École Doctorale des Sciences Économiques, Juridiques, Politiques et de Gestion
Centre d'Études et de Recherches sur le Développement International (CERDI)
Université Clermont Auvergne, CNRS, IRD, CERDI, F-63000 Clermont-Ferrand, France

NUTRITION, MIGRATION AND FAMILY ECONOMICS IN DEVELOPING COUNTRIES: THREE ESSAYS

Thèse présentée et soutenue publiquement le 17 février 2025
pour l'obtention du titre de Docteur en Sciences Économiques

Par
Adrien GOSSELIN-PALI

Sous la direction de
Théophile AZOMAHOU, Simone BERTOLI & Bity DIENE

	Membres du Jury	
Théophile AZOMAHOU	Professeur, Université Clermont Auvergne	Co-directeur de thèse
Jean-Louis COMBES	Professeur, Université Clermont Auvergne	Suffragant
Sylvie DÉMURGER	Directrice de Recherche, CNRS, ENS Lyon	Rapporteure
Bity DIENE	Maître de Conférences, Université Clermont Auvergne	Co-directrice de thèse
Catherine GUIRKINGER	Professeure, Université de Namur	Rapporteure
Élodie MAÎTRE D'HÔTEL	Chargée de Rercherche, CIRAD	Suffragante
	Invité	
Simone BERTOLI	Professeur, Université Clermont Auvergne	Co-directeur de thèse

Avant-propos

Rédiger une thèse est une étape importante pour tout chercheur, offrant l'opportunité de partager la passion qui anime ses travaux, mais c'est aussi une occasion d'exprimer ses réflexions personnelles. C'est dans cet esprit que je me permets d'évoquer brièvement l'origine de mon intérêt pour les thématiques liées à l'alimentation et à la nutrition.

> *Food is life. [...] Our choices for what to eat, and how to obtain the foods we want, are among our most frequent and important decisions, both individually and for each household, community and country.*[1]
>
> — W.A. Masters, A.B. Finaret, *Food Economics: Agriculture, Nutrition and Health* (2024)

L'alimentation est un pilier essentiel de notre vie et de notre quotidien, aussi bien à l'échelle individuelle qu'au sein de nos sociétés. Chaque jour, les choix alimentaires rythment nos vies, et rares sont les aspects qui mobilisent autant notre réflexion. Bien que cette thèse n'aborde pas directement les choix alimentaires, elle s'inscrit dans une réflexion qui leur est étroitement liée.

Quand je parle de mes recherches, on me demande souvent pourquoi je m'intéresse aux questions d'alimentation et de nutrition. Je crois que c'est avant tout le désir de relier ce qui me passionne avec un travail de thèse en économie. Depuis longtemps, je suis intrigué par ce que nous consommons, par les effets de l'alimentation sur notre corps et notre cerveau, ainsi que par ses liens avec l'activité physique. Il m'a donc semblé naturel, en poursuivant mon parcours en économie du développement, de me tourner vers les questions de sécurité alimentaire, de nutrition et d'alimentation. Non seulement ce sujet m'intéresse profondément, mais je pense qu'il s'inscrit également dans une réflexion économique pertinente, notamment dans le contexte des pays en développement.

Dans cette optique, j'ai toujours cherché à faire en sorte que ce travail ne soit pas uniquement une démarche personnelle, mais qu'il contribue aussi à une réflexion sur des enjeux plus globaux. Bien que mes contributions soient modestes, ces quelques années de recherche sur ces sujets m'ont permis de mesurer l'ampleur de ces problématiques et leur importance pour le bien-être des populations les plus vulnérables. Je me rends compte également de la chance que représente le fait de mener ces études depuis une position privilégiée, où j'ai accès à une alimentation en quantité et de qualité. Ce décalage entre ma réalité et celle des individus que j'étudie m'incite à aborder ces sujets avec encore plus de rigueur et d'humilité. Ainsi, mon engagement dépasse la production académique et vise à contribuer, même modestement, à des solutions concrètes pour améliorer la sécurité alimentaire. J'espère que, tout au long de ma carrière de chercheur, je pourrai continuer à enrichir cette réflexion et à apporter ma pierre à l'édifice, en gardant toujours en tête l'objectif de rendre la recherche utile.

[1] Traduction proposée : La nourriture, c'est la vie. [...] Nos choix alimentaires, ainsi que la manière d'obtenir les aliments que nous désirons, figurent parmi nos décisions les plus fréquentes et importantes, tant au niveau individuel que pour chaque ménage, communauté et pays.

Remerciements

Je souhaite d'abord remercier mes directeurs de thèse : Théophile Azomahou, Simone Bertoli et Bity Diene. Bity et Théophile, merci pour votre accompagnement et pour m'avoir offert une précieuse liberté intellectuelle. Je suis également reconnaissant d'avoir pu co-écrire ma première publication avec vous. Simone, votre arrivée en tant que co-directeur au cours des deux dernières années a été une véritable chance. Votre rigueur et vos analyses m'encouragent toujours à progresser et vos retours ont été essentiels à l'aboutissement de ce travail.

Je remercie chaleureusement les membres du jury, Jean-Louis Combes, Sylvie Démurger, Catherine Guirkinger et Élodie Maître d'Hôtel, d'avoir accepté d'évaluer ce travail. Votre expertise sur les thématiques abordées dans cette thèse est une grande source de valeur et d'inspiration. Je suis honoré de bénéficier de vos retours, qui me permettront d'approfondir et d'enrichir mon analyse.

Je tiens ensuite à remercier mes co-autrices. Christelle, merci d'avoir rejoint le projet, c'est un privilège de collaborer avec toi et de bénéficier de tes connaissances sur ces sujets essentiels. Elsa, je te remercie pour tes conseils, ta disponibilité et ton soutien tout au long de cette thèse.

