

Une Revue Des Approches Et Methodes D'évaluation D'impact Appliquees En Microfinance

Gnoudanfoly A. Soro

Enseignant-Chercheur-UFR des Sciences Économiques et de Gestion (UFR-SEG)

Université Félix Houphouët Boigny de Cocody-Abidjan-Côte d'Ivoire

Doi:10.19044/esj.2019.v15n13p272 [URL:http://dx.doi.org/10.19044/esj.2019.v15n13p272](http://dx.doi.org/10.19044/esj.2019.v15n13p272)

Résumé

Les débats sur la microfinance se concentrent le plus souvent sur son impact sur les bénéficiaires. L'évaluation de l'impact de la microfinance est devenue une préoccupation majeure ces dernières années, vu les doutes émis par de nombreux chercheurs sur ses effets réels. Cet article fournit une revue des approches et méthodes de l'évaluation de l'impact de la microfinance. La méthodologie adoptée consiste en une analyse descriptive et critique des principales méthodes tout en mettant en relief les avantages et les limites de chacune d'elle. Les résultats de cette revue révèlent qu'il n'existe pas de méthode clé en main pour tous les types d'évaluation. Le choix des méthodes dépend fortement du contexte et de la spécificité des biais à traiter. L'analyse relève également l'importance de ces évaluations dans le secteur de la microfinance et les problèmes du financement de ces études dans les pays à faibles revenus.

Mots Clés: Microfinance, Evaluation d'Impact, Approche non-expérimentale, Expérimentations aléatoires, Biais de sélection.

A Review of Impact Evaluation Approaches and Methods Applied in Microfinance

Gnoudanfoly A. Soro

Enseignant-Chercheur-UFR des Sciences Économiques et de Gestion (UFR-SEG)

Université Félix Houphouët Boigny de Cocody-Abidjan-Côte d'Ivoire

Abstract

Microfinance debates most often focus on its impact on beneficiaries. The assessment of the impact of microfinance has become a major concern in recent years, given the doubts expressed by many researchers about its real effects. This article provides a descriptive and critical review of approaches and methods for assessing the impact of microfinance. This paper analyzes the main methods of evaluation used in microfinance while highlighting the advantages and limitations of each of them. The results of this review reveal that there is no a ready-made method for all types of evaluation. The choice of methods strongly depends on the context and the specificity of the biases to be treated. The analysis also highlights the importance of these evaluations in the microfinance sector and the problem of financing these studies in low incomes countries.

Keywords: Microfinance, Impact Evaluation, Non-Experimental Approach, Randomized Evaluation, Selection Bias.

Introduction

Ces dernières décennies, face à la rareté relative des subventions et des sources de financement, le respect du « *double bottom line* »¹ est devenu une exigence des partenaires au développement. Il ne suffit plus de montrer la performance financière et la viabilité des institutions de microfinance (IMF) pour espérer obtenir des subventions, des dons et des prêts provenant des partenaires. De plus en plus, ceux-ci exigent que les performances financières soient couplées aux performances sociales. Dans ce contexte, les évaluations d'impact s'avèrent nécessaires pour les IMF, notamment celles dont une grande partie des financements dépendent des partenaires pour continuer leurs activités.

Les dérives de la microfinance tant décriées par bon nombre de travaux confortent l'idée de la conduite d'évaluations rigoureuses afin de

mieux cerner son impact sur les bénéficiaires (Fouillet et *al.*, 2007; Mersland et Storm, 2010, etc.). De plus, la crise survenue en 2011 en Inde et la forte financiarisation avec l'entrée en bourse de SKS² en 2010 a suscité beaucoup d'interrogation sur ce virage néolibéral de la microfinance vers la recherche du profit au détriment de la lutte contre la pauvreté. L'espoir mis dans ces institutions par les gouvernants des pays en développement et les organisations internationales mérite que des évaluations soient conduites afin d'apprécier leurs effets réels en matière d'inclusion financière et de réduction de la pauvreté. Ces évaluations sont également indispensables pour une meilleure orientation des activités de ces institutions d'une part, et d'autre part, pour des ajustements cohérents dans les politiques nationales de lutte contre la pauvreté. Mais, comment évaluer un programme de microfinance ? Quels sont les méthodes et outils d'évaluation couramment utilisés en microfinance ? Quels sont les avantages et les limites de ces méthodes. C'est à ces questions que nous souhaitons apporter des éléments, non pas de réponse, mais tout au moins d'éclairage. L'évaluation est une nécessité commune à tous les programmes de développement et particulièrement en microfinance au vu des objectifs et missions qui lui sont assignés. L'objectif de ce travail est d'apporter un éclairage sur les méthodes et les outils statistiques et économétriques utilisés dans les évaluations de la microfinance. Il s'agira d'une part de comprendre l'analyse de l'impact de la microfinance et d'autre part d'analyser les approches et méthodes d'évaluation et d'en tirer leurs avantages et leurs limites. Notre démarche consiste en une revue de littérature descriptive et critique des principales méthodes utilisées dans les évaluations de l'impact en microfinance. Nous supposons qu'il n'existe pas de méthodes clés en mains, même si certaines de ces méthodes se révèlent plus pertinentes que d'autres dans la prise en compte des probables biais qui pourraient apparaître dans le processus d'évaluation. Au-delà des coûts et du temps que nécessitent certaines méthodes, le contexte propre à chaque observation est déterminant dans le recours à chacune des méthodes pour réduire les biais spécifiques au problème de l'évaluation. Dans la suite de ce travail, nous analysons l'impact en microfinance (1), puis nous exposons les méthodes des approches quasi-expérimentale (2) et d'expérimentation aléatoire (3), incluant, leurs avantages et leurs limites, enfin nous concluons. Ce papier s'inscrit dans la suite des contributions à la compréhension des évaluations d'impact, en mettant l'accent sur le secteur de la microfinance et les difficultés de l'usage de ces méthodes et leur influence sur les résultats.

1. L'impact en microfinance

1.1. La nécessité de l'évaluation de la microfinance

L'évaluation de l'impact de la microfinance n'est pas une priorité pour les IMF dont l'objectif est la recherche des profits. La capacité des clients

à pouvoir rembourser ou à emprunter des prêts et l'accroissement du nombre de client constitue pour ces institutions une justification évidente de leur impact sur les bénéficiaires. Dans cette logique, la conduite d'évaluations apparaît injustifiée pour ces institutions, surtout quand celles-ci sont financièrement coûteuses. Quant aux institutions à but non lucratif dont les interventions mettent en priorité l'inclusion financière et la lutte contre la pauvreté, et dont les financements dépendent des donateurs, elles sont contraintes de justifier l'impact de leurs interventions sur le bien-être des bénéficiaires pour prétendre à de nouveaux financements. De fait, ces divergences de la nécessité de l'évaluation opposent deux logiques d'approche en microfinance: l'approche institutionnaliste qui met l'accent sur la rentabilité des IMF et l'approche *welfariste* qui met l'accent sur le bien-être des bénéficiaires. Mais, quoi qu'il en soit, l'évaluation de l'impact est justifiée pour diverses raisons quel que soit l'approche de l'IMF. En suivant Karlan et Goldberg (2007), nous énumérons quatre bonnes raisons qui justifient l'intérêt de conduire des évaluations d'impact pour l'ensemble des acteurs du secteur. Pour les donateurs, l'évaluation des programmes s'avère nécessaire pour une allocation efficace et efficiente des ressources qui se raréfient dans le contexte de crises dans les pays donateurs. En outre, l'évaluation d'impact peut s'apparenter à une étude de marché ou de la clientèle. De telles études permettent à une IMF de concevoir de meilleurs produits adaptés aux besoins des clients, d'élargir la gamme de ses services et de fidéliser sa clientèle. Il y a donc un intérêt pour les IMF lucratives de conduire de telles études qui permettent d'améliorer leur rentabilité. Un second intérêt de l'évaluation est qu'il permet d'apprécier si des subventions indirectes sous la forme de prêts à taux réduit ou de l'assistance technique gratuite accordés par des gouvernements à des IMF ou à des ONG sont justifiés en termes d'impact sur les bénéficiaires par rapport à d'autres usages alternatifs de ces fonds publics. L'évaluation est également un moyen d'apprentissage des exemples de succès et des bonnes pratiques, de même que les échecs de certains programmes. Enfin, une évaluation peut permettre d'apprécier l'impact d'un programme de microfinance selon qu'il soit géré par une IMF ou une agence gouvernementale ou qu'il migre vers un programme à but lucratif. Il s'agit d'analyser quel type de gouvernance est mieux adapté pour générer des effets positifs significatifs sur les bénéficiaires. Comme on peut le constater, il existe différentes raisons qui peuvent motiver une évaluation d'impact en microfinance. Si de plus en plus leur importance est reconnue par tous, la question des coûts et des résultats parfois sujets de discussions sont invoqués. Ce qui découlerait des approches et des méthodes utilisées dans ces évaluations.

Après cette présentation des motivations des évaluations en microfinance, la section suivante appréhende les types d'évaluation conduits en microfinance.

1.2. Types d'évaluation en microfinance

Plusieurs types de variables peuvent être identifiés et faire l'objet d'évaluation en microfinance. Il faut pour cela que les variables soient définies avec précision et soient mesurables (Hashemi et *al.*, 1996; Hulme, 2000). Toutefois, il est possible comme l'on fait Karlan et Goldberg (2007) de faire une typologie d'ensemble des évaluations en microfinance. Ces auteurs proposent trois grands ensembles d'évaluations: les évaluations de programme, les évaluations de processus ou de produits et les évaluations de politiques.

- ***L'évaluation des programmes***

De façon générale, l'évaluation d'impact fait référence à l'évaluation des programmes de microfinance. Ces évaluations portent sur l'évaluation des prêts ou subventions octroyés à un programme de microfinance dans une région donnée. L'évaluation peut consister à analyser l'efficacité des prêts octroyés par l'IMF sur le bien-être de leurs clients. Elle doit prendre en compte les modalités et les conditions de prêt, la qualité des produits, etc. Globalement, l'évaluation de programme nécessite la prise en compte de plusieurs composantes (éducation, constitution d'un capital social, santé, assurance, etc.) afin de mieux cerner les effets directs et indirects.

- ***L'évaluation de produit ou de processus***

L'évaluation de processus ou de produits consiste à évaluer l'efficacité relative d'une IMF particulière dans la mise en œuvre d'un produit par rapport à un autre, ou d'un processus par rapport à un autre. Il peut s'agir d'analyser l'impact des prêts sur l'éducation des enfants des bénéficiaires, sur la santé des ménages bénéficiaires, de l'impact des prêts de groupe par rapport aux prêts individuels, etc. Il peut s'agir aussi d'analyser l'impact de nouveaux produits que l'IMF propose à des clients volontaires ou choisis. Elle peut porter également sur un changement dans une branche de l'institution. Ces types d'analyse sont risqués aussi bien pour les prêteurs que pour les clients et les déductions sur les avantages des changements peuvent être trompeurs, car il est difficile d'établir si le changement ou l'innovation va conduire à une amélioration de la situation des clients ou de la rentabilité de l'institution. Pour Karlan et Goldberg (2007), la méthode des randomisations est plus appropriée à ce type d'évaluation, car elles sont moins risquées pour les IMF et les leçons apprises de ces liens de causalité peuvent être diffusées à d'autres IMF et utilisées dans d'autres programmes.

- ***L'évaluation de politiques***

Elle concerne l'évaluation de politiques de réglementation ou de réformes dans le secteur de la microfinance. Le plafonnement du crédit, la réglementation de l'épargne et des taux d'intérêt, la création d'agences de contrôle de crédit, l'interdiction de collecter de l'épargne sont autant de politiques qui peuvent faire l'objet d'évaluation. La difficulté de concevoir des études pour mesurer les effets macroéconomiques de ces politiques suggère de procéder par des analyses au micro des comportements spécifiques en réaction à des politiques. Ces analyses permettront de renseigner les politiques macro.

1.3. Méthodes d'évaluation communes en microfinance

En suivant Hulme (2000), on distingue quatre méthodes d'évaluation communes en microfinance. La première méthode concerne *les évaluations basées sur des enquêtes par questionnaires*. L'échantillon est choisi de manière aléatoire en distinguant un groupe témoin et un groupe traité, avant et après la mise en œuvre du traitement. Dans cette catégorie, on retrouve aussi bien des évaluations expérimentales que quasi-expérimentales. Ensuite, *les évaluations dites rapides* regroupent une gamme d'outils et de techniques développés à l'origine sous les évaluations rurales rapides. Elles impliquent l'utilisation de groupes de discussions, des entretiens semi-structurés et avec des informateurs clés, d'études de cas, d'observations de participants et de sources secondaires. Dans cette catégorie, on peut citer les évaluations de type *credit scoring*, de mesure de la portée, coût bénéfice, de retour social sur investissement (voir Weber, 2013). La méthode d'*observation des participants* consiste à un séjour prolongé dans une communauté des chercheurs d'un programme en utilisant des techniques qualitatives et des enquêtes par sondages à petite échelle. Cette méthode relève beaucoup du domaine de l'anthropologie basée sur une approche inductive, une focalisation sur les informateurs clés, l'enregistrement par notes ou par image, et l'analyste de données est généralement directement (et fortement) impliqué dans la collecte de données. On ne cherche pas à prouver l'impact, mais plutôt à fournir une interprétation des processus impliqués dans l'intervention et des impacts plausibles. Enfin, les *études de cas* font une étude détaillée sur une unité spécifique, un groupe, une communauté, une localité ou un organisme de microfinance et comporte un questionnement ouvert et de détermination de l'histoire. Enfin, nous avons les apprentissages et actions participatives. Elle se fait par une préparation du programme par les bénéficiaires en partant de l'élaboration du chronogramme, de graphiques d'impacts, de cartes de villages et des ressources nécessaires, du classement du bien-être et de la richesse, de l'élaboration de diagramme saisonniers, au classement des problèmes et

d'évaluation institutionnelles par le biais de groupes assistés par un facilitateur.

Toutes ces méthodes permettent d'apporter une compréhension de l'influence des services de microfinance ou bien d'autres aspects de la microfinance sur le bien-être ou sur diverses autres dimensions³. Mais, notre analyse se concentre ici sur les méthodes qui cherchent à « prouver l'impact », notamment celles basées sur des méthodes scientifiques rigoureuses utilisant des données d'enquêtes généralement par questionnaire: l'approche quasi-expérimentale et l'approche des expérimentations aléatoires.

2. Approche non expérimentale ou quasi-expérimentale

L'approche non expérimentale de l'évaluation regroupe des méthodes statistiques et économétriques qui utilisent des bases de données dont les enquêtes n'ont pas suivi un processus aléatoire de choix des enquêtés. En général, les méthodes d'évaluation utilisées sont des méthodes quantitatives. Les méthodes couramment utilisées sont les scores de propensions, la méthode des variables instrumentales, la méthode de régression discontinue et la méthode de la différence double. Avant de présenter les différentes méthodes nous posons l'un des problèmes cruciaux de l'évaluation d'impact, la question du biais de sélection, ensuite, nous définissons les principes de base des méthodes, puis nous analysons de manière brève ces différentes méthodes et enfin, nous résumons les avantages et les limites de cette approche.

2.1. Le problème des biais dans les évaluations quasi-expérimentales

Différents types de biais peuvent apparaître dans le processus d'évaluation dans l'approche quasi-expérimentale (biais de sélection, d'attrition, biais de contagion, etc.)⁴. Le biais le plus souvent mis en relief est le biais de sélection. En s'inspirant Hulme (2000), on peut citer quatre raisons de l'existence probable de biais. Le biais de sélection traduit les difficultés à trouver un endroit où un environnement économique et social du groupe de contrôle (le contrefactuel) et du groupe de traitement qui soient identiques (i). Au biais de sélection d'autres biais peuvent subvenir et entacher les résultats s'ils ne sont pas pris en compte. En effet, le groupe de traitement possède systématiquement des caractéristiques inobservables que le groupe de contrôle n'a pas, notamment l'esprit entrepreneuriale et la motivation dans la conduite d'activités génératrices de revenus (ii) ; le fait de recevoir différents traitement peut entraîner un effet positif à court terme sur le groupe de traitement, notamment un *effet Hawthorne*⁵ (iii) ; le groupe de contrôle peut être contaminé par son contact avec le groupe de traitement, bien que les effets de cette contagion se produisent à long terme dans le programme, c'est le biais de contagion (iv). À cela peut-on ajouter la fongibilité du traitement, qui

survient lorsque le prêt est transféré d'une personne à une autre, ou lorsque le prêt est utilisé à d'autres fins (*cf.* tableau en annexe). C'est pour tenter de résoudre ces problèmes de biais que diverses méthodes statistiques et économétriques ont été développées que nous analysons à la section suivante.

2.2. Méthodes statistiques des évaluations non expérimentales

Nous commençons par présenter brièvement les principes de base qui ont permis l'élaboration de ces différentes méthodes statistiques de l'évaluation.

- **Principes de base de l'appariement**

La méthode de l'appariement part de l'hypothèse que les différences entre les deux groupes d'individus (traités et non traités) proviennent des seules caractéristiques observables et de l'effet du traitement. Afin d'avoir des individus qui ont des caractéristiques observables similaires, on sélectionne dans le groupe des non traités, un sous-ensemble d'individus ayant les caractéristiques similaires à ceux du groupe traité. Le principe est de considérer qu'on élimine le biais de sélection lorsqu'on compare des individus ayant des caractéristiques observables identiques. On part de l'hypothèse forte que les différences entre les deux groupes de population résultent uniquement des caractéristiques individuelles observables et de l'effet du traitement. Ainsi, il est possible de connaître l'effet du traitement si on neutralise les caractéristiques individuelles observables. Deux possibilités se présentent s'agissant de chaque individu de la population: accès ou non aux services du programme de microfinance. Ainsi, on représente cette situation par une variable aléatoire binaire D pour chaque individu i de manière suivante :

$D_i = 1$ si l'individu a accès au programme de microfinance

$D_i = 0$ sinon

L'efficacité du programme est analysée à partir des variables latentes de résultat Y_{1i} et Y_{0i} . tel que :

Y_{1i} correspond à un individu i reçoit le traitement et alors $D = 1$ et,

Y_{0i} correspond à un individu ne reçoit pas le traitement et on a $D = 0$

L'effet causal pour chaque individu est déterminé par Δ_i tel que :

$$\Delta_i = Y_{1i} - Y_{0i} \quad (1)$$

Cet écart représente la situation d'un individu selon qu'il est ou non traité. Or cette situation ne peut être observée de manière spontanée sur un même individu. On se contente donc de déterminer les effets moyens du traitement dans la population. Deux effets sont généralement analysés: l'effet moyen du traitement dans la population, Δ^{ATE} et l'effet moyen du traitement sur les individus traités, Δ^{ATT} .

À partir de ces principes de base de l'appariement, diverses méthodes ont été développées dans les études d'évaluation. Nous faisons une présentation brève des plus couramment utilisées.

- ***La méthode des scores de propension***

En pratique, l'appariement sur un ensemble de variables pose un problème de dimension du vecteur des caractéristiques observables (X). En effet, lorsque le nombre de variables de conditionnement est élevé, il est difficile de trouver pour chaque individu traité un individu non traité qui lui est exactement semblable du point de vue des caractéristiques observables. Pour résoudre ce problème, Rubin et Rosenbaum (1983) ont développé la méthode des scores de propension. Cette méthode repose sur deux hypothèses fondamentales dites hypothèses de « forte ignorabilité »⁶ qui sont importantes pour la résolution du problème des biais de sélection (Rubin and Rosenbaum, 1983). Ce sont l'hypothèse d'indépendance conditionnelle (CIA⁷) et l'hypothèse de support commun. La condition d'indépendance conditionnellement aux caractéristiques observables est définie comme suit:

$$(Y_0, Y_1) \perp D | X \quad (2)$$

Cette hypothèse suppose qu'étant donné un ensemble de variables observables X non affecté par le traitement, les résultats latents sont indépendants de l'affectation au traitement. Ce qui implique que les variables qui influencent l'affectation au traitement et les résultats du traitement sont observables directement par le chercheur (Caliendo et Kopeinig, 2008). C'est une hypothèse forte qui doit être vérifiée par la qualité des données.

La seconde hypothèse est la condition de support commun définie formellement comme suit:

$$0 < P(D = 1 | X) < 1 \quad (3)$$

Cette condition assure que tous les individus ayant les mêmes caractéristiques observables X ont la même probabilité d'accès au programme de microfinance.

À partir de ces hypothèses fondamentales, l'idée de Rubin et Rosenbaum (1983) des scores de propensions consiste à ramener le nombre de variables de conditionnement de dimension k à une seule et unique variable qui est un résumé univarié de l'ensemble des covariables X . Cette variable est le score de propension noté :

$$P(X) \equiv Pr(D = 1 | X) \quad (4)$$

Cette méthode de Rubin et Rosenbaum (1983) énonce que si la variable de résultat Y_0 est indépendante de l'accès au traitement (accès aux IMF dans notre cas), conditionnellement aux observables X , alors elle est également indépendante du traitement D , conditionnellement au score de propension $P(X)$ tel que :

$$Y_0 \perp D | X \Rightarrow Y_0 \perp D | p(X) \quad (5)$$

Ce résultat suggère que plutôt qu'apparier les individus sur l'ensemble des variables de conditionnement, il suffit de les apparier à leur score de propension qui est un résumé unidimensionnel de l'ensemble de ces variables. Ainsi, un individu non traité \bar{i} , apparié avec un individu i , est défini par son score de propension comme suit: $P(X_{\bar{i}}) = P(X_i)$. Considérons une population d'individus i . Si le score de propension $P(X_i)$ est connu alors, l'effet du traitement sur les traités peut être défini comme suit :

$$\begin{aligned} \Delta^{TT} &\equiv E[Y_{1i} - Y_{0i} | D = 1] & (6) \\ &= E[E(Y_{1i} - Y_{0i}) | D = 1, p(X_i)] \\ &= E[E(Y_{1i} | D_i = 1), p(X_i)] - E[E(Y_{0i} | D_i = 0), p(X_i) | D_i = 1] \end{aligned}$$

Une fois que les conditions de « forte ignorabilité » sont respectées, on peut alors estimer à partir d'un modèle de probabilité standard l'effet du traitement.

On peut définir à partir de : $\Pr(D_i = 1 | X_i) = F\{h(X_i)\}$ avec $F(\cdot)$, une fonction de répartition normale ou logistique et $h(X_i)$ une fonction linéaire des variables pour lesquelles les conditions d'équilibre sont satisfaites. Les fonctions $F(\cdot)$ utilisées sont les modèles Probit ou Logit pour l'estimation des paramètres nécessaires à la détermination de l'effet du traitement.

Divers méthodes d'estimation peuvent alors être utilisées pour déterminer l'effet du traitement, notamment les méthodes proposées par Becker et Ichino (2002), Abadie et Imbens (2006), etc.

- **Le score de propension généralisé**

L'analyse des scores de propensions va être prolongé avec Imbens (2000) avec les scores de propensions généralisées (SPG) pour prendre en compte les cas de traitements multiples et par Hirano et Imbens (2004) pour le cas de traitement continu (Lee et Salanié, 2018). Heckman et al. (2006, 2008) ont montré comment (en fonction de restrictions et des instruments), une variété d'effet pouvait être identifiée dans les modèles de choix discret et qui peuvent séparer de manière additive dans les instruments et les erreurs. D'autres auteurs, notamment Cattaneo (2010) et Cattaneo et al. (2013) ont défini deux estimateurs qui permettent de déterminer de manière efficace une variété de traitements à effets multiples. Ils élaborent la commande *poparms* sur Stata qui permet d'effectuer ces estimations. Plus récemment, Heckman et Pinto (2018) ont défini une condition de « monotonie non ordonnée », plus faible lorsque le traitement est multiple. Ils montrent que compte tenu de la monotonie non ordonnée du traitement, plusieurs effets du traitement peuvent être identifiés. En microfinance, il s'agirait par exemple, de pouvoir identifier les effets induits de différents services sur les bénéficiaires. On pourrait ainsi pouvoir séparer les effets du microcrédit des services de micro-assurances.

- **La méthode des variables instrumentales**

Lorsque dans une équation de résultat, le terme de l'erreur et la variable indicatrice de participation ou non au programme sont corrélés alors, il y a un biais de sélection. La méthode des variables instrumentales tente de résoudre ce problème. Globalement, la méthode consiste à trouver une variable appelée instrument qui influe sur la variable indicatrice de sélection à la participation ou non au programme, mais qui n'affecte pas les résultats ou l'effet du programme. Cette variable n'étant pas corrélée au terme d'erreur, on peut donc faire une estimation non biaisée de l'effet du programme. Plus la corrélation entre la variable instrumentale et la variable de sélection de la participation ou non au programme est précise ou forte, plus l'estimation de l'effet du programme donne des résultats précis. La difficulté de cette méthode réside dans la détermination d'un instrument « valide » qui ait une corrélation forte avec la variable indicatrice de sélection sans avoir d'effet sur le résultat. Lorsque l'instrument utilisé est faible, c'est-à-dire lorsque la corrélation entre la variable endogène et l'instrument tend vers 0, le risque que les estimateurs soient imprécis est élevé. En suivant Kono et Takahashi (2010), de manière formelle, considérons une équation linéaire standard:

$$y_i = \alpha + \beta * d_i + e_i \tag{7}$$

avec y_i , le revenu ou les gains des ménages i , α est le terme constant correspondant à $E(Y_0|d = 0)$, β le paramètre égal à $E(Y_1|d = 1) - E(Y_0|d = 0)$ et d le traitement.

Si l'on considère que d_i est endogène, c'est-à-dire qu'il est corrélé au terme d'erreur e alors l'estimateur des moindres carrés ordinaires est biaisé et il existe donc des effets liés au biais de sélection. Pour résoudre ce problème, il faut trouver une variable qui soit corrélée à la variable d_i et non corrélée aux résidus.

Soit Z la variable instrumentale définie telle que:

$cov(Z_i, d_i) \neq 0$, la variable instrumentale est corrélée à la variable endogène d_i ,

$cov(Z_i, e_i) = 0$, la variable instrumentale n'est pas corrélée au résidu e ,

La variable Z est donc aléatoire et non corrélée aux déterminants inobservés du traitement.

On détermine d_i comme suit :

$$d_i = \alpha_2 + \beta_2 * z_i + \mu_i \tag{8}$$

En remplaçant d_i dans l'équation (11), on obtient:

$$y_i = \alpha + \beta(\alpha_2 + \beta_2 * z_i + u_i) + e_i \tag{9}$$

$$y_i = (\alpha + \beta\alpha_2) + (\beta\beta_2 * z_i) + (\beta u_i + e_i)$$

Par hypothèse, on a: $cov(Z_i, e_i) = 0$ et $(Z_i, u_i) = 0$, alors la forme réduite estimée sera non biaisée. On peut donc estimer le paramètre β de manière consistante en divisant $\beta\beta_2$ par β_2 .

De plus, en supposant que d est un indice unique suivant l'hypothèse de monotocité, c'est-à-dire $d_{1i} \geq d_{0i}$ et que z est une variable binaire, l'estimateur β couramment utilisé s'écrit :

$$\beta = \frac{E(y_i|z_i = 1) - E(y_i|z_i = 0)}{E(d_i|z_i = 1) - E(d_i|z_i = 0)} \quad (10)$$

β correspond à l'estimateur de Wald. Il estime le changement de y relatif à d induit par le changement de z compris entre 0 et 1. En d'autres termes, il correspond à la différence moyenne empirique du revenu du groupe des individus pour lesquels $Z = 1$ et pour $Z = 0$ rapporté à l'équivalent de ces groupes pour l'instrument. Cela correspond à une estimation des doubles moindres carrés à partir de l'équation 11. Il faut noter qu'avec cette méthode, l'impact estimé ne reflète pas l'effet du traitement sur l'ensemble de la population mais l'effet du traitement sur la sous population qui a accès au traitement quand celui-ci est administré (Imbens et Angrist, 1994). Autrement dit, l'impact obtenu correspond à celui de la sous-population qui a été traité lorsque le traitement a été offert. Dans ce cadre, l'estimateur de la variable instrumentale calculé est appelé effet moyen du traitement local ou *Local Average Treatment Effect* (LATE) (Imbens et Angrist, 1994). Plus précisément, l'impact déterminé par l'effet moyen du traitement est local car, il ne porte que sur une sous-population, celle qui réagit à l'instrument ou celle pour laquelle l'instrument a un effet.

- ***La méthode de régression discontinue***

Le modèle de régression discontinue part de l'idée de l'existence d'une variable de sélection qui a un impact discontinu sur la probabilité d'être éligible ou non à un programme. Elle suppose que les individus sont sensiblement identiques autour du seuil ou du point de discontinuité. Ainsi, il est possible d'identifier l'effet causal du programme en comparant les individus qui ont bénéficié du programme et les autres qui n'ont pas été sélectionnés. De manière triviale les conditions de Rubin et Rosembaum (1983) sont satisfaites car, s'il est impossible d'observer la situation de traité et de non traité pour une même valeur de X , il nous est possible d'observer ces deux résultats pour des valeurs de X autour d'un point seuil. Par exemple, supposons qu'un programme de microfinance n'octroie les prêts qu'aux personnes disposant de moins d'un hectare de terre cultivable, ceux qui en possèdent plus étant inéligibles. Cette variable supposée exogène au programme de microfinance n'est pas *a priori* corrélée aux caractéristiques inobservables du programme. Il est donc possible, à partir de cette variable de déterminer l'impact du programme en comparant ces deux groupes de personnes.

La difficulté de cette méthode réside dans le choix de la variable seuil qui soit discontinue et qui respecte le critère d'exogénéité par rapport au

programme et de non corrélation avec les caractéristiques inobservables des individus. Par ailleurs, la méthode nécessite, pour rendre l'analyse beaucoup plus pertinente et précise, que l'on dispose d'une base de données importante dans laquelle il existe un nombre d'observations suffisantes autour du seuil.

De manière formelle, soit c le seuil critique de la variable indicatrice de sélection C de la participation ou non à un programme de microfinance. $d = 1$ si $C > c$ et 0 si non, c'est-à-dire, les individus au-dessus du seuil critique c sont éligibles ou ont accès au programme de microfinance tandis que ceux en dessous de ce seuil ($C < c$) ne sont pas éligibles ou n'ont pas accès aux services du programme de microfinance.

Si cette règle est établie et que le traitement (accès au programme de microfinance) dépend de manière déterministe de la variable de sélection C , l'estimateur de régression discontinue est dit *sharp regression discontinuity* et peut être formulé selon Kono et Takahashi (2010) comme suit:

$$E(y_1|c \leq C \leq c + e) - E(y_0|c - e \leq C < c) \quad (11)$$

En revanche, si la variable de sélection ($C < c$) affecte la probabilité d'être traité, l'affectation n'est plus parfaite et est donc exécutée avec une erreur. Alors, on aura une régression discontinue stochastique dite (*fuzzy regression discontinuity*). On amplifie les différences par division de la probabilité de traitement en tenant compte de l'erreur e . L'estimateur se présente alors comme suit:

$$\beta = \frac{E(y_i|c_i \leq C < c_i + e) - E(y_i|c_i - e \leq C < c_i)}{E(d_i|c_i \leq C < c_i + e) - E(d_i|c_i - e \leq C < c_i)} \quad (12)$$

Il est équivalent à l'estimateur de Wald avec une variable dichotomique ($C < c$) comme instrument pour le statut de traité (Kono et Takahashi, 2010). Cet estimateur permet d'évaluer l'impact de la sous population autour du seuil critique et non l'impact moyen du traitement sur l'ensemble de la population (Ravallion, 2008). Il correspond à l'effet moyen du traitement local (LATE) défini plus haut. La généralisation des résultats requiert donc une appréciation selon le cas étudié. La méthode de régression discontinue a pour avantage de produire des estimations non biaisées de l'effet du traitement au point de discontinuité. Il permet d'utiliser des règles connues dans l'évaluation des avantages communs dans les politiques sociales. On n'exclut pas dans l'évaluation du traitement les ménages ou les individus éligibles au programme mais qui n'en bénéficient pas. Cependant, l'impact obtenu reste localisé autour du seuil et n'est pas toujours généralisable. L'effet du traitement obtenu autour du point de discontinuité utilise moins d'observations que dans une expérience aléatoire avec un échantillon de taille similaire. En outre, la spécification du modèle peut être sensible à la forme fonctionnelle de l'équation de régression y compris les interactions et les relations non linéaires possibles.

- **La méthode de la double différence**

On l'a vu plus haut, comparer simplement des individus qui ont accès à un programme de microfinance à d'autres qui n'y ont pas accès et déduire de cette comparaison l'impact dudit programme conduit à des résultats biaisés. De même, comparer simplement la situation des individus avant et après la mise en œuvre du programme de microfinance et en déduire l'impact n'est pas approprié. D'autres facteurs indépendamment du traitement sont susceptibles d'expliquer les changements dans le temps. Ces deux manières de procéder prises individuellement ne suppriment pas les éventuels biais. Mais on peut y remédier en combinant ces deux situations. Cette combinaison est appelée la méthode de la double différence ou la méthode DID⁸. Elle détermine les différences systématiques et temporelles en fonction du traitement entre les deux groupes (groupe traité et de contrôle). Une première comparaison permet d'éliminer les différences systématiques entre les deux groupes. Celle-ci compare la différence entre les groupes de traitement et de contrôle avant la mise en œuvre du programme dont on souhaite analyser l'impact. Le résultat obtenu de ce calcul constitue le biais de sélection. La seconde comparaison élimine l'évolution temporelle qui est supposée identique pour les deux groupes en l'absence du traitement. Elle est calculée en prenant la situation avant et après la mise en œuvre du programme pour chaque bénéficiaire. Enfin, de ces deux comparaisons, on fait la différence des deux différences et on obtient l'effet du programme. À la suite de Kono et Takahashi (2010), une formalisation de l'équation de l'impact d'un programme de microfinance peut être définie comme suit:

$$y_{it} = \beta_0 d_i + \beta_1 T_t + \beta_2 (T_t * d_i) + e_{it} \quad (13)$$

avec y_{it} , la variable d'intérêt déterminant l'effet du programme d'un ménage i à la période t , β_0 , β_1 , β_2 les paramètres à estimer et e , le terme d'erreur. T prend la valeur 0 et 1, avec $T = 0$ correspondant à la période dite de pré-traitement et $T = 1$, la période de post-traitement. L'estimateur DID (β_2) sera biaisé si le statut de traité est corrélé au terme d'erreur de sorte qu'on ait: $cov(e_{it}, d_i) \neq 0$. Supposons que le terme d'erreur e comprend une composante invariable dans le temps v , et une autre variable ϵ tel que: $e_{it} = v_i + \epsilon_{it}$. De plus, supposons que e_{it} est indépendant de la participation au programme de microfinance; c'est-à-dire que pour tout ménage donné:

$\epsilon_{it} \perp d_i$ et $cov(\epsilon_{it}, \epsilon_{it+1}) = 0$. La différence première se présente alors comme suit :

$$\Delta y_i = \beta_1 + \beta_2 d_i + \Delta \epsilon_i \quad (14)$$

où Δ représente le changement des variables dans le temps. Cette équation montre clairement que les caractéristiques restent inchangées dans le temps et ne sont pas corrélées avec la différence première. L'estimateur DID est donc sans biais car, par hypothèse $\Delta \epsilon_i$ n'est pas corrélé à d_i . L'estimateur

DID est analogue à l'effet moyen du traitement sur les traités et s'écrit comme suit :

$$DID = [E(\Delta y_1 | d = 1) - E(\Delta y_0 | d = 0)] - [E(\Delta y_1 | d = 0) - E(y_0 | d = 0)] \quad (15)$$

Il y aura biais si $E(\Delta y_1 | d = 0) \neq E(\Delta y_0 | d = 0)$. Ainsi, sous l'hypothèse que l'effet temporel est commun et donc sans effet, alors, le résultat de l'impact est fonction des changements des résultats du groupe de comparaison entre les périodes pré et post-traitement identiques à ceux du groupe de traitement en l'absence de traitement.

L'une des difficultés de cette méthode est liée au biais de « sélection invariante » dans le temps. La validité de cette méthode repose en partie sur l'hypothèse forte qu'il y a une absence d'effet temporel. Le fait de postuler que le résultat n'aurait pas connu de tendances différentes dans les régions qui ont reçu le traitement si le programme n'avait pas été mis en place constitue une hypothèse forte qui doit être considérée avec réserves comme le souligne Duflo (2005). Cette hypothèse est difficile à tester et à vérifier de manière plausible car elle nécessite une base de données sur des périodes assez longues, antérieures à la mise en place du programme indispensable à des comparaisons tendancielles. Encore faut-il s'assurer qu'aucun programme n'ait été mis en place dans le même temps. La survenue de chocs imprévus ou passés peut être confondus avec les effets du programme.

- **La méthode « pipeline matching »**

L'hypothèse de base considère que les caractéristiques inobservables du groupe de contrôle sont identiques à celles du groupe traité dans la mesure où tous les individus sont éligibles au programme. Ainsi, la différence entre les deux groupes proviendrait de l'effet du programme. Le *pipeline matching* peut être combinée avec les méthodes précédentes pour déterminer l'impact d'un programme. Vu l'importance et la nécessité d'évaluation récurrente dans le secteur de la microfinance contrariée par des évaluations coûteuses financièrement et nécessitant un temps long et des techniques complexes, ce projet a adopté cette méthode simplifiée d'évaluation des IMF. Elle diffère principalement des méthodes précédentes par la méthode de collecte des données et par la différenciation du groupe traité et non traité. En effet, l'analyse porte essentiellement sur les clients bénéficiaires du programme que l'on essaie de distinguer selon certains critères. Le critère de distinction généralement adopté consiste à opposer anciens et nouveaux clients. Les anciens clients sont des personnes qui ont accès aux services de l'IMF sur une période relativement plus longue (un an, mais variable selon les études) et qui ont bénéficié de prêts ou d'autres services de l'IMF. Les nouveaux clients sont, soit des personnes éligibles au programme et en attente de recevoir ses services, soit des personnes qui viennent d'entrer dans le programme mais,

n'ont pas encore reçu de prêts. Cette distinction des clients diffère selon les études, mais globalement, les anciens clients sont considérés comme le groupe traité et les nouveaux clients, le groupe de contrôle. Cette méthode a été mise en œuvre par le projet AIMS (*Assessing the Impact of Microenterprise*) de l'USAID. Elle a été beaucoup utilisée dans l'évaluation des programmes.

L'avantage de cette méthode est qu'elle est simple, moins coûteuse et nécessite moins de temps en termes de collecte de données. Les clients sont connus et enregistrés au sein de l'IMF et l'on peut utiliser leur livret de compte et avoir accès à leur historique d'épargne et de crédit. La méthode ne nécessite pas de données de panel et d'interviews à réaliser sur des non clients, ce qui permet au départ un ciblage de la population à enquêter. La démarche a été largement utilisée dans plusieurs études⁹. En général, ces études concluent à un effet positif et significatif de l'impact du microcrédit sur les profits des entreprises et sur le bien-être des clients. Toutefois, la méthode n'est pas exempte de critiques. En pratique, l'on ne peut exclure l'existence de probables biais. Il peut exister entre les clients qui ont reçu les prêts et ceux qui sont en attente des différences inobservées liées à l'esprit entrepreneurial, à la prise de risque, etc. Par ailleurs, les individus peuvent être clients d'un programme pour épargner sans pour autant demander du crédit. Dans ce cadre, il est possible de contrôler l'hétérogénéité inobservée à travers des effets fixes ou des estimations des différences doubles si les données le permettent. Quant à l'hétérogénéité observée, elle peut être prise en compte par la méthode des scores de propension avant de faire le *pipeline matching* (Galasso et Ravallion, 2004)¹⁰. Ce problème de biais de sélection a été critiqué à l'égard de certaines études qui ont adopté cette méthode comme outil d'évaluation d'impact de programme de microfinance (Hulme et Mosley, 1996; Coleman, 1999), etc.

2.3. Avantages et limites de l'approche quasi-expérimentales

Les avantages de l'approche quasi-expérimentale peuvent être perçus en termes de coûts, de gains de temps et pertinence des résultats obtenus du fait de la rigueur de l'analyse qui permet d'obtenir des preuves scientifiques de l'existence ou non d'effets induits par le programme de microfinance. Le coût de réalisation de ces analyses sont relativement moins coûteuses en comparaison aux expérimentations aléatoires. En effet, ces études peuvent être conduites avec des données d'enquêtes existantes améliorées, ce qui peut réduire significativement le coût, mais aussi le temps de réalisation de ces évaluations. En outre, les résultats de ces analyses améliorent les connaissances sur le secteur de la microfinance et permettent de renforcer, d'améliorer et d'orienter le secteur vers ses objectifs prioritaires, l'inclusion financière et l'amélioration du bien-être des communautés.

Quant aux limites, elles sont diverses et concernent essentiellement la pertinence des résultats obtenus. La difficulté de trouver des contrefactuels

pertinents pour des programmes déjà mises en œuvre causes d'énormes difficultés. La question des problèmes des divers biais auxquels les différentes méthodes tentent de répondre est sujet à discussions. L'usage des techniques statistiques et économétriques que nécessitent ces méthodes d'évaluation requiert des hypothèses fortes sur la distribution des erreurs. En outre, la diversité des méthodes et leur complexité rend difficile la compréhension des résultats pour les décideurs. Par ailleurs, les résultats peuvent varier en fonction de la spécification du modèle et surtout lorsque la théorie propose plusieurs analyses alternatives. Par ailleurs, les méthodes d'appariement sur des variables de résultats sont inadaptés avec la persistance des effets individuels inobservés, même avec l'introduction d'un grand nombre de caractéristiques individuelles observables. Cette situation peut être comblée avec l'estimation des doubles différences, mais sa pertinence dépend de la richesse des variables de conditionnement. L'usage des méthodes d'appariement devient difficile voire impossible lorsqu'il n'est pas possible d'éliminer les effets individuels avec l'application des différences premières. Si les scores de propension permettent de mettre en relief l'hétérogénéité du traitement, elles posent la difficulté d'extrapolation des résultats. Il faut noter également que le choix des méthodes dépend du contexte de l'évaluation pour réduire les biais spécifiques liés à l'évaluation. Cela conditionne la précision des estimations et donc la qualité et la pertinence des résultats.

Au terme de cette revue des outils méthodologiques, il ressort que chacune de ces méthodes tente de répondre au problème de biais de sélection et d'hétérogénéité auxquels on est confronté dans les évaluations. Ces évaluations portant généralement sur des groupes restreints de la population (bénéficiaires et non bénéficiaires) posent la question de la généralisation des effets induits de ces programmes. Toutefois, les enseignements et les informations tirées de ces analyses sont importantes en termes de leçons apprises pour des projets et programmes futurs et pour une amélioration de ceux existants en cours de mises en œuvre. Mais, qu'en est-il des expérimentations aléatoires en microfinance ? La section suivante présente brièvement l'approche expérimentale, notamment celles des évaluations aléatoires en vogue aujourd'hui dans l'évaluation des programmes de développement et en microfinance en particulier.

3. L'approche des expérimentation aléatoires ou randomisations

D'un point de vue historique, la méthode issue des sciences médicales a été introduite en agronomie puis en sciences sociales. L'incursion de cette méthodologie dans les sciences sociales a suscité des débats et des controverses entre sceptiques et défenseurs sur sa portée et ses limites, notamment dans l'évaluation des politiques de développement¹¹. Inspirée des sciences des essais cliniques la méthode des assignations aléatoires a été

transposée dans les sciences sociales. On est ainsi passé d'une médecine des preuves (*evidence-based medicine*) à une politique des preuves (*evidence-based policy*) (Harry, 1997). L'évaluation par assignation aléatoire se distingue des techniques non expérimentales par la qualité des données qu'elle permet de produire. L'évaluation réalisée est plus précise, fiable et « la qualité de la preuve est comparable à celle d'un flagrant-délit pour une cour de justice, alors que les méthodes qui mobilisent des données non expérimentales sont équivalentes à une accumulation de présomptions de culpabilité sans jamais véritablement administrer une preuve définitive » (L'Horty et Petit, 2010). Ainsi, permet-elle de fonder l'aide publique au développement non plus sur des preuves non convaincantes, sans consistance (*Wishy-Washy evidence*) des analyses en coupe internationale, mais sur des preuves solides, concrètes et convaincantes (*Hard evidence*) (Labrousse, 2010). Les expérimentations aléatoires se positionnent comme une alternative à la suite des difficultés rencontrées par les méthodes non expérimentales dans la suppression des biais de sélection. En effet, la suppression du biais de sélection constitue la difficulté majeure dans toute évaluation qu'elle soit aléatoire ou non expérimentale. L'évaluation par assignation aléatoire tente d'y remédier en sélectionnant à l'avance et avec minutie, de manière aléatoire les groupes traités et de comparaison au sein d'une population de bénéficiaires potentiels. On est a priori assuré que les groupes traités et les groupes de contrôle ne sont pas si différents. On peut alors attribuer en toute confiance aux effets du programme toutes les différences statistiquement significatives entre les deux groupes après administration du traitement. La méthode des évaluations par assignation aléatoire est particulièrement récente en microfinance. Ce n'est que récemment que cette méthode vulgarisée par les chercheurs du *J-PAL*¹² qui l'ont utilisé dans le secteur de la microfinance. La randomisation connaît ces dernières années une reconnaissance tant au niveau des chercheurs qu'au niveau des ONG et des organismes internationaux de développement comme un nouveau paradigme dans le champ de la nouvelle microéconomie du développement. Cependant, elle n'est pas nouvelle, son extension en économie date depuis la fin des années 1960 avec les premières expérimentations réalisées aux Etats-Unis, en Grande-Bretagne et dans certains pays d'Europe du Nord. Pour Jatteau, (2013), la résurgence des expérimentations aléatoires est loin d'être une innovation méthodologique, mais plutôt, le changement des terrains d'expérimentation, principalement, les pays pauvres et la volonté de promouvoir cette méthode comme *gold standard* de l'évaluation. Cette approche de l'évaluation est désormais mise en œuvre dans divers domaines du développement, notamment dans le secteur de la microfinance, de la santé et de l'éducation dans les pays pauvres. Dans la suite de cette section nous présentons brièvement, les principes, les méthodes

d'estimation de l'assignation aléatoire. Ensuite nous présentons ces avantages et ces limites.

3.1. Conditions de l'expérimentation aléatoire

Ces conditions ou principes sont des éléments à prendre en compte dans le processus de l'évaluation. Elles concernent le contrefactuel, la taille de l'échantillon et des principes de collecte de données.

- ***La détermination du contrefactuel***

Le problème central des expérimentations aléatoires est la détermination du contrefactuel encore appelé groupe de contrôle ou de comparaison. Dans le cadre d'un programme de microfinance, l'évaluation du programme consisterait à savoir comment les conditions de vie des bénéficiaires diffèrent de celles qui auraient prévalu si le programme n'avait pas été mis en œuvre. En d'autres termes, quelle serait la situation des bénéficiaires en l'absence du programme? Ou quelle aurait été la situation de ceux qui n'ont pas bénéficié du programme s'ils y avaient été exposés? Seul, le choix d'un contrefactuel précis permet de répondre à cette interrogation. Il désigne ce qui se serait passé pour un individu s'il avait fait partie du groupe auquel il n'avait pas été assigné (Jatteau, 2013). Toute la difficulté de l'évaluation consiste à déterminer le bon contrefactuel qui permet d'isoler l'effet propre du traitement. Ceci est très important en ce sens que les effets peuvent provenir d'autres facteurs liés par exemple, à la composition des groupes, à l'existence de biais de sélection.

- ***La taille de l'échantillon***

La mise en œuvre d'une expérimentation aléatoire requiert un nombre d'individus nécessaires pour déterminer avec précision l'impact. L'utilisation de techniques statistiques permettent de déterminer la taille adéquate de l'échantillon. Mais en général, plus les groupes sont de grande taille, plus les chances de comparaison augmentent. Ainsi, de la taille de l'échantillon dépend la confiance que l'on peut avoir dans les résultats et leur précision. En d'autre termes, plus le niveau de confiance est élevé, plus l'intervalle de confiance est réduit, c'est-à-dire moins les différences entre les deux groupes sont de faible ampleur. Des formules de calcul précis de la taille des échantillons existent et sont inspirées de la médecine et de la théorie des sondages (Jatteau, 2013).

- ***Principes de collecte des données***

Nous nous inspirons de Fougère (2010) pour distinguer les principes nécessaires dans le processus de collecte des données d'une expérimentation aléatoire. Il est nécessaire que les participants à l'expérience soient informés

des avantages et risques que leur participation peut leur faire courir. C'est le principe de consentement éclairé (i); ensuite, il faut prévoir des compensations pour les individus qui seraient affectés au groupe de traitement si l'on anticipe que le traitement peut leur causer des risques ou des torts. À l'inverse, si le tort est causé au groupe de contrôle, alors il faut leur affecter une compensation, c'est le principe de compensation (ii); enfin, si l'expérimentation est jugée préjudiciable, il est préférable d'observer les effets sur un nombre limité d'individus consentants que sur un très grand nombre d'individus contraints, c'est le principe de précaution (iii).

La prise en compte de la taille et de ces principes dans la collecte des données sont nécessaires pour une adhésion des participants et pour l'obtention de résultats précis qui reflètent l'effet du traitement. Nous nous limitons à ces différents points, mais il faut noter que le processus d'expérimentation suit généralement différentes étapes qui peut être en sept en huit, commençant du montage institutionnel et financière et l'expérimentation à la publication des résultats¹³.

3.2. Méthodes d'estimation des expérimentations aléatoires

Dans la pratique de l'assignation aléatoire, il se pose le problème de l'adhésion partielle (*partial compliance*) au protocole de l'expérimentation. En effet, pour diverses raisons certains individus du groupe de traitement ne seront pas traités tandis que certains du groupe de contrôle peuvent recevoir le traitement. Prenons le cas d'une IMF qui proposent des prêts à taux réduit aux individus du groupe de traitement. Certains individus du groupe de traitement peuvent décider d'emprunter et d'autres non, pour des raisons liées à leurs attitudes risquophobes. Pour ces raisons et bien d'autres encore, deux indicateurs ont été formulés pour permettre de mesurer l'impact dans les expérimentations aléatoires: l'effet moyen en intention de traiter (*Intention to treat effect*, ITT) et l'effet moyen sur les traités (*Treatment on the treated*, TT). Dans le premier (ITT), on s'intéresse à l'impact du traitement sur ceux que l'on souhaitait traiter sans se soucier qu'ils ne l'ont pas tous et dans le second (TT), on s'intéresse uniquement à ceux qui ont été traités, c'est-à-dire qui ont reçu le microcrédit. En suivant Jatteau (2013), on formalise ces estimateurs comme suit : Notons $A = 1$, si l'individu est assigné au groupe de traitement et $A = 0$ s'il est assigné au groupe de contrôle, alors on a : l'effet moyen en intention de traiter s'écrit comme suit :

$$ITT = E(Y|A = 1) - E(Y|A = 0) \quad (16)$$

Ce résultat s'obtient facilement : il suffit de comparer les moyennes des *outcomes* du groupe de contrôle et du groupe de traitement. De cette façon, on prend en compte les *outcomes* des individus du groupe de traitement qui n'ont pas été traités, de même que ceux du groupe de contrôle qui l'ont été du fait de l'adhésion partielle au traitement. Dans cette phase, ce n'est pas

l'impact précis du traitement qui est apprécié, mais plutôt l'impact de sa diffusion. Pour s'intéresser de manière spécifique à l'impact du traitement lui-même, on s'intéresse qu'aux individus qui ont effectivement été traités, ceux qui ont reçu le microcrédit. On calcule TT:

$$TT = E(Y_1 - Y_0|T = 1) = E(Y_1|T = 1) - E(Y_0|T = 1) \quad (17)$$

Comme l'adhésion au programme n'est pas faite comme on l'avait supposé jusqu'alors, on a :

$$E(Y|A = 1) = E(Y_1|T = 1)E(T = 1|A = 1) + E(Y_0|T = 0)E(T = 0|A = 1)$$

$$E(Y|A = 0) = E(Y_1|T = 1)E(T = 1|A = 0) + E(Y_0|T = 0)E(T = 0|A = 0)$$

Grâce à la randomisation, les hypothèses suivantes sont vérifiées :

$$E(Y_0|T = 1) = E(Y_0|T = 0) \text{ et } E(Y_1|T = 1) = E(Y_1|T = 0)$$

Quand on remplace $E(Y_0|T = 0)$ par $E(Y_0|T = 1)$, on obtient après calculs :

$$ITT = E(Y|A = 1) - E(Y|A = 0) \\ = [E(Y_1|T = 1) - E(Y_0|T = 1)][E(T = 1|A = 1) - E(T = 1|A = 0)]$$

D'où

$$ITT = TT \times [E(T = 1|A = 1) - E(T = 1|A = 0)] \quad (18)$$

Donc on obtient une forme calculable de TT telle que:

$$TT = ITT/[E(T = 1|A = 1) - E(T = 1|A = 0)] \quad (19)$$

On remarque que si $E(T = 1|A = 1) = 1$ et $E(T = 1|A = 0) = 0$, c'est-à-dire que si l'assignation est parfaitement respectée, alors ITT et TT sont égaux : l'effet moyen en intention de traité et l'effet moyen sur les traités sont identiques. Ces deux indicateurs sont intéressants car ils informent sur deux situations différentes du protocole. L'ITT donne la situation réelle de ce qui se passe réellement quand on applique le programme de microcrédit. Quant à TT, il donne avec précision l'effet du traitement et est utile dans une logique d'extension ou de généralisation de l'octroi de prêts. Toutefois, l'assignation comporte aussi bien des avantages que des limites que nous énumérons dans la section suivante.

3.3. Avantages et limites des expérimentations aléatoires

Les problèmes liés aux expérimentations aléatoires sont multiples, mais nous contentons d'énumérer quelques-unes : l'éthique et le contexte environnemental et la validité externe des résultats, les biais de randomisation.

- *L'éthique en question*

Les expérimentations aléatoires soulèvent le plus souvent des questions éthiques. L'administration aléatoire d'un traitement ou d'un programme peut être contraire à l'éthique. En effet, l'identification d'un

groupe de contrôle et d'un groupe de traitement suppose que l'on choisisse en toute conscience de traiter un groupe de personnes et non tout le monde. Or, le traitement peut être une politique, une aide, un traitement médical ou l'accès au microcrédit que l'on juge *a priori* bénéfique pour la population. La privation du groupe témoin pour les besoins de l'étude peut aller à l'encontre même de l'objet de l'institution demandeur de l'expérimentation dont la finalité est l'amélioration du bien-être des personnes en situation de difficultés économiques. Dans ce cadre, la réalisation d'une expérimentation par assignation aléatoire nécessite au préalable une réflexion éthique. Cette réflexion peut parfois aboutir à une réorientation de l'expérimentation ou à son abandon surtout quand il s'agit d'expérimentations qui mettent en jeu des ressources vitales pour les personnes ou qui peuvent créer des situations de frustration pouvant mettre à mal la cohésion sociale existante avant le programme. La difficulté est réelle dans les pays en développement où la plupart des programmes concernent la mise en place de services indispensables à la satisfaction des besoins vitaux en faveur des pauvres. D'un point de vue éthique, l'évaluation est possible que lorsque le programme apporte une amélioration substantielle par rapport aux ressources de droit commun. En effet, on suppose que le groupe traité et de contrôle ont tous les deux accès aux ressources de droit commun. On peut alors affecter le traitement ou les ressources supplémentaires aux personnes sur lesquelles le test d'efficacité est réalisé. Or, ce processus qui consiste à s'assurer que tous les citoyens ont un accès garanti aux ressources de droit commun n'est pas envisageable dans des pays où les inégalités sont de fait très fortes et persistantes. L'accès aux ressources de droit commun est très limité et la satisfaction des besoins reste très faible quand il existe un minimum de ces droits. Pour contourner ces problèmes, plusieurs stratégies sont possibles.

On peut dans un programme administrer des traitements différents à une population sans exclure un groupe de personnes. Une méthode consiste également à différer ou à administrer graduellement dans un ordre aléatoire le traitement de sorte que les groupes témoins d'aujourd'hui soient les groupes traités de demain (L'Horty et Petit, 2010).

Le respect des préoccupations éthiques suggère que la mise en œuvre d'expérimentations aléatoires nécessite une mise en place de comités d'éthique. Ces comités ont pour objectif d'analyser la conformité d'un point de vue éthique en rapport avec le contexte socioéconomique et environnemental dans lequel s'inscrit l'évaluation.

- ***L'environnement comme limite à l'assignation aléatoire***

La dépendance à l'environnement est une autre critique généralement faite aux évaluations aléatoires. Cette dépendance soulève la question de la généralisation ou de la validité externe des résultats. Pour

Banerjee et Duflo (2009), une manière de généraliser un résultat consiste à conduire d'autres expérimentations supplémentaires dans des lieux différents, avec des équipes différentes. De cette façon, on peut connaître la distribution des effets du traitement d'un site à un autre à partir de n'importe quelle variable de contrôle, ce qui n'est pas possible avec les méthodes non expérimentales¹⁴. Mais, pour Rodrik (1998), les méthodes non expérimentales sont moins dépendantes de ce biais car elles couvrent dans la plupart des cas des domaines beaucoup plus larges, ce qui permet de faire aisément des généralisations.

- ***La validité externe***

La validité externe pose le problème de la généralisation de résultats de l'expérimentation. C'est la possibilité de généraliser les résultats à des personnes, sites ou époques différents de ceux qui faisaient l'objet de l'expérimentation (Monnier, 1992). Pour Rodrik (2008), les résultats des expérimentations aléatoires sont difficilement généralisables au-delà du contexte. L'expérimentation dans ce cas ne donne que des connaissances que sur les groupes qui ont participé à l'expérience, dans le temps et le contexte considéré. Mais la notion de généralisation pose elle-même problème. Que faut-il entendre par généralisation ? Faut-il opérer un choix entre les évaluations par assignation aléatoire dont la validité interne est forte ou conduire des évaluations non expérimentales où la validité externe est large, s'interrogent Banerjee et Duflo (2009). Si l'on définit la généralisation d'une étude comme l'impact de la moyenne de l'ensemble des variables ou des indicateurs que l'on peut regrouper à partir de données agrégées alors cela ne veut pas forcément dire grand-chose selon Banerjee et Duflo (2009). L'impact moyen obtenu à partir de données agrégées en coupe transversale ne saurait être considéré comme un justificatif de généralisation d'un programme. Un autre problème lié à la généralisation évoquée par ces auteurs est l'implication des équipes de terrain de l'expérimentation, par exemple de par sa rigueur peut être différente de celles qui vont mettre en œuvre la généralisation du programme dans une zone beaucoup plus vaste. Les effets peuvent se révéler faible ou nuls voire changer de sens.

- ***Le biais de randomisation***

Ce biais limite les possibilités de généralisation des effets du programme. En effet, la constitution de groupe traité et de groupe témoin induit parfois des modifications de comportements chez les acteurs. Conscients du fait qu'ils sont testés (*effet Hawthorne*) ou pris comme témoins d'une expérimentation (*effet de John Henry*), les individus adoptent des comportements tendant à favoriser leur inscription ou non dans le programme.

Tout comme dans les évaluations non expérimentales, des études ont été réalisées dans divers domaines avec cette méthodologie. Les expérimentations aléatoires sont beaucoup appréciées par les partenaires au développement qui l'ont intégrée dans leur démarche d'évaluation. Mais leurs coûts très élevés ne sont à la portée des pays pauvres à fortiori les petites IMF. Un arbitrage évident s'impose aux gouvernants des pays pauvres : utiliser ces ressources pour d'autres projets ou conduire des évaluations aux issues incertaines. Pour les IMF l'arbitrage est identique : octroyer du crédit à plus de clients pauvres ou conduire des évaluations coûteuses.

Conclusion

Cette revue des méthodes d'évaluation en microfinance comme dans toutes les évaluations de programme sont confrontés au problème du biais de sélection qui peut être traité de plusieurs façons. Le choix des méthodes dans la résolution de ce problème est spécifique aux hypothèses adaptées et au type de biais auquel on est confronté. Notre analyse de ces différentes et des méthodes confirme l'idée qu'il n'existe pas de méthode clé en main adaptée à toutes à toutes les circonstances. Les méthodes s'inscrivant dans l'approche quasi-expérimentales tentent de traiter la question du bien de sélection sans totalement résoudre les problèmes liés aux effets individuels inobservés. Les méthodes d'appariement, si elles produisent les effets moyens des traitements, l'existence d'effets inobservables, même avec un échantillon de grande taille restent inadaptés. Globalement, avec ces méthodes, il subsiste un biais, même si celui-ci n'est pas trop important, mais peut constituer une partie relativement significative de l'effet du traitement. L'usage des méthodes quasi-expérimentales dépend fortement également du contexte de l'évaluation. Ce contexte est déterminant dans le choix du type de méthode pouvant réduire de manière significative les biais spécifiques auxquels on peut être confronté.

Face aux limites des évaluations quasi-expérimentales, il s'est manifesté un regain d'intérêt ces dernières années pour les expérimentations aléatoires dans l'analyse des programmes et politiques de développement dans les pays à faibles revenus. La pertinence scientifique, la qualité et l'aisance de présentation et de compréhension des résultats de ces expérimentations ont suscité un intérêt auprès des partenaires au développement. Elles sont de plus en plus sollicitées par les organismes internationaux et les grandes organismes non gouvernementaux et certaines fondations pour évaluer leurs actions en faveur du développement. Cependant, elle n'est pas exempte de critiques. Ces évaluations posent des problèmes éthiques et les résultats dépendent fortement du contexte environnemental comme dans l'évaluation quasi-expérimentales. De plus, les coûts élevés de réalisation de ces évaluations font qu'elles sont hors de portée des gouvernements des pays à faibles revenus et pour les IMF

de ces pays. Face à la rareté des ressources, de telles évaluations coûteuses ne peuvent être réalisées par des gouvernants qui ont d'autres priorités et dont l'agenda politique qui ne coïncide pas avec le temps long que nécessite ces évaluations pour obtenir des résultats.

Au-delà de toutes ces difficultés que présentent ces méthodes, l'apport des évaluations est nécessaire pour une meilleure orientation des programmes et politiques de développement, et une meilleure allocation des ressources rares. En concentrant l'analyse que sur la revue des méthodes, nous avons occulté l'usage de ces méthodes dans les travaux empiriques en microfinance pour la simple raison que de telles revues ont été réalisées dans d'autres travaux (Hulme, 2000 ; Hermes et Lensink, 2007, etc.).

References :

1. Abadie, A. and Imbens, G.W. (2006), "Large sample properties of matching estimators for average treatment effects", *Econometrica*, Vol. 74, pp. 235–267.
2. Banerjee, A. V et Duflo, E. (2009), "L'approche expérimentale en économie du développement", *Revue d'économie Politique*, Vol. 119 No. 5, pp. 691–726.
3. Becker, S.O. and Ichino, A. (2002), "Estimation of average treatment effect based on propensity scores", *Stata Journal*, Vol. 2 No. 4, pp. 358–377.
4. Caliendo, M. and Kopeinig, S. (2008), "Some Practical Guidance for the Implementation of Propensity Score Matching", *Journal of Economic Surveys*, Vol. 22 No. 1, pp. 31–72.
5. Cattaneo, M.D. (2010), "Efficient semiparametric estimation of multi-valued treatment effects", *Journal of Econometrics*, Vol. 155 No. 2, pp. 138–154.
6. Cattaneo, M.D., Drukker, D.M. and Holland, A.D. (2013), "Estimation of multivalued treatment effects under conditional independence", *Stata Journal*, Vol. 13 No. 3, pp. 407–450.
7. Charpail, C., Klein, T. and Serge, Z. (2005), "Évaluation des politiques d'emploi: la deuxième génération des Panels des bénéficiaires", DARES, Ministère des affaires sociales, du travail et de la solidarité, *Document de Travail*, France.
8. Coleman, B.E. (1999), "The impact of group lending in Northeast Thailand", *Journal of Development Economics*, Vol. 60, pp. 105–141.
9. Crump, R., Hotz, J., Imbens, G.W. and Mitnik, O. (2008), "Nonparametric tests for treatment effect heterogeneity", *The Review of Economics and Statistics*, Vol. 90 No. 3, pp. 389–405.
10. Duflo, E. (2005), "Évaluer l'impact des programmes d'aide au développement: le rôle des évaluations par assignation aléatoire",

- Revue d'économie du Développement*, Vol. 19 No. 2, pp. 185–226.
11. Fay, R.G. (1996), “Enhancing the Effectiveness of Active Labour Market Policies”, *OECD Labour Market and Social Policy Occasional Papers* N°18.
 12. Fougère, D. (2010), “Les méthodes économétriques d'évaluation”, *RFAS*, No. 1999, pp. 105–128.
 13. Fouillet, C., Guérin, I., Morvant-Roux, S., Roesch, M. and Servet, J.M. (2007), “Le microcrédit au péril du néolibéralisme et de marchands d'illusions : manifeste pour une inclusion financière socialement responsable”, *Revue Mauss*, Vol. 29 No. 1, pp. 329–350.
 14. Galasso, E. and Ravallion, M. (2004), “Social Protection in a Crisis: Argentina's Plan Jefes y Jefas”, *World Bank Economic Review*, Vol. 18 No. 3, pp. 367–399.
 15. Harry, M. (1997), *La médecine des preuves. histoire et anthropologie des essais cliniques (1900-1990)*, Synthélabo-Les empêcheurs de penser en rond, Paris.
 16. Hashemi, S.M., Schuler, S.R. and Riley, A.P. (1996), “Rural credit programs and womens empowerment in Bangladesh”, *World Development*, Vol. 4 No. 24, pp. 635–654.
 17. Heckman, J.J. and Pinto, R. (2018), “Unordered Monotonicity”, *Econometrica*, Vol. 86 No. 1, pp. 1–35.
 18. Heckman, J.J., Schmierer, D. and Urzua, S. (2010), “Testing the correlated random coefficient model”, *Journal of Econometrics*, Vol. 158 N° 2, pp. 177–203.
 19. Heckman, J.J., Urzua, S. and Vytlacil, E. (2006), “Understanding instrumental variables in models with essential heterogeneity”, *Review of Economics and Statistics*, Vol. 88 No. 3, pp. 389–432.
 20. Heckman, J.J., Urzua, S. and Vytlacil, E. (2008), “Instrumental variables in models with multiple outcomes: The general unordered case”, *Annales d'économie et de Statistique*, Vol. 91/92, pp. 151–174.
 21. Hermes, N. and Lensink, R. (2007), “Impact of Microfinance: A Critical Survey”, *Economic and Political Weekly*, pp. 462–465.
 22. Hirano, K. and Imbens, G.W. (2004), “The propensity score with continuous treatments”, in A. Gelman, et X.M. (Ed.), *Applied Bayesian Modeling and Causal Inference from Incomplete Data Perspectives*, Wiley-Black., pp. 73–84.
 23. Hulme, D. (2000), “Impact Assessment Methodologies for Microfinance: Theory , Experience and Better Practice”, *World Development*, Vol. 28 No. 1, pp. 79–98.
 24. Hulme, D. et Mosley, P. (1996), *Finance Against Poverty*, Vol. 2., Routledge.
 25. Imbens, G.W. (2000), “The role of the propensity score in estimating

- dose-response functions”, *Biometrika*, Vol. 87 No. 3, pp. 706–710.
26. Imbens, G.W. et Angrist, J.D. (1994), “Identification and Estimation of Local Average Treatment Effects”, *Econometrica*, Vol. 62 No. 2, p. 467.
 27. Jatteau, A. (2013), *Les expérimentations Aléatoires en économie*, Repères; La Découverte, Paris.
 28. Karlan, D.S. and Goldberg, N. (2007), “Impact Evaluation for Microfinance: Review of Methodological Issues”, *World Bank, Doing Impact Evaluation Series N° 7*.
 29. Khandker, S.R., Koolwal, G.B. and Samad, H.A. (2010), *Handbook on Impact Evaluation Quantitative Methods and Practices*, World Bank.
 30. Kono, H. and Takahashi, K. (2010), “Microfinance revolution: its effects, innovations, and challenges”, *The Developing Economies*, Vol. 48 No. 1, pp. 15–73.
 31. L’Horty, Y.L. et Petit, P. (2010), “Évaluation aléatoire et expérimentations sociales”, Centre d’Études de l’Emploi, *Document de Travail*, Vol. 135 No. 12, p. 29.
 32. Labrousse, A. (2010), “Nouvelle économie du développement et essais cliniques randomisés : une mise en perspective d’un outil de preuve et de gouvernement”, *Revue de la régulation*, Vol. 7 No. 1, pp. 1–32.
 33. Lee, S. et Salanié, B. (2018), “Identifying effects of multivalued treatments”, Center for Microdata and Practice (CEMMAP), *Working Paper CWP34/18*.
 34. Mersland, R. and Storm, R.O. (2010), “Microfinance Mission Drift?”, *World Development*, Vol. 38 No. 1, pp. 28–36.
 35. Monnier, É. (1992), *Évaluations de l’action des pouvoirs publics*, Economica, Paris.
 36. Ravallion, M. (2008), “Evaluating Anti-Poverty Programs”, *Handbook of Development Economics*, Elsevier., Vol. 4, Schultz, Paul, Strauss, John, North- Holland, Amsterdam, pp. 3788–3840.
 37. Rodrik, D. (1998), “Who needs capital-account convertibility?”, Unpublished.
 38. Rubin, D.B. and Rosenbaum, P.R. (1983), “The central role of the propensity score in observational studies for causal effects”, *Biometrika*, Vol. 70 No. 1, pp. 41–55.
 39. Weber, O. (2013), “Impact measurement in microfinance: Is the measurement of the social return on investment an innovation in microfinance?”, *Journal of Innovation Economics & Management*, Vol. 11 No. 1, pp. 149–171.

Annexe: Résumé de quelques avantages et limites des approches

Approche non expérimentale	Approche expérimentale ou randomisation
Points positifs	Points positifs
<p>Résultats: Elle permet de déterminer les différences moyennes des résultats entre bénéficiaires et non bénéficiaires, mais aussi l'effet marginal sur les bénéficiaires.</p> <p>Connaissance : Elle permet de renforcer et d'améliorer les institutions évaluées.</p> <p>Coût : Coût relativement inférieur aux évaluations aléatoires.</p>	<p>Résultats : Précision et pertinences des résultats de l'évaluation et facilité d'interprétation et de compréhension.</p> <p>Élimination du biais de sélection ;</p> <p>Connaissance du terrain d'étude;</p> <p>Production des données par l'évaluateur.</p> <p>Connaissances : elle permet d'améliorer la qualité des services et d'adapter les services aux besoins des populations.</p>
Points faibles ou limites	Points faibles ou limites
<p>1. Estimations diverses : Les méthodes d'estimation sont diverses et difficiles à comprendre pour le décideur. De plus, les résultats peuvent varier en fonction de la spécification du modèle surtout lorsque la théorie propose plusieurs alternatives.</p> <p>2. Complexité : L'usage de techniques économétriques variées requiert des hypothèses fortes sur la distribution du terme d'erreur.</p> <p>3. Effets d'entrée du programme: Cela se produit lorsque la population inscrite dans le programme et le groupe de contrôle ne sont pas représentatifs de la population qui serait touchée si celui-ci était un programme national en cours.</p> <p>4. Le problème du groupe de contrôle : Les estimations sont sensibles à la sélection du groupe de contrôle est sélectionné. Il est souvent difficile d'obtenir un bon groupe de comparaison dans les enquêtes générales, notamment lorsque le programme porte sur un petit groupe dont la représentativité est très faible dans l'échantillon.</p> <p>5. Biais de contagion: Il est fonction des données du groupe de comparaison. Les individus peuvent avoir reçu un traitement similaire au groupe de contrôle qui n'est pas pris en compte par la base de donnée utilisée par l'évaluateur.</p> <p>6. Base de données: En général, les bases de données ne sont pas construites dans l'optique de conduire une évaluation d'impact. Ainsi, l'évaluateur ne fait qu'à adapter les données à la situation par l'usage de variables <i>proxy</i>.</p> <p>7. Biais de sélection : Le biais de sélection est aléatoire et il existe une incertitude sur sa taille. Il est donc difficile de savoir quel est son niveau d'influence sur le résultat.</p> <p>8. Des ajustements non expérimentaux: On a parfois recours à des ajustements non</p>	<p>1. Éthique en question : La limitation de l'accès à un service utile au bien-être des populations érigées en groupes de contrôle en raison de l'évaluation</p> <p>2. Problèmes de mise en œuvre: La méthode est difficile à mettre en œuvre dans une institution où les individus sont réfractaires aux changements d'habitude et plus facile dans le cadre de nouveaux programmes.</p> <p>3. Incertitude de la période d'évaluation: Il est difficile de savoir combien de temps peut prendre un programme pour modifier les comportements et donc de déterminer le temps nécessaire pour conduire une évaluation.</p> <p>4. Contamination du traitement: L'évaluateur peut être tenté d'inscrire autant de personnes que possible dans l'expérience. Ce qui peut rendre le groupe différent de celui qui aurait été choisi en l'absence de traitement.</p> <p>5. Expériences multiples: Plusieurs évaluations sont nécessaires pour comprendre et mieux adapter les services d'un programme. Par exemple, une première évalue l'impact et une seconde analyse les services à mettre en place.</p> <p>6. Auto-sélection: Les problèmes d'auto-sélection se produisent lorsque le choix d'être traité ou non ou d'être bénéficiaire ou non n'est pas aléatoire et dépend du choix des personnes.</p> <p>7. Équilibre partiel et généralisation L'impact obtenu est partiel et non global. La généralisation des résultats de micro-dispositifs est difficile dans la mesure où le système des marchés et des prix diffère d'une zone à une autre.</p> <p>8. Contamination de l'échantillon: Elle se produit lorsque les personnes inscrites dans le traitement sont différentes de ce qu'est le groupe en l'absence de traitement.</p>

<p>expérimentaux pour corriger certains problèmes. C'est le cas des problèmes d'attrition qui peuvent rendre l'échantillon moins représentatif de la population initiale à l'origine de l'expérimentation.</p> <p>9. Biais de randomisation : Il englobe un certain nombre de biais (sélection de site, les abandons ou les décrocheurs, et l'effet <i>Hawthorne</i>).</p>	<p>9. Coût : Elles sont coûteuses en temps, en argent et en ressources humaines.</p> <p>10. Les résultats: L'effet mesuré est la différence moyenne entre les bénéficiaires et les non bénéficiaires et non la distribution de l'impact.</p> <p>11. Biais de croisement: Il survient lorsque le groupe de contrôle se confond avec le groupe de traitement ou se retrouve à un moment donné dans la zone du traitement. Cela peut entraîner des risques de contamination.</p> <p>12. Effet d'entrée du programme: Il se produit lorsque la population inscrite dans le programme et le groupe de contrôle n'est pas assez représentatif si on l'on veut étendre le programme à l'échelle nationale.</p> <p>13. Biais de substitution: Il se produit lorsque le groupe de contrôle participe à un programme similaire comparable à celui reçu par le groupe de traitement.</p>
---	--

Source : Soro(2009) adapté de Charpail *et al.* (2005) et Fay (1996)

Notes

1. Le *double bottom line* consiste au respect des objectifs financiers mais également des objectifs sociaux notamment le bien-être des bénéficiaires des programmes de microfinance.
2. *Swayam Krishi Sangam* est une des plus grandes institutions de microcrédit en Inde et dans le monde.
3. Pour plus de détails, le lecteur peut consulter le tableau 2 et l'appendix A de Hulme (2000) sur les conditions d'utilisation ou non de ces différentes méthodes qui est une, adaptation de Montgomery *et al.* (1996).
4. Cf. *tableau en annexe*.
5. L'effet Hawthorne décrit la situation où le sujet modifie son comportement, car il est conscient qu'il fait l'objet d'une expérimentation, ce qui a des conséquences sur sa motivation et sa productivité. Dans ce cas les résultats obtenus ne sont plus dû aux facteurs expérimentaux, mais à la prise de conscience des sujets qu'ils sont testés.
6. Traduit du terme anglais: *strong ignorability*.
7. *Conditional independence assumption*.
8. *Difference in Difference Method*.
9. On peut citer les études de: Barnes *et al.* (2001) en Ouganda, Dunn et Arbuckle (2001) au Pérou, Mosley (2001) en Bolivie, en Thaïlande et les études de cas de *The United Nations Capital Development Fund* (l'UNCDF) réalisées en 2003 en Haïti, au Kenya, au Malawi et au Nigeria sur le secteur de la microfinance.
10. Cité par Khandker *et al.* (2010).
11. *Jameel Poverty Action Lab (J-PAL)*.
12. Les sceptiques sont les tenants de l'approche structurelle des politiques publiques (Dani Rodrik par exemple.) et les défenseurs sont les adeptes de la nouvelle génération d'économistes du développement *New Development Economists* (Banerjee A; Duflo E., etc.).
13. Voir Jatteau (2013); Miguel E. *et*, Kremer M. (2004), «Worms: identifying impacts on education and health in the presence of treatment externalities », *Econometrica*, vol. 72, n° 1, p. 159-217.
14. Cela est possible avec les enseignements produits par la recherche sur les effets des traitements hétérogènes (Crump *et al.*, 2008; Heckman *et al.*, 2010)