)	0,876*** (0,143)	-0,457*** (0,143)	1,100*** (0,299)
âge	0,072*** (0,020)	-0,019 (0,024)	0,084*** (0,021)	-0,022 (0,022)	-0,004 (0,018)	-0,010 (0,020)	0,058*** (0,019)	-0,041* (0,024)	0,072** (0,032)	0,036 (0,037)	0,016 (0,019)	0,047* (0,025)	-0,015 (0,023)	0,025 (0,042)
âge au carré	-0,001*** (0,000)	0,000 (0,000)	-0,001*** (0,000)	0,000 (0,000)	0,000 (0,000)	0,000 (0,000)	-0,001*** (0,000)	0,001* (0,000)	-0,001** (0,000)	-0,000 (0,000)	-0,000 (0,000)	-0,000 (0,000)	0,000 (0,000)	-0,000 (0,000)
Au moins niveau primaire	0,380*** (0,113)	0,013 (0,183)	0,407*** (0,099)	-0,071 (0,131)	0,421*** (0,099)	-0,042 (0,127)	0,370*** (0,100)	-0,108 (0,142)	0,727*** (0,137)	-0,331* (0,186)	0,270*** (0,091)	-0,001 (0,136)	0,452*** (0,120)	0,130 (0,270)
Ent. de 6 personnes et +	1,054*** (0,179)	-0,929*** (0,186)	0,194 (0,128)	-0,672*** (0,134)	0,852*** (0,130)	-0,908*** (0,154)	0,679*** (0,145)	-0,810*** (0,226)	1,100*** (0,154)	-1,214*** (0,222)	1,479*** (0,128)	-1,008*** (0,177)	0,483** (0,193)	
entre 1 et 2 salaires minimaux	-0,353** (0,147)	0,162 (0,168)	0,395*** (0,113)	-0,220 (0,140)	-0,241* (0,127)	0,254** (0,117)	0,264* (0,135)	0,105 (0,146)	-0,064 (0,185)	-0,108 (0,216)	0,165 (0,105)	-0,160 (0,141)	-0,134 (0,173)	0,011 (0,273)
entre 2 et 4 salaires minimaux	0,268** (0,120)	0,157 (0,159)	0,694*** (0,110)	-0,125 (0,146)	0,399*** (0,117)	0,208 (0,153)	0,630*** (0,124)	0,407** (0,176)	0,396** (0,155)	-0,120 (0,209)	0,387*** (0,112)	0,233 (0,169)	0,194 (0,140)	-0,121 (0,274)
plus de 4 salaires minimaux	0,939*** (0,213)		1,955*** (0,247)	-0,012 (0,114)	1,423*** (0,211)		1,269*** (0,175)	0,621 (0,387)	0,926*** (0,224)	-0,123 (0,393)	1,494*** (0,209)	0,119 (0,146)	1,228*** (0,206)	-0,414 (0,082)
5 ans et plus d'expérience	-0,051 (0,111)	-0,174 (0,146)	0,150 (0,095)	0,424 (0,327)	-0,129 (0,104)	-0,227** (0,104)	0,091 (0,108)	0,033 (0,134)	-0,004 (0,132)	-0,216 (0,195)	-0,114 (0,089)	-0,193 (0,294)	0,114 (0,125)	-0,082 (0,260)
Cadres, ingénieurs et intellectuels	-0,387 (0,240)	-0,139 (0,314)	-0,390* (0,205)	0,480 (0,372)	0,181 (0,166)	0,233 (0,241)	-0,323* (0,190)		1,310*** (0,201)	0,037 (0,343)	0,156 (0,226)	0,442 (0,283)	0,449* (0,231)	
Prof. intermédiaires et administratives	-0,823** (0,392)	-0,261 (0,366)	0,034 (0,274)	0,473 (0,393)	0,041 (0,184)	0,287 (0,251)	-0,212 (0,212)	0,043 (0,306)	0,113 (0,216)	0,715* (0,431)	0,173 (0,164)	0,453 (0,375)	0,349 (0,223)	0,272 (0,480)
Agriculteurs		-0,225 (0,225)		0,265** (0,133)	-0,327** (0,146)	0,525* (0,303)	-0,632*** (0,132)	0,061 (0,346)	-0,094 (0,176)		-0,133 (0,116)	0,520*** (0,174)	0,115 (0,166)	-0,190 (0,457)
Artisans	-0,335*** (0,124)	0,047 (0,160)	-0,349*** (0,109)	0,472* (0,242)	-0,049 (0,280)	0,469*** (0,166)	-0,296 (0,198)	0,263 (0,169)	0,137 (0,304)	0,556* (0,285)	-0,150 (0,233)	0,040 (0,307)	-0,437 (0,415)	0,783** (0,365)
Ouvriers et manœuvres	-0,512** (0,233)	0,537* (0,302)	-0,617** (0,280)	0,463 (0,422)	-0,923*** (0,331)	0,032 (0,209)	-0,866** (0,372)	-0,114 (0,213)	-0,172 (0,477)	-0,074 (0,359)		0,346 (0,446)	-0,007 (0,252)	
Constante	-2,531*** (0,378)	1,684*** (0,471)	-2,969*** (0,395)	2,238*** (0,423)	-1,744*** (0,374)	2,131*** (0,398)	-2,762*** (0,400)	1,688*** (0,429)	-4,280*** (0,690)	0,699 (0,701)	-2,138*** (0,374)	0,325 (0,449)	-1,829*** (0,439)	0,612 (0,802)
Test de nullité jointe des coeff. des variables de collecte		92,24***		61,11***		121,91***		116,10***		30,52***		86,32***		52,99***
Observations	1754	1899	2200	2340	2593	2662	2617	2602	2962	2928	2574	2710	2085	1882
Pseudo R2	0,261	0,251	0,224	0,183	0,226	0,193	0,299	0,254	0,377	0,287	0,264	0,277	0,245	0,376

Champ : Indépendants de 10 ans et plus, déclarant travailler dans une entreprise privée. Contrôles : Statut migratoire et variables de collecte, Données non pondérées
 Les tranches de salaires proposées correspondent exactement à des multiples du salaire minimum en vigueur pour les villes de Niamey, Dakar et Bamako (cf les questionnaires nationaux).

Tableau 8 : Modèles CI_I et $C2_I$: Comptabilité formelle (y_i^2) des travailleurs indépendants, sans prise en compte de la sélection, et mécanisme de réponse à la question de comptabilité (a_i^2)

MD4 & MD5	Niamey		Ouagadougou		Dakar		Bamako		Cotonou		Lomé	
	y_i^2	a_i^2	y_i^2	a_i^2	y_i^2	a_i^2	y_i^2	a_i^2	y_i^2	a_i^2	y_i^2	a_i^2
contrat de travail écrit	0,656** (0,317)	-0,507 (0,322)	0,141 (0,314)	0,080 (0,326)	-0,190 (0,217)		0,384* (0,213)	0,087 (0,220)	0,534 (0,406)	-0,809* (0,478)	0,891** (0,426)	-1,736*** (0,206)
femme	-0,292*** (0,111)	0,132 (0,130)	-0,239 (0,159)	-0,006 (0,146)	-0,242*** (0,094)	0,572* (0,344)	-0,700*** (0,151)	-0,108 (0,082)	-0,007 (0,078)	0,168 (0,213)	-0,452*** (0,149)	1,111*** (0,107)
âge	0,093*** (0,025)	0,041** (0,019)	0,090*** (0,033)	-0,009 (0,022)	0,040** (0,017)	-0,058 (0,058)	0,067*** (0,024)	0,052*** (0,011)	0,022 (0,014)	0,017 (0,031)	0,011 (0,027)	0,043* (0,025)
âge au carré	-0,001*** (0,000)	-0,000 (0,000)	-0,001*** (0,000)	0,000 (0,000)	-0,000** (0,000)	0,001 (0,001)	-0,001** (0,000)	-0,000*** (0,000)	-0,000 (0,000)	0,000 (0,000)	-0,000 (0,000)	-0,000 (0,000)
Au moins niveau primaire	0,538*** (0,119)	-0,193 (0,139)	0,639*** (0,131)	0,021 (0,156)	0,201** (0,086)	0,232 (0,186)	0,373*** (0,120)	-0,131 (0,080)	0,465*** (0,062)	-0,217 (0,155)	0,329** (0,139)	-0,090 (0,087)
Ent. de 6 personnes et +	0,841*** (0,173)	-0,700*** (0,184)	0,178 (0,155)	-0,872*** (0,162)	0,393*** (0,119)	-0,647** (0,253)	0,931*** (0,170)	-0,530*** (0,137)	0,554*** (0,113)	-0,508** (0,238)	0,459** (0,181)	-1,097*** (0,146)
entre 1 et 2 salaires minimaux	0,055 (0,138)	-0,014 (0,154)	0,356* (0,183)	0,097 (0,187)	0,096 (0,097)	0,577 (0,365)	-0,172 (0,188)	0,057 (0,085)	0,143* (0,075)	0,217 (0,182)	0,033 (0,174)	-0,106 (0,090)
entre 2 et 4 salaires minimaux	0,341** (0,133)	0,308* (0,160)	0,902*** (0,155)	0,285 (0,234)	0,530*** (0,101)	0,015 (0,340)	0,477*** (0,140)	0,219** (0,105)	0,445*** (0,078)	0,423* (0,244)	0,302* (0,161)	0,476*** (0,126)
plus de 4 salaires minimaux	0,768*** (0,249)	0,841* (0,473)	1,679*** (0,238)	0,643* (0,388)	1,281*** (0,202)		1,116*** (0,187)	0,290 (0,240)	0,534*** (0,158)	0,129 (0,311)	1,321*** (0,246)	0,991*** (0,370)
5 ans et plus d'expérience	-0,100 (0,110)	-0,361*** (0,135)	0,146 (0,142)	0,152 (0,141)	-0,040 (0,086)	0,555** (0,251)	0,166 (0,135)	0,161** (0,078)	0,005 (0,064)	0,184 (0,155)	-0,037 (0,140)	0,225** (0,099)
Cadres, ingénieurs et intellectuels	0,128 (0,259)	-0,417 (0,304)	-0,009 (0,241)	0,199 (0,510)	0,246* (0,141)		-0,030 (0,213)	-0,062 (0,222)	0,438** (0,171)	-0,313 (0,342)	0,604*** (0,225)	0,049 (0,252)
Prof. intermédiaires et administratives	-0,247 (0,388)	-0,401 (0,321)	0,497* (0,286)	-1,382*** (0,271)	-0,086 (0,171)	-0,825*** (0,286)	-0,007 (0,235)	-0,054 (0,209)	-0,360** (0,144)	0,021 (0,341)	-0,332 (0,248)	-0,249 (0,170)
Agriculteurs		0,019 (0,220)			-0,585** (0,263)		-0,377 (0,332)	-0,233 (0,199)	-0,170 (0,230)	-0,179 (0,456)	-0,231 (0,426)	0,609** (0,307)
Artisans	-0,445*** (0,152)	0,085 (0,154)	-0,074 (0,154)	0,010 (0,172)	-0,451*** (0,121)	-0,060 (0,266)	-0,898*** (0,173)	0,051 (0,096)	-0,172** (0,085)	0,339 (0,226)	-0,403** (0,174)	0,879*** (0,132)
Ouvriers et manœuvres	-0,148 (0,216)	-0,102 (0,218)	0,403 (0,272)	-0,056 (0,321)	-0,423* (0,256)	-0,286 (0,408)	-0,017 (0,266)	-0,502*** (0,120)	-0,319* (0,173)	-0,126 (0,266)	0,191 (0,472)	-1,307*** (0,181)
Constante	-3,095*** (0,484)	1,170*** (0,401)	-4,337*** (0,629)	2,291*** (0,429)	-2,268*** (0,330)	2,991*** (0,902)	-3,207*** (0,479)	0,453** (0,229)	-1,726*** (0,275)	1,496*** (0,578)	-2,180*** (0,572)	-0,175 (0,416)
Test de nullité jointe des coeff. des variables de collecte		129,53***		11,35***		0,64		97,75***		0,06		84,52***
Observations	1744	1947	2251	2302	2696	2388	2350	2686	2957	2993	2393	2710
Pseudo R2	0,165	0,230	0,276	0,130	0,116	0,188	0,301	0,132	0,0866	0,0796	0,230	0,377

Champ : Dépendants de 10 ans et plus, déclarant travailler dans une entreprise privée Contrôles : Statut migratoire et variables de collecte Données non pondérées
Les tranches de salaires proposées correspondent exactement à des multiples du salaire minimum en vigueur pour les villes de Niamey, Dakar et Bamako (cf les questionnaires nationaux).

Tableau 9 : Modèle $E3_D$: ESF des travailleurs dépendants, avec prise en compte de la sélection.

MD3H	Niamey		Ouagadougou		Dakar		Bamako		Cotonou		Lomé		Abidjan	
	y_i^1	a_i^1	y_i^1	a_i^1	y_i^1	a_i^1	y_i^1	a_i^1	y_i^1	a_i^1	y_i^1	a_i^1	y_i^1	a_i^1
contrat de travail écrit	0,532*** (0,184)	0,478*** (0,157)	0,264 (0,166)	0,351** (0,137)	0,517*** (0,120)	0,259** (0,117)	0,890*** (0,156)	0,117 (0,180)	0,202 (0,162)	0,263* (0,138)	0,371*** (0,120)	-0,041 (0,100)	0,472*** (0,113)	0,336* (0,174)
bulletin de paie	0,530*** (0,204)	0,343** (0,164)	1,031*** (0,245)	0,293* (0,160)	1,509*** (0,127)	0,670*** (0,125)	0,698*** (0,160)	0,550*** (0,202)	1,236*** (0,234)	0,759*** (0,169)	0,642*** (0,174)	0,453** (0,178)	0,751*** (0,128)	0,208 (0,180)
femme	-0,261** (0,122)	0,111 (0,099)	-0,333*** (0,121)	0,124 (0,100)	-0,028 (0,109)	-0,156 (0,097)	-0,534*** (0,149)	0,213 (0,157)	-0,213 (0,153)	0,261** (0,125)	-0,027 (0,117)	0,047 (0,100)	0,002 (0,120)	0,067 (0,151)
âge	0,077*** (0,024)	0,004 (0,019)	0,039* (0,021)	-0,020 (0,018)	0,022 (0,020)	-0,007 (0,018)	0,092*** (0,021)	0,038** (0,018)	0,167*** (0,029)	-0,001 (0,027)	0,012 (0,025)	0,053** (0,021)	0,054** (0,027)	-0,037 (0,032)
âge au carré	-0,001*** (0,000)	-0,000 (0,000)	-0,001* (0,000)	0,000 (0,000)	-0,000 (0,000)	0,000 (0,000)	-0,001*** (0,000)	-0,000** (0,000)	-0,002*** (0,000)	0,000 (0,000)	-0,000 (0,000)	-0,001** (0,000)	-0,001 (0,000)	0,001 (0,000)
Au moins niveau primaire	0,113 (0,121)	0,105 (0,100)	0,335*** (0,105)	0,085 (0,087)	0,298*** (0,094)	0,117 (0,092)	-0,198 (0,126)	0,035 (0,132)	0,151 (0,148)	-0,135 (0,131)	0,365*** (0,110)	-0,010 (0,103)	0,262*** (0,100)	0,125 (0,133)
Ent, de 6 personnes et +	0,940*** (0,134)	-0,410*** (0,095)	1,041*** (0,105)	-0,415*** (0,090)	0,940*** (0,155)	-0,122 (0,087)	0,538*** (0,127)	-0,333** (0,132)	0,878*** (0,132)	-0,472*** (0,128)	0,893*** (0,113)	-0,437*** (0,094)	1,179*** (0,106)	0,247* (0,142)
entre 1 et 2 salaires minimaux	0,338** (0,140)	0,203 (0,125)	0,357*** (0,136)	-0,010 (0,112)	0,357*** (0,104)	0,142 (0,101)	0,040 (0,146)	-0,094 (0,147)	0,245 (0,150)	-0,026 (0,125)	0,620*** (0,155)	-0,049 (0,141)	0,620*** (0,134)	-0,191 (0,161)
entre 2 et 4 salaires minimaux	0,197 (0,193)	0,018 (0,146)	0,552*** (0,167)	0,055 (0,142)	0,287* (0,169)	0,344** (0,163)	0,380** (0,159)	0,144 (0,173)	0,662*** (0,200)	0,222 (0,176)	0,820*** (0,166)	0,419*** (0,162)	0,478*** (0,132)	-0,132 (0,168)
plus de 4 salaires minimaux	-0,541 (0,378)	0,177 (0,323)	-0,887** (0,413)	0,023 (0,390)	-0,238 (0,417)	-0,658** (0,301)	0,915** (0,405)	6,749*** (0,564)	0,328 (0,562)	0,859 (0,819)	0,902** (0,407)	0,811** (0,353)	1,050*** (0,257)	0,440 (0,464)
5 ans et plus d'expérience	-0,113 (0,121)	0,148 (0,098)	-0,055 (0,121)	0,302*** (0,100)	-0,206** (0,090)	-0,033 (0,078)	-0,275** (0,128)	0,026 (0,132)	-0,311** (0,152)	0,187 (0,139)	0,061 (0,135)	-0,069 (0,126)	-0,123 (0,110)	0,204 (0,140)
Cadres, ingénieurs et intellectuels	0,215 (0,257)	0,170 (0,212)	0,227 (0,298)	0,324 (0,285)	0,116 (0,227)	0,415* (0,216)	-0,455* (0,261)	0,226 (0,331)	0,458* (0,278)	0,382 (0,270)	0,631** (0,302)	-0,402* (0,217)	0,171 (0,216)	-0,468* (0,263)
Prof, intermédiaires et administratives	0,375* (0,199)	-0,059 (0,157)	0,286 (0,179)	0,130 (0,142)	0,389*** (0,135)	0,200 (0,134)	0,124 (0,171)	0,045 (0,182)	0,476** (0,186)	0,010 (0,167)	0,314* (0,171)	-0,216 (0,147)	0,479*** (0,133)	0,072 (0,195)
Agriculteurs	-2,001*** (0,280)	-0,303* (0,168)	-1,309*** (0,436)	0,258 (0,434)	-1,077*** (0,305)	0,181 (0,253)	-1,158*** (0,259)	7,136*** (0,963)	-0,555 (0,763)	9,114 (-)	-0,922*** (0,267)	0,077 (0,278)	-0,316 (0,234)	0,452 (0,452)
Artisans	-0,274** (0,138)	-0,327*** (0,104)	-0,276** (0,115)	0,003 (0,096)	-0,690*** (0,133)	-0,044 (0,101)	-0,517*** (0,157)	0,086 (0,145)	-0,505*** (0,168)	0,094 (0,142)	-0,254* (0,132)	-0,161 (0,114)	-0,114 (0,129)	-0,475*** (0,164)
Ouvriers et manœuvres	-0,587*** (0,189)	-0,181 (0,167)	0,051 (0,192)	-0,075 (0,144)	0,399** (0,182)	0,088 (0,173)	-0,206 (0,174)	0,042 (0,196)	0,319 (0,219)	-0,170 (0,204)	-0,137 (0,167)	0,145 (0,182)	0,548*** (0,183)	-0,265 (0,193)
Constante	-2,004*** (0,342)	1,764*** (0,323)	-1,259*** (0,352)	1,992*** (0,350)	-1,891*** (0,303)	1,347*** (0,295)	-2,180*** (0,353)	0,936*** (0,326)	-4,134*** (0,595)	1,202*** (0,435)	-1,591*** (0,353)	1,120*** (0,335)	-2,789*** (0,429)	2,578*** (0,535)
Test de nullité jointe des coeff, des variables de collecte	156,25***		134,14***		100,01***		125,30***		131,98***		230,01***		111,63***	
Corrélation des résidus	0,162 (0,321)		-0,868*** (0,314)		0,441 (0,944)		0,427 (0,306)		-0,346 (0,576)		0,109 (0,192)		0,461 (0,330)	
Test de la sélection	0,613		0,006		0,641		0,163		0,548		0,570		0,162	
Observations	1329		1345		2103		921		1014		1317		1435	
Observations censurées	381		389		293		170		180		335		112	

Champ : Dépendants de 10 ans et plus, déclarant travailler dans une entreprise privée

Contrôles : Statut migratoire et variables de collecte Données non pondérées

Les tranches de salaires proposées correspondent exactement à des multiples du salaire minimum en vigueur pour les villes de Niamey, Dakar et Bamako (cf les questionnaires nationaux).

Tableau 10 : Modèle $E3_I$: ESF des travailleurs indépendants, avec prise en compte de la sélection.

MI3H	Niamey		Ouagadougou		Dakar		Bamako		Cotonou		Lomé		Abidjan	
	y_i^1	a_i^1	y_i^1	a_i^1	y_i^1	a_i^1	y_i^1	a_i^1	y_i^1	a_i^1	y_i^1	a_i^1	y_i^1	a_i^1
contrat de travail écrit	0,293 (0,307)	-0,470 (0,305)	0,198 (0,260)	-0,272 (0,378)	-0,211 (0,225)	0,018 (0,238)	0,408** (0,181)	-0,409 (0,302)	1,425*** (0,393)	5,289*** (0,327)	0,619*** (0,185)	-0,173 (0,272)	0,458 (0,281)	0,207 (0,470)
femme	-1,282*** (0,136)	0,441*** (0,148)	-0,602*** (0,102)	0,675*** (0,124)	-0,511*** (0,124)	0,255** (0,118)	-1,051*** (0,134)	0,618*** (0,143)	-0,141 (0,149)	0,459** (0,233)	-0,324*** (0,114)	0,876*** (0,143)	-0,471*** (0,151)	1,094*** (0,300)
âge	0,073*** (0,020)	-0,023 (0,025)	0,082*** (0,020)	-0,025 (0,022)	-0,003 (0,018)	-0,010 (0,020)	0,056*** (0,019)	-0,039* (0,023)	0,070** (0,031)	0,042 (0,039)	0,016 (0,019)	0,047* (0,025)	-0,015 (0,023)	0,027 (0,042)
âge au carré	-0,001*** (0,000)	0,000 (0,000)	-0,001*** (0,000)	0,000 (0,000)	0,000 (0,000)	0,000 (0,000)	-0,001*** (0,000)	0,000* (0,000)	-0,001* (0,000)	-0,000 (0,000)	-0,000 (0,000)	-0,000 (0,000)	0,000 (0,000)	-0,000 (0,000)
Au moins niveau primaire	0,378*** (0,112)	0,012 (0,184)	0,400*** (0,097)	-0,117 (0,129)	0,421*** (0,099)	-0,039 (0,128)	0,362*** (0,099)	-0,123 (0,140)	0,712*** (0,135)	-0,355* (0,187)	0,270*** (0,091)	-0,003 (0,138)	0,448*** (0,122)	0,125 (0,272)
Ent, de 6 personnes et +	1,154*** (0,194)	-0,929*** (0,185)	0,057 (0,121)	-0,637*** (0,133)	0,881*** (0,143)	-0,912*** (0,154)	0,617*** (0,145)	-0,832*** (0,227)	0,999*** (0,153)	-1,232*** (0,230)	1,487*** (0,137)	-1,009*** (0,177)	0,474** (0,196)	5,028*** (--)
entre 1 et 2 salaires minimaux	-0,373** (0,145)	0,150 (0,165)	0,371*** (0,111)	-0,204 (0,140)	-0,250** (0,127)	0,252** (0,117)	0,269** (0,134)	0,121 (0,147)	-0,060 (0,182)	-0,113 (0,215)	0,166 (0,105)	-0,159 (0,142)	-0,132 (0,172)	0,012 (0,271)
entre 2 et 4 salaires minimaux	0,229* (0,131)	0,151 (0,155)	0,668*** (0,108)	-0,102 (0,146)	0,392*** (0,117)	0,203 (0,154)	0,643*** (0,123)	0,440** (0,176)	0,391** (0,154)	-0,105 (0,217)	0,386*** (0,113)	0,233 (0,169)	0,200 (0,140)	-0,109 (0,289)
plus de 4 salaires minimaux	0,843*** (0,243)	8,737 (--)	1,942*** (0,237)	0,407 (0,460)	1,390*** (0,223)	6,932*** (--)	1,260*** (0,176)	0,685* (0,414)	0,897*** (0,223)	-0,009 (0,417)	1,492*** (0,209)	0,348 (0,448)	1,234*** (0,206)	-0,412 (0,547)
5 ans et plus d'expérience	-0,031 (0,113)	-0,170 (0,147)	0,138 (0,093)	-0,014 (0,114)	-0,121 (0,105)	-0,226** (0,104)	0,090 (0,107)	-0,001 (0,133)	-0,007 (0,130)	-0,216 (0,195)	-0,115 (0,089)	0,118 (0,146)	0,112 (0,125)	-0,091 (0,265)
Cadres, ingénieurs et intellectuels	-0,399* (0,240)	-0,199 (0,320)	-0,302 (0,199)	0,371 (0,337)	0,175 (0,168)	0,237 (0,242)	-0,280 (0,187)	31,315 (--)	1,256*** (0,205)	0,042 (0,342)	0,157 (0,226)	-0,196 (0,296)	0,434* (0,237)	4,597*** (0,521)
Prof, intermédiaires et administratives	-0,809** (0,392)	-0,251 (0,359)	0,060 (0,263)	0,637* (0,386)	0,029 (0,186)	0,283 (0,251)	-0,196 (0,208)	0,034 (0,304)	0,136 (0,212)	0,758* (0,433)	0,171 (0,164)	0,445 (0,285)	0,340 (0,224)	0,243 (0,505)
Agriculteurs	-3,902,384 (--)	-0,289 (0,225)	-7,682 (--)	0,479 (0,399)	-0,928*** (0,330)	0,533* (0,305)	-0,841** (0,366)	0,104 (0,346)	-0,140 (0,468)	6,379*** (0,811)	-6,484*** (--)	0,453 (0,375)	-0,010 (0,252)	-0,201 (0,469)
Artisans	-0,346*** (0,121)	0,032 (0,160)	-0,316*** (0,106)	0,274** (0,133)	-0,341** (0,151)	0,472*** (0,166)	-0,620*** (0,130)	0,303* (0,174)	-0,077 (0,172)	0,598** (0,294)	-0,136 (0,118)	0,518*** (0,175)	0,103 (0,171)	0,767** (0,362)
Ouvriers et manœuvres	-0,561** (0,241)	0,515* (0,294)	-0,581** (0,277)	0,483** (0,245)	-0,054 (0,279)	0,030 (0,208)	-0,303 (0,195)	-0,164 (0,202)	0,135 (0,300)	-0,037 (0,365)	-0,149 (0,233)	0,044 (0,310)	-0,453 (0,424)	6,843 (--)
Constante	-2,481*** (0,384)	1,741*** (0,493)	-2,969*** (0,390)	2,317*** (0,425)	-1,728*** (0,375)	2,130*** (0,397)	-2,758*** (0,395)	1,682*** (0,418)	-4,230*** (0,677)	0,562 (0,728)	-2,126*** (0,387)	0,325 (0,449)	-1,796*** (0,448)	0,591 (0,809)
Test de nullité jointe des coeff, des variables de collecte	91,20***		64,31***		122,05***		116,86***		30,24***		86,38***		53,60***	
Corrélation des résidus	-0,514 (0,601)		6,193 (49,936)		-0,178 (0,336)		0,895** (0,365)		4,829 (17,711)		-0,047 (0,263)		-0,263 (0,719)	
Test de la sélection	0,392		0,901		0,597		6,03**		0,785		0,859		0,714	
Observations	1947		2340		2707		2686		2993		2710		2108	
Observations censurées	86		103		114		69		31		77		23	

Champ : Indépendants de 10 ans et plus, déclarant travailler dans une entreprise privée Contrôles : Statut migratoire et variables de collecte Données non pondérées
Les tranches de salaires proposées correspondent exactement à des multiples du salaire minimum en vigueur pour les villes de Niamey, Dakar et Bamako (cf les questionnaires nationaux).

Tableau 11 : Modèle $C3_I$: CF des travailleurs indépendants, avec prise en compte de la sélection.

MI6H	Bamako		Lomé	
	y_i^2	a_i^2	y_i^2	a_i^2
contrat de travail écrit	0,375* (0,209)	0,101 (0,223)	1,378*** (0,406)	-1,741*** (0,204)
Femme	-0,694*** (0,149)	-0,109 (0,082)	-0,707*** (0,155)	1,117*** (0,106)
Age	0,071*** (0,024)	0,052*** (0,011)	-0,014 (0,026)	0,041 (0,025)
âge au carré	-0,001*** (0,000)	-0,000*** (0,000)	0,000 (0,000)	-0,000 (0,000)
Au moins niveau primaire	0,353*** (0,120)	-0,130 (0,080)	0,321** (0,129)	-0,091 (0,086)
Ent. de 6 personnes et +	0,859*** (0,172)	-0,531*** (0,137)	0,682*** (0,182)	-1,122*** (0,145)
entre 1 et 2 salaires minimaux	-0,161 (0,185)	0,059 (0,085)	0,056 (0,152)	-0,096 (0,090)
entre 2 et 4 salaires minimaux	0,490*** (0,139)	0,220** (0,105)	0,123 (0,153)	0,495*** (0,127)
plus de 4 salaires minimaux	1,126*** (0,185)	0,301 (0,241)	1,004*** (0,270)	0,876** (0,341)
5 ans et plus d'expérience	0,178 (0,133)	0,162** (0,078)	-0,090 (0,127)	0,227** (0,099)
Cadres, ingénieurs et intellectuels	-0,022 (0,209)	-0,076 (0,222)	0,574*** (0,220)	0,076 (0,248)
Prof. intermédiaires et administratives	-0,013 (0,232)	-0,056 (0,210)	-0,172 (0,234)	-0,239 (0,169)
Agriculteurs	-0,381 (0,329)	-0,226 (0,201)	-0,361 (0,409)	0,574* (0,300)
Artisans	-0,876*** (0,171)	0,048 (0,096)	-0,576*** (0,171)	0,912*** (0,134)
Ouvriers et manœuvres	-0,066 (0,264)	-0,502*** (0,120)	0,586 (0,431)	-1,300*** (0,180)
Constante	-3,353*** (0,472)	0,454** (0,229)	-1,130 (0,697)	-0,137 (0,415)
Test de nullité jointe des coeff. des variables de collecte		99,47***		91,97***
Corrélation des résidus		0,497* (0,283)		-0,884*** (0,313)
Test de la sélection		3,07*		7,99***
Observations		2686		2710
Observations censurées		336		317

Champ : Indépendants de 10 ans et plus, déclarant travailler dans une entreprise privée

Contrôles : Statut migratoire et variables de collecte

Données non pondérées

Les tranches de salaires proposées correspondent exactement à des multiples du salaire minimum en vigueur pour les villes de Niamey, Dakar et Bamako (cf les questionnaires nationaux).

Tableau 12 : Sensibilité des probabilités prédites à la spécification des modèles d'imputation (dans le texte)Tableau 13 : Modèles servant à l'imputation (dans le texte)

Tableau 14 : Robustesse de l'imputation probabiliste des réponses aux deux critères

	Taux moyen des 500 imputations (%)	Ecart-type des 500 imputations (%)	Intervalle à 95%	Taux minimum obtenu parmi les 500 imputations (%)	Taux maximum obtenu parmi les 500 imputations (%)	Taux choisi (%)
Niamey	70,886	0,105	[70,68 ; 71,10]	70,606	71,197	70,887
Ouagadougou	72,443	0,151	[72,14 ; 72,75]	71,932	72,837	72,443
Dakar	75,462	0,055	[75,35 ; 75,57]	75,317	75,614	75,462
Bamako	77,807	0,073	[77,66 ; 77,95]	77,570	78,019	77,807
Cotonou	80,235	0,091	[80,05 ; 80,42]	80,006	80,475	80,235
Lomé	80,444	0,100	[80,24 ; 80,64]	80,186	80,750	80,443
Abidjan	74,354	0,058	[74,24 ; 74,47]	74,133	74,509	74,354

Données pondérées
Ensemble des actifs occupés de 10 ans et plus

Tableau 15 : Taux de non-réponse avant et après imputation

	ESF		CF	
	Taux de non-réponse avant imputation	Taux de non-réponse après imputation	Taux de non-réponse avant imputation	Taux de non-réponse après imputation
Niamey	14,55	0,48	4,74	4,23
Ouagadougou	13,90	1,17	2,15	1,94
Dakar	9,20	1,30	0,47	0,47
Bamako	6,59	0,05	12,60	11,93
Cotonou	7,10	2,46	1,20	1,13
Lomé	10,34	0,22	11,65	9,82
Abidjan	4,86	1,63	0	0

Données non pondérées, Ensemble des actifs occupés de 10 ans et plus, déclarant travailler pour une entreprise privée

Tableau 16 : Répartition par secteurs formel et informel selon le programme de construction de la variable SI ayant été la base des calculs figurant dans Statéco n°99.

	Comptabilité	ESF=Oui		ESF=Non		non-réponse	
		<=20	> 20	<=20	> 20	<=20	> 20
Patrons et comptes propres	=Oui	Formel	Formel	Informel	Manquant	Informel	Manquant
	=Non	Informel	Formel	Informel	Manquant	Informel	Manquant
	non-réponse	Formel	Formel	Informel	Manquant	Informel	Manquant
Salariés, apprentis et aides familiaux		Formel	Formel	Informel	Manquant	Informel	Manquant

Tableau 17 : Sensibilité du taux d'emploi dans le secteur informel au traitement de la non-réponse (dans le texte)

VII. Références

- Amegashie, F., Brilleau, A., Coulibaly, S., Koriko, O., Ouédraogo, E., Roubaud, F., & Torelli, C. (2005). La conception et la mise en oeuvre des enquêtes 1-2-3 en UEMOA. Les enseignements méthodologiques. *Statéco*, 2005(99), 21-41.
- Backiny-Yetna, P., & Bardon, R. (1999). *Concepts et indicateurs du marché du travail et du secteur informel* (Série Méthode 2). Bamako: AFRISTAT.
- Brilleau, A., Roubaud, F., & Torelli, C. (2005). L'emploi, le chômage et les conditions d'activité, Enquête 1-2-3 Phase 1. *Statéco*, 2005(99), 43-64.
- Chavance, M. (2004). Handling Missing Items in Quality of Life Studies. *Communications in Statistics - Theory and Methods*, 33(6), 1371-1383.
- Diggle, P. J. (1989). Testing for random dropouts in repeated measurement data. *Biometrics*, 45(4), 1255-1258.
- Dubin, J. A., & Rivers, D. (1989). Selection Bias in Linear Regression, Logit and Probit Models. *Sociological Methods and Research*, 18(2-3), 360-390.
- Gautier E. (2006). *Éléments sur les mécanismes de sélection dans les enquêtes et sur la non-réponse non-ignorable*. Acte de congrès. Journées de Méthodologie Statistique 2005.
- Heckman, J. J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables, and a simple estimator for such models. *Annals of Economic and Social Measurement*, 1976(5), 475-492.
- Heitjan, D. F., & Rubin, D. B. (1991). Ignorability and coarse data. *The Annals of Statistics*, 19(4), 2244-2253.
- Henley, A., Arabsheibani, R. G., & Carneiro, F. G. (2006). On Defining and Measuring the Informal Sector. *IZA Discussion Papers*(2473).
- Little, R. J. A. (1988). Missing Data Adjustments in Large Surveys (with discussion). *Journal of Business and Economic Statistics*, 6(3), 287-297.
- Little, R. J. A., & Rubin, D. B. (2002). *Statistical Analysis with Missing Data, Second Edition*. New York: John Wiley & Sons.
- Park, T., & Davis, C. S. (1993). A Test of the Missing Data Mechanism for Repeated Categorical Data. *Biometrics*, 49(2), 631-638.
- Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(3), 581-592.
- Sales, A. E., Plomondon, M. E., Magid, D. J., Spertus, J. A., & Rumsfeld, J. S. (2004). Assessing response bias from missing quality of life data: The Heckman method. *Health and Quality of Life Outcomes*, 2(49), 1-10.