Je suis reconnaissant envers les personnes de l'Université Clermont Auvergne et du CERDI, un environnement à taille humaine où j'ai eu la chance de travailler avec des collègues compétents et passionnés. Merci à tous les chercheurs pour les discussions constructives, les conseils avisés et l'ambiance stimulante. Pendant mes années de magistère, j'ai aussi eu le privilège de bénéficier de l'enseignement de qualité des maîtres de conférences et professeurs de l'École d'Économie, et je leur en suis reconnaissant. Un immense merci aux équipes administratives, notamment à Chantal, dont la communication met si bien en valeur notre travail, et à Johan, pour son aide et sa patience. Merci à toutes les personnes de l'Université qui, directement ou indirectement, ont contribué à rendre ces années enrichissantes. Je tiens aussi à remercier les participants aux conférences, ateliers et séminaires où j'ai eu l'opportunité de présenter mes travaux. Vos retours et suggestions ont grandement contribué à l'amélioration de mes recherches.

I am deeply grateful to Sebastian Vollmer for the opportunity to conduct a research stay in Göttingen. During my time there, I was fortunate to meet incredible people who made those months truly special. The friendships I built are very dear to me. Thank you, Bruno, Jonas, Leonard, Luisa, Marlene, Ruchira, and Wisnu, as well as all the members of the RTG, for making this experience so memorable. I always enjoy coming back to Göttingen.

Un mot particulier pour les doctorants et post-doctorants, anciens ou actuels, du CERDI, avec qui j'ai partagé cette aventure, pas toujours simple, qu'est le doctorat. Merci à tous ceux qui m'ont accompagné à un moment ou un autre : Alima, Alistair, Arouna, Aristide, Bao-We, Camille, Capucine, Chanrithy, Cheick, Dorian, Édouard, Habibou, Harouna, Hugues, Idrissa, Isaac, Ismael, Jérémy, Julie, Lompo, Mahamady, Mariz, Marine, Mohamed, Naseem, Pauline, Peter, Rachid, Vincent K., Vincent N., Yannick, ainsi qu'à tous ceux que je n'ai pas cités. Merci également à mes anciens camarades de promotion du magistère, qui ont eux aussi contribué à rendre ces années à Clermont inoubliables. Une pensée particulière pour Étienne, avec qui je suis heureux de voir notre amitié durer. Une mention spéciale aussi pour ceux avec qui j'ai tissé des liens plus proches durant mes années de doctorat. Melchior, partager ces quelques années dans le même bureau a été un vrai plaisir. Entre discussions sérieuses et productives, et d'autres beaucoup moins, mais tout aussi importantes, j'espère qu'on n'aura pas trop dérangé au fond du couloir. Aubin, merci de créer si souvent des occasions pour se changer les idées. Habibou, c'était cool de partager un temps à Göttingen avec toi et ça fait plaisir d'avoir trouvé quelqu'un qui travaille aussi sur les mêmes thématiques. Pierre et Paul, merci pour les kilomètres partagés en courant dans la nature auvergnate. Je tiens finalement à remercier toutes celles et ceux que je n'ai pas mentionnés explicitement, mais dont le soutien et les échanges ont marqué ce bout de chemin parcouru pendant ma thèse.

Merci également à mes amis les plus proches et de toujours : Arsène, Axel, Éric, Gaëtan, Jérémy, Lucas et Yohann. Les moments que nous passons ensemble sont toujours géniaux, et j'ai toujours hâte aux prochaines soirées, aux prochains bars, aux prochains concerts, aux prochaines vacances, aux prochains mariages, finalement à tous les moments qui nous rassemblent. La diversité de nos chemins de vie et de nos horizons est notre force et je suis fier de ce que chacun accomplit. Je sais que nos chemins resteront toujours liés et que notre amitié durera.

Enfin, mes derniers mots vont à ma famille, sans qui rien de tout cela n'aurait été possible. À mes parents, Christian et Corinne, pour leur amour, leur soutien indéfectible et leur confiance. Vous m'avez toujours donné la liberté de suivre mes envies et mes ambitions. Je ne pourrai jamais exprimer à quel point je vous suis reconnaissant pour tout ce que vous m'avez apporté, car aucun merci ne sera jamais à la hauteur. À mon frère et mes sœurs, Alban, Anaïs et Manon, dont les vies, les parcours et les réussites me remplissent de fierté. Merci évidemment pour votre amour, votre affection et votre présence, qui me rendent plus heureux et plus fort. À Nolan, qui me fait revivre l'insouciance de l'enfance. À Laurent, à Roman, à Joseph, et à Betty, qui nous a quittée, mais reste dans mon cœur. Et enfin, à tous les membres de ma famille plus éloignée, que je n'ai pas cités, mais dont je ressens toujours le soutien et l'amour.

Resumé

Cette thèse est composée de trois essais empiriques en économie, situés à l'intersection de l'économie du développement, de l'économie de la santé, ainsi que de l'économie de la migration et de la famille, tous liés par un thème commun : la nutrition en Afrique subsaharienne.

Le chapitre 1 constitue une introduction générale qui présente le contexte actuel de la malnutrition dans les pays en développement, les motivations de cette thèse, ainsi que les concepts transversaux communs, tels que le ménage, la mobilité et les dimensions géographiques. Il synthétise également les principaux résultats et contributions des chapitres qui suivent.

Le chapitre 2 analyse les dynamiques temporelles du double fardeau de la malnutrition (DFM) ainsi que du surpoids et de l'obésité (SUOB) à partir de données longitudinales provenant d'Afrique du Sud. Afin d'étudier si le DFM (la coexistence, au sein d'un même ménage, d'au moins une personne en surpoids ou obèse et d'une personne en insuffisance pondérale) est un phénomène transitoire ou persistant, un modèle probit dynamique à effets aléatoires est utilisé. Les résultats révèlent que le double fardeau est un phénomène transitoire, car la plupart des ménages concernés ne conservent pas ce statut lors des vagues d'enquêtes suivantes. En revanche, les ménages SUOB (avec au moins une personne en surpoids ou obèse mais aucune en insuffisance pondérale) présentent une forte persistance de leur statut, reflétant également la stabilité dans le temps du surpoids et de l'obésité au niveau individuel. À l'inverse, le caractère transitoire du DFM s'explique par la tendance des personnes en insuffisance pondérale à sortir progressivement de cette condition.

Le chapitre 3 explore l'effet de la migration sur les statuts nutritionnels des individus restés sur place, c'est-à-dire les personnes qui cohabitaient auparavant avec un migrant. À l'aide de données longitudinales et en combinant des méthodes de *matching* et de doubles différences, les résultats suggèrent que la migration interne au Ghana affecte négativement le statut nutritionnel des personnes restées dans les ménages d'origine, en particulier les enfants. Le principal canal de transmission expliquant l'impact nutritionnel négatif sur les enfants réside dans l'effet perturbateur à court terme causé par la migration, qui est susceptible d'entraîner un choc de revenu négatif. Même si les transferts de fonds ne parviennent pas toujours à compenser ces effets négatifs, ils peuvent potentiellement contribuer à améliorer la nutrition des enfants restés derrière, mais à long terme. Néanmoins, le choc initial pourrait engendrer des conséquences durables sur les trajectoires de croissance des enfants.

Le chapitre 4 examine l'impact du fosterage, une pratique répandue en Afrique subsaharienne consistant à envoyer un enfant vivre temporairement ou de façon permanente dans un ménage autre que celui de ses parents biologiques, sur le statut nutritionnel des enfants. En utilisant les mêmes données longitudinales que dans le chapitre 2, qui suit des individus en Afrique du Sud, des modèles de *machine learning* sont utilisés pour corriger les biais liés à la sélection dans le confiage et à l'attrition endogène. Les résultats montrent que le confiage améliore le statut nutritionnel des enfants en réduisant la probabilité de retard de croissance de 7 points de pourcentage, ce qui représente une diminution de 45 % par rapport à la prévalence moyenne. Cette amélioration semble s'expliquer par la relocalisation des enfants dans des ménages plus petits situés en milieu rural, généralement composés de personnes âgées, typiquement des grands-parents bénéficiant d'une pension de retraite. De plus, le confiage profite non seulement aux enfants confiés, mais aussi à la nutrition des frères et sœurs restés dans le ménage d'origine, suggérant que cette pratique peut être mutuellement bénéfique.

Summary

This dissertation is composed of three empirical essays in economics, positioned at the intersection of development economics, health economics, and migration and family economics, with a shared focus on nutrition in sub-Saharan Africa.

Chapter 1 serves as a general introduction, presenting the current context of malnutrition in developing countries, the motivations behind this thesis, and the overarching concepts such as the household, mobility, and geographic dimensions. It also provides a summary of the results and contributions of each chapter.

Chapter 2 analyzes the dynamics of the double burden of malnutrition (DBM) and overweight or obesity (OVOB) using South African longitudinal data. To investigate whether DBM (the coexistence, within the same household, of at least one overweight or obese person and one underweight person) is a transient or persistent phenomenon, a dynamic random-effects probit model is employed. The findings suggest that the double burden is a transitory phenomenon: most affected households do not retain this status in subsequent survey waves. In contrast, OVOB households (with at least one overweight or obese person but no underweight individuals) exhibit strong persistence of their status, also reflecting the long-term stability of overweight and obesity at the individual level. Conversely, the transient nature of DBM can be explained by the tendency of underweight individuals to transition out of this condition over time.

Chapter 3 investigates the effect of migration on the nutritional outcomes of the left behind—individuals who previously co-resided with a migrant. Utilizing longitudinal data from Ghana and a combination of matching and difference-in-differences, the analysis reveals that internal migration negatively impacts the nutritional status of left-behind individuals, particularly children. The primary channel driving the adverse nutritional impact appears to be the short-term disruptive effect caused by migration, likely leading to a negative income shock. While remittances do not consistently offset these negative effects, they may potentially contribute to improved outcomes for left-behind children in the long term. However, the initial shock could have lasting consequences for children's growth trajectories.

Chapter 4 examines the impact of child fostering, a widespread practice in sub-Saharan Africa in which a child is sent to live temporarily or permanently in a household different from that of their biological parents, on the nutritional status of children. Using the same longitudinal dataset as in Chapter 2, which tracks individuals in South Africa, machine learning techniques are employed to address biases related to selection into fostering and endogenous attrition. The results indicate that fostering reduces the probability of being stunted by 7 percentage points, equivalent to a 45% decrease compared to the mean prevalence. This improvement likely results from foster children moving to smaller, rural households, often including retired individuals—typically grandparents receiving a pension. Additionally, fostering not only enhances the nutritional status of foster children but also improves the nutrition of siblings who remain in the sending household, suggesting that fostering can produce mutually beneficial outcomes.

Contents

List of Tables

List of Figures

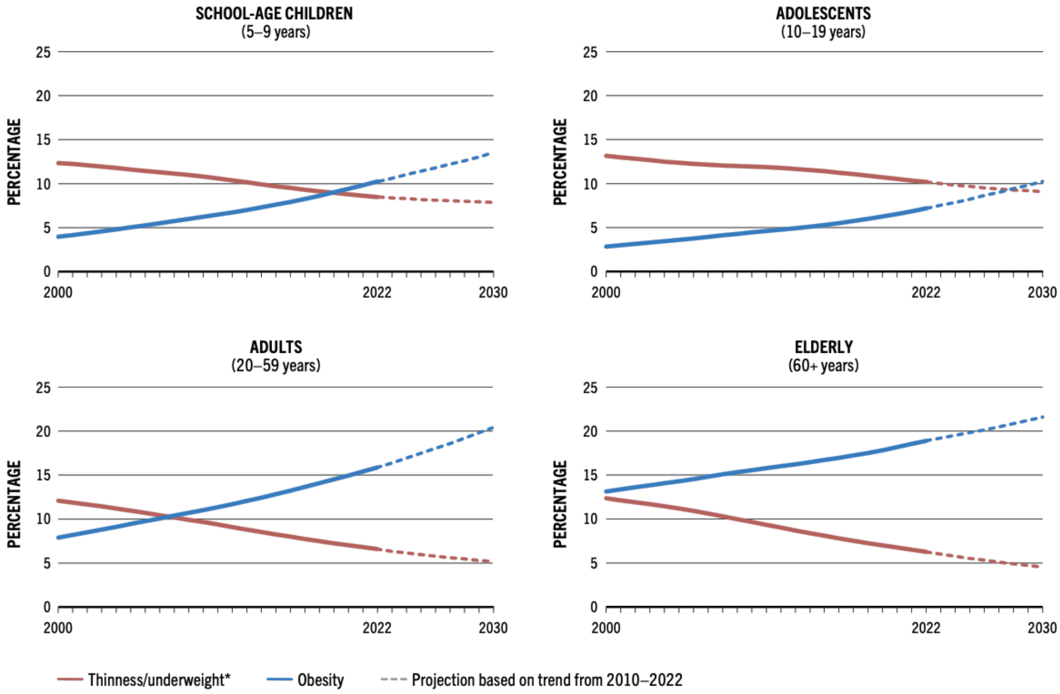
Chapter 1

General Introduction

1.1 Background and motivation

By late 2024, the world remains significantly off track in achieving Sustainable Development Goal 2, "Zero Hunger," established by the United Nations in 2015. Despite global efforts, hunger, food insecurity, and malnutrition persist (United Nations, 2024). According to the FAO's 2024 report on food security and nutrition, approximately 733 million people—roughly one in eleven globally—were undernourished in 2023, meaning they did not consume enough calories for a healthy and active life (FAO et al., 2024). Hunger levels in 2023 mirrored those of 2008-2009, reversing fifteen years of progress. As a result of conflicts, climate shocks, and the COVID-19 pandemic, the number of food-insecure people has increased by 152 million since 2019, with this impact being especially pronounced in countries of the global South. While some progress is projected in regions like Asia and Latin America, where the prevalence of undernourishment is expected to decline by 2030, the outlook for Africa is particularly concerning. In 2023, one in five people in Africa faced hunger, and projections indicate that by 2030, an additional 10 million Africans will suffer from chronic hunger, representing 18% of the continent's population. In addition, more than half of the world's hungry population is expected to be concentrated in Africa (FAO et al., 2024). Given these trends, addressing global hunger continues to be a pressing challenge.

However, hunger and undernutrition represent only one dimension of the broader picture of malnutrition. Malnutrition, in all its forms, includes not only undernutrition (wasting, stunting, underweight), but also micronutrient deficiencies, overweight, obesity, and diet-related noncommunicable diseases (WHO, 2024). Over the past decades, economic development and urbanization have led to a nutrition transition, increasing the prevalence of overweight and obesity, particularly in low- and middle-income countries (LMICs) (Popkin et al., 2020). This new nutrition reality has been described as the "double burden of malnutrition," highlighting the coexistence of undernutrition and overweight or obesity at various levels (Shrimpton and Rokx, 2012; Popkin et al., 2020; Shekar and Popkin, 2020). Figure 1.1 below clearly illustrates the phenomenon of the double burden of malnutrition: while thinness and underweight rates have been slowly decreasing since 2000, obesity rates have risen sharply, with projections indicating further increases by 2030. Furthermore, this rise in obesity is expected to primarily affect LMICs, where, by 2035, 79% of adults and 88% of children with overweight or obesity are projected to live (World Obesity Federation, 2024).



Notes: * Thinness for school-age children and adolescents; underweight for adults and elderly. Adapted from Figure 2.13 in FAO et al. (2024).

Figure 1.1: Projected trends in thinness and obesity by age group by 2030

In light of the data presented above, it is beyond dispute that malnutrition in all its forms remains a significant challenge for LMICs (Popkin et al., 2020). Addressing this issue is crucial not only due to the immediate health impacts that undernutrition can have at the individual level on mortality (Mosley and Chen, 1984; Pelletier, 1994; Black et al., 2008), but also for mitigating the long-term consequences that poor nutrition can have on human capital, social development, or economic growth (Shekar et al., 2006; Alderman et al., 2006; Grantham-McGregor et al., 2007; Victora et al., 2008; Hoddinott et al., 2013; Black et al., 2013). Additionally, the rising prevalence of overweight and obesity introduces a new dimension to this challenge, with significant economic repercussions at multiple levels. While the adverse health impacts of overweight and obesity are well established (Guh et al., 2009; Global BMI Mortality Collaboration, 2016), these conditions also lead to substantial economic costs, including lower wages (Cawley, 2004), reduced productivity (Goettler et al., 2017), and increased disability-related costs (Tremmel et al., 2017), among others. On a broader macro level, these issues place additional strain on healthcare systems and may hinder economic growth (Finkelstein et al., 2005; Cawley, 2015; Tremmel et al., 2017). Ultimately, investing in better nutrition is not only an essential marker of development but also a driver of progress (Horton et al., 2009), with potential positive spillover effects across households, communities, and nations. Conversely, persistent malnutrition, whether through undernutrition or overnutrition, threatens progress and compromises the health and well-being of future generations (Fanzo et al., 2018). Addressing malnutrition in all its forms is therefore essential not only for immediate health outcomes but also for sustainable development across multiple levels. In this context, a dissertation focused on nutrition issues in developing countries remains both timely and relevant.

1.2 Analytical framework: nutrition, household, movement, and geographic context

This section provides an overview of how nutrition is conceptualized and analyzed in this work, highlights the central role of the household and movement across the chapters, and defines the geographic scope.

The approach to nutrition Given the background and motivation outlined in the previous section, it is evident that this dissertation focuses on the issue of malnutrition in developing countries. However, this is a vast and multifaceted subject, and it is essential to outline how the chapters are connected through the common theme of nutrition, which will be discussed in more detail here.

The central theme connecting the chapters is their shared focus on anthropometric indicators and the nutritional statuses derived from these measures. Anthropometry refers to the measurement of the human body, such as height and weight, which are typically used to assess nutritional well-being. The measurements can be used to calculate indices to identify low birth weight, stunting, wasting, body mass index (BMI), head circumference for age, and acute malnutrition.[1] In this thesis, while food consumption choices are not directly examined, the nutritional outcomes used are, in part, shaped by such decisions. Overall, the chapters focus systematically on nutritional outcomes, which are widely recognized as key indicators of human health (Willett et al., 2019). Thus, while nutrition serves as the central theme of this work, it inherently underscores a broader concern for the well-being and overall health of individuals.

More specifically, Chapter 2 examines the double burden of malnutrition, introduced earlier but defined at the household level in this context. Within the same household, this phenomenon refers to the coexistence of at least one overweight or obese individual and one underweight individual. In this chapter, a nutritional status is thus defined at the household level, derived from individual nutritional statuses, which are themselves determined by anthropometric indicators.

In subsequent chapters, the focus shifts from the household level to individual nutritional outcomes. Chapters 3 and 4 analyze various anthropometric indicators, also allowing for the definition of individual nutritional statuses. The outcomes explored range from the anthropometric indicators themselves, such as BMI-for-age z-scores for children (a measure comparing a child's BMI to a reference population of the same age and sex), or the body weight of adults in Chapter 3, to specific nutritional statuses, like being considered stunted, underweight, healthy, or overweight or obese, in both Chapters 3 and 4.

The central role of the household and movement Beyond its shared focus on nutrition, this dissertation also highlights the household as another central theme. Households are fundamental units in economic analysis, particularly in food and nutrition-related decisions. These decisions are among the most basic yet frequent economic choices, involving the daily allocation of time and resources (Becker, 1965; Strauss and Thomas, 1995; Deaton, 1997). Therefore, the centrality of the household is evident throughout this research. Chapter 2 explicitly positions the household as the primary unit of analysis in examining the double burden of malnutrition. While subsequent chapters shift their focus to individual-level outcomes, they explore dynamics influencing and influenced by household structures. Indeed, in Chapter 3, which focuses on migration, although the mobility is experienced at the individual level, it often stems from collective household strategies (Stark and Lucas, 1988), with decisions and consequences

[1] The definition of anthropometric indicators is derived from the World Bank's DIME Wiki, available at: `https://dimewiki.worldbank.org/Anthropometric_Indicators`. Last accessed: 18 October 2024.

reverberating through the entire household unit. Similarly, in Chapter 4, the practice of child fostering, which in developing countries often involves informal arrangements where children are placed in the care of relatives or other caregivers, represents an arrangement negotiated between households (Ainsworth, 1996; Serra, 2009).

A last common theme emerges in Chapters 3 and 4: movement, which is intrinsically linked to household dynamics. This concept takes two different forms in this thesis. In Chapter 3, movement is reflected through migration, specifically the departure of a household member, which reorganizes family dynamics and impacts both the migrants and those who remain behind (Stark and Bloom, 1985; Stark, 1991; Antman, 2013), who are the focus of this chapter. Chapter 4 explores a different type of movement through child fostering, where children relocate between households. Although these two forms of mobility differ in their nature and scope, both chapters highlight how movement is intricately connected to household decision-making and dynamics. These movements, whether through migration or fostering, create ripple effects that can impact household composition, resource allocation, caregiving dynamics, decision-making processes, and so forth, but most importantly in the context of this work, nutritional outcomes.

The geographic scope The geographic focus of this dissertation is centered on developing countries, broadly classified by the World Bank as low- and middle-income countries (LMICs), or, geographically speaking, as countries in the global South. More specifically, through its three case studies, the thesis concentrates on countries in sub-Saharan Africa. In particular, Chapters 2 and 4 draw on South Africa as a case study, while Chapter 3 focuses on Ghana. This regional focus is driven by several considerations. First, as emphasized earlier, sub-Saharan Africa is likely to remain in the coming years and decades, the continent most affected by the double burden of malnutrition, whether it be persistent hunger or given the rising prevalence of overweight and obesity. Second, as will be further detailed when discussing the case studies of the different chapters, the countries examined are particularly relevant in illustrating the various research questions addressed. Finally, given the empirical nature of this dissertation, the choice is also largely dictated by the availability of data for the countries selected. For these reasons, while the title of this thesis refers to developing countries broadly, the focus will primarily be on sub-Saharan Africa.

1.3 Chapters overview and contributions

This dissertation comprises three essays in development economics, all of which share a common focus on themes related to the nutrition of individuals in sub-Saharan Africa and utilizing longitudinal data. However, each chapter stands on its own, with its own introduction, data set, empirical strategy, discussion, and results.

This dissertation makes several contributions to the existing literature. At a broader level, all chapters enhance our understanding of the impacts of various economic and social phenomena on nutritional outcomes in sub-Saharan Africa, while contributing to multiple strands of literature in development economics, health and nutrition economics, and family and migration economics. In particular, each chapter addresses specific gaps in its respective field of research, offering unique contributions. These will be outlined in the following paragraphs, along with a summary of the chapters.

Chapter 2 *"Transition and Persistence in the Double Burden of Malnutrition and Overweight or Obesity: Evidence from South Africa"* (joint with Théophile Azomahou and Bity Diene) examines the dynamics of the double burden of malnutrition (DBM) at the household level in South Africa. Using five survey waves of the National Income Dynamics Study (NIDS) and a dynamic random-effects probit model, this chapter analyzes how household and individual nutritional statuses evolve over time. The results suggest that while the double burden is transitory, with most DBM households (those with at least one underweight and one overweight/obese individual) not maintaining this status across survey waves, households with at least one overweight or obese member (OVOB households) tend to persist in this state. Furthermore, DBM households are more likely to transition toward having only overweight or obese members in subsequent periods, inducing that underweight individuals do not remain in this condition over time. Therefore, these household-level dynamics stem from individual-level dynamics. More specifically, what this chapter also shows is that while underweight individuals do not necessarily remain so over time, overweight and obese individuals tend to stay in that condition for extended periods. Additionally, underweight individuals have a high likelihood of becoming overweight or obese over time. These patterns collectively contribute to the increasing risks associated with the obesity pandemic.

This chapter makes several contributions to the literature. i) It provides the first empirical analysis of DBM dynamics at the household level, establishing causal relationships in nutritional transitions. While existing research acknowledges the double burden, the mechanisms driving its dynamics remain underexplored. This study helps fill this gap by examining these transitions empirically. ii) By focusing on household-level analysis while examining individual-level implications, we uncover how the transitory and persistence patterns of nutritional statuses drive household-level patterns. iii) The use of panel data also represents a methodological contribution to a literature dominated by cross-sectional analyses, allowing us to control for unobserved heterogeneity. iv) Applied to South Africa, a country advanced in the nutrition transition, this approach offers insights into potential future patterns in other sub-Saharan African countries as they may undergo similar issues.

Chapter 3 *"Migration and Nutrition of the Left Behind: Evidence from Ghana"* analyzes the effects of internal migration on the nutritional status of left-behind individuals in Ghana. The empirical strategy relies on two waves of panel data from the Ghana Socioeconomic Panel Survey (GSPS) covering the period 2013/2014-2017/2018. The identification strategy combines kernel matching with a difference-in-differences approach. Findings show that migration has a negative effect on the nutritional status of left-behind individuals, particularly pronounced among children who experience a deterioration in their nutritional status. Analysis of mechanisms suggests that this negative effect primarily operates through the short-term negative income shock following the migrant's departure, often referred to as the disruptive effect of migration, although remittances may have a positive offsetting effect in the longer run.

The contributions of this chapter are varied and address key gaps identified in the literature. i) It sheds new light on the heterogeneous effects of migration documented in the literature by distinguishing between short- and long-term mechanisms, particularly the role of initial income shocks and subsequent received transfers, thereby reconciling mixed results found in the literature. ii) It provides a detailed analysis of effect heterogeneity by considering all household members and documenting heterogeneities across individual characteristics (gender, age, initial nutritional status, parental migration). iii) It develops a rigorous identification strategy exploiting the longitudinal dimension of the data to control for both household and individual-level selection bias related to migration. iv) It extends the limited evidence on internal migration and on sub-Saharan Africa, and Ghana in particular, where research on the nutritional impacts of internal migration has been notably scarce.

Chapter 4 *"Child Fostering and Nutrition in South Africa"* (joint with Christelle Dumas and Elsa Gautrain) investigates the effects of child fostering on the nutritional status of both foster children and those remaining in the household of origin. Similar to Chapter 2, this chapter uses a longitudinal dataset from the NIDS in South Africa, combined with machine learning techniques, to address both selection into fostering and endogenous attrition. The results indicate that fostering significantly reduces the child's probability of being stunted. This positive impact on child nutrition operates through two main channels: improved living conditions in host households and caregiving arrangements that enhance the child's well-being. Additionally, the findings suggest that fostering creates a mutually beneficial dynamic, as children who remain in the household of origin also experience improved nutritional outcomes, primarily due to reduced household size and lower competition for resources.

While assessing the effects of fostering on children's nutrition in South Africa, this chapter also seeks to make methodological contributions to the literature by addressing gaps in previous research. i) This study capitalizes on longitudinal data that track individuals, allowing for the observation of pre- and post-fostering characteristics, a dual perspective that is rarely achieved in the literature. ii) Even when pre- and post-fostering information is available, it is often drawn from non-representative datasets or small sample sizes. In contrast, the data used in this chapter are nationally representative and based on a larger sample, enhancing the generalizability of the findings. iii) By leveraging the specific survey design of the NIDS, this study identifies exogenous instruments to address endogenous attrition. iv) The chapter also provides robust causal evidence on the impact of child fostering by employing a double machine learning (DML) framework, which effectively addresses both selection into fostering and survey attrition. More broadly, it introduces the recent DML methodology to the literature, offering reliable results in the context of double selection issues. v) Finally, this study sheds light on the indirect effects of fostering for non-foster children remaining in the sending household.

References of Chapter 1

Ainsworth, M. (1996): "Economic aspects of child fostering in Cote d'Ivoire," *Research in Population Economics*, 8, 25–62.

Alderman, H., J. Hoddinott, and B. Kinsey (2006): "Long term consequences of early childhood malnutrition," *Oxford Economic Papers*, 58, 450–474.

Antman, F. M. (2013): "The impact of migration on family left behind," *International Handbook on the Economics of Migration*, 293–308.

Becker, G. S. (1965): "A Theory of the Allocation of Time," *The Economic Journal*, 75, 493–517.

Black, R. E., L. H. Allen, Z. A. Bhutta, L. E. Caulfield, M. De Onis, M. Ezzati, C. Mathers, J. Rivera, Maternal, and C. U. S. Group (2008): "Maternal and child undernutrition: Global and regional exposures and health consequences," *The Lancet*, 371, 243–260.

Black, R. E., C. G. Victora, S. P. Walker, Z. A. Bhutta, P. Christian, M. d. Onis, M. Ezzati, S. Grantham-McGregor, J. Katz, R. Martorell, and R. Uauy (2013): "Maternal and child undernutrition and overweight in low-income and middle-income countries," *The Lancet*, 382, 427–451.

Cawley, J. (2004): "The Impact of Obesity on Wages," *Journal of Human Resources*, 39, 451–474.

——— (2015): "An economy of scales: A selective review of obesity's economic causes, consequences, and solutions," *Journal of Health Economics*, 43, 244–268.

Deaton, A. (1997): *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy*, Washington, DC: The World Bank.

Fanzo, J., C. Hawkes, E. Udomkesmalee, A. Afshin, L. Allemandi, O. Assery, P. Baker, J. Battersby, Z. Bhutta, K. Chen, C. Corvalan, M. Di Cesare, C. Dolan, J. Fonseca, L. Grummer-Strawn, C. Hayashi, J. McArthur, A. Rao, C. Rosenzweig, and D. Schofield (2018): *2018 Global Nutrition Report: Shining a light to spur action on nutrition.*

FAO, IFAD, UNICEF, WFP, and WHO (2024): *The State of Food Security and Nutrition in the World 2024 – Financing to end hunger, food insecurity and malnutrition in all its forms.*, The State of Food Security and Nutrition in the World (SOFI), Rome: FAO; IFAD; UNICEF; WFP; WHO.

Finkelstein, E. A., C. J. Ruhm, and K. M. Kosa (2005): "Economic causes and consequences of obesity," *Annual Review of Public Health*, 26, 239–257.

Global BMI Mortality Collaboration (2016): "Body-mass index and all-cause mortality: individual-participant-data meta-analysis of 239 prospective studies in four continents," *The Lancet*, 388, 776–786.

Goettler, A., A. Grosse, and D. Sonntag (2017): "Productivity loss due to overweight and obesity: a systematic review of indirect costs," *BMJ Open*, 7, e014632.

Grantham-McGregor, S., Y. B. Cheung, S. Cueto, P. Glewwe, L. Richter, and B. Strupp (2007): "Developmental potential in the first 5 years for children in developing countries," *The Lancet*, 369, 60–70.

Guh, D. P., W. Zhang, N. Bansback, Z. Amarsi, C. L. Birmingham, and A. H. Anis (2009): "The incidence of co-morbidities related to obesity and overweight: A systematic review and meta-analysis," *BMC Public Health*, 9, 88.

Hoddinott, J., J. R. Behrman, J. A. Maluccio, P. Melgar, A. R. Quisumbing, M. Ramirez-Zea, A. D. Stein, K. M. Yount, and R. Martorell (2013): "Adult consequences of growth failure in early childhood," *The American Journal of Clinical Nutrition*, 98, 1170–1178.

Horton, S., M. Shekar, C. McDonald, A. Mahal, and J. K. Brooks (2009): *Scaling Up Nutrition: What Will It Cost?*, World Bank Publications.

Mosley, W. H. and L. C. Chen (1984): "An Analytical Framework for the Study of Child Survival in Developing Countries," *Population and Development Review*, 10, 25–45.

Pelletier, D. L. (1994): "The Relationship Between Child Anthropometry and Mortality in Developing Countries: Implications for Policy, Programs and Future Research1," *The Journal of Nutrition*, 124, 2047S–2081S.

Popkin, B., C. Corvalan, and L. M. Grummer-Strawn (2020): "Dynamics of the double burden of malnutrition and the changing nutrition reality," *The Lancet*, 395, 65–74.

Serra, R. (2009): "Child fostering in Africa: When labor and schooling motives may coexist," *Journal of Development Economics*, 88, 157–170.

Shekar, M., R. Heaver, and Y. K. Lee (2006): *Repositioning Nutrition as Central to Development: A Strategy for Large Scale Action*, New York, NY: World Bank Publications.

Shekar, M. and B. Popkin (2020): *Obesity: Health and economic consequences of an impending global challenge*, The World Bank.

Shrimpton, R. and C. Rokx (2012): *The Double Burden of Malnutrition: A Review of Global Evidence*, World Bank.

Stark, O. (1991): *The Migration of Labor*, Cambridge, MA: Basil Blackwell.

Stark, O. and D. E. Bloom (1985): "The New Economics of Labor Migration," *American Economic Review*, 75, 173–178.

Stark, O. and R. E. B. Lucas (1988): "Migration, Remittances, and the Family," *Economic Development and Cultural Change*, 36, 465–481.

Strauss, J. and D. Thomas (1995): "Chapter 34 Human resources: Empirical modeling of household and family decisions," in *Handbook of Development Economics*, Elsevier, vol. 3, 1883–2023.

Tremmel, M., U.-G. Gerdtham, P. M. Nilsson, and S. Saha (2017): "Economic Burden of Obesity: A Systematic Literature Review," *International Journal of Environmental Research and Public Health*, 14, 435.

United Nations (2024): *The Sustainable Development Goals Report 2024*, New York: United Nations.

Victora, C. G., L. Adair, C. Fall, P. C. Hallal, R. Martorell, L. Richter, and H. S. Sachdev (2008): "Maternal and child undernutrition: consequences for adult health and human capital," *The Lancet*, 371, 340–357.

WHO (2024): "Fact sheets: Malnutrition," Definition of malnutrition available at: https://www.who.int/news-room/fact-sheets/detail/malnutrition. Last accessed: 10 October 2024.

Willett, W. et al. (2019): "Food in the Anthropocene: The EAT–Lancet Commission on healthy diets from sustainable food systems," *The Lancet*, 393, 447–492.

World Obesity Federation (2024): *World Obesity Atlas 2024*, London: World Obesity Federation.

Chapter 2

Transition and Persistence in the Double Burden of Malnutrition and Overweight or Obesity: Evidence from South Africa

This chapter is a joint work with Théophile Azomahou (Professor, Université Clermont Auvergne, CERDI) and Bity Diene (Associate Professor, Université Clermont Auvergne, CERDI), published in *Food Policy*, https://doi.org/10.1016/j.foodpol.2022.102303

2.1 Introduction

The double burden of malnutrition (hereafter, DBM or double burden) is defined as the coexistence of undernutrition (i.e., micronutrient deficiencies, underweight, or childhood stunting and wasting) and overweight, obesity, and diet-related non-communicable diseases.[1] Over 70% of countries face the DBM, the overwhelming majority of which are low- and middle-income countries (LMICs). In the 1990s, these were countries in Central America, Francophone Africa, South Africa, and Central Asia. Nowadays, most countries facing DBM are in sub-Saharan Africa, South Asia, and South-East Asia (Shekar and Popkin, 2020). Significant economic development and urbanization have led to a nutrition transition that has increased the prevalence of obesity in LMICs (Popkin et al., 2020). The nutrition transition is used to describe "progressive shifts in the stages of eating, drinking, and moving from traditional, nutrient-rich diets to energy-dense, nutrient-poor, ultra-processed foods, sugary calorie-laden beverages, and increased sedentary lifestyle that coincides with or is preceded by economic, demographic, and epidemiological changes" (Shekar and Popkin, 2020). Unlike developed countries, LMICs are experiencing the nutrition transition over a shorter period. The double burden is thus due to simultaneous increases in economic development and nutrition transition (Popkin, 2004; Shrimpton and Rokx, 2012). It is also driven by a decrease in undernutrition that is not proportional to the increase in overweight and obesity. As a result, undernutrition is stagnating or decreasing while obesity is increasing significantly. To investigate the double burden and its dynamics, we use the example of South Africa, where the prevalence of stunting is 27.4%, while the prevalence of overweight or obesity among women is 67.2% which is considered a "high" level (Shekar and Popkin, 2020).[2] These figures make South Africa a particularly interesting context to analyze. Indeed, it is the most affected country by the obesity epidemic in Sub-Saharan Africa and undernutrition persists, particularly in the poorest communities (Kimani-Murage et al., 2010; Otterbach et al., 2021).

There are several ways to measure the double burden depending on the scope. It can be assessed at the population level (a population with both undernutrition and overweight or obesity prevalent in the same community, region, or nation). It can also be assessed at the household level when household members are affected by different forms of malnutrition and at the individual level when individuals experience the simultaneous occurrence of two or more types of malnutrition, for instance, when obesity is coupled with micronutrient deficiency. Recent years have witnessed a growing academic interest in the household level approach, although studies referring to it started in the early 2000s (Doak et al., 2002, 2005; Garrett and Ruel, 2005). A household can be considered in DBM if there is at least one overweight or obese individual and one underweight individual in the household (Doak et al., 2005; Roemling and Qaim, 2013). Alternatively, a household can also be considered in DBM if a child suffers from undernutrition with an overweight mother (Conde and Monteiro, 2014; Jehn and Brewis, 2009; Kimani-Murage et al., 2015).

This paper aims to investigate the dynamics of two nutritional statuses (DBM and overweight or obesity (OVOB))[3] measured at the household level. However, the dynamics of nutritional status at the household

[1]Sometimes the DBM can be defined as the triple burden of malnutrition as it also encompasses micronutrient deficiencies. Here, we refer to the coexistence of undernutrition and overweight or obesity.

[2]Shekar and Popkin (2020) proposed the cut-offs: low or none, moderate, high, or very high. In the rest of Africa, the prevalence of stunting ranges from 13.6% (Mauritius) to 55.9% (Burundi) and the prevalence of overweight or obesity varies from 29% (Ethiopia) to 67.2% (South Africa).

[3]Hereafter, we use the term OVOB to refer to overweight or obese households, i.e. those with at least one overweight or obese individual in the household (individuals with a BMI ≥ 25). We will discuss this issue in more detail in Section 2.2. In this way, if we write "overweight/obese" or "overweight or obese", it implies that we do not refer to the classification at

level imply intra-household and individual behaviors. The literature on the persistence of obesity focuses on its measurement at the individual level. Persistence refers to the fact that obesity is observed over a long period of people's lives. Indeed, once obese, the body becomes used to having extra fat reserves and tries to maintain them (Rosenbaum et al., 2010). The persistence of obesity has been analyzed for American individuals, among whom most obese adults remain so for a long lifetime (Daouli et al., 2014). Individuals with persistent obesity are also those who, as children, were already obese (Gordon-Larsen et al., 2010). The persistence of overweight or obesity may also be related to genetic factors (O'Rahilly and Farooqi, 2006; Rohde et al., 2019). A double-burdened household has at least one individual who suffers from undernutrition. At the national level, the prevalence of stunting in children under five years of age has been stagnant or slowly decreasing in South Africa over the past ten years, as the prevalence of stunting was 24.9% in 2008[4] compared to 27.4%, according to the latest global report on obesity (Shekar and Popkin, 2020). Therefore, undernutrition is persistent at the national level, particularly for the most disadvantaged, lower-income, and food-insecure households. However, the story differs at the individual level, it is not the same people who continue to experience undernutrition and underweight might not persist over time. Indeed, according to Barker's hypothesis, adverse nutrition in childhood (for instance, being stunted or underweight) could increase the likelihood of developing chronic and non-communicable conditions later on, such as overweight or obesity (Edwards, 2017; Barker, 1990).

This paper focuses on the persistence and transition of nutritional status at the household level. We aim to further explore the dynamics of the DBM and OVOB at the household level using survey data from South Africa. Despite the growing literature on the DBM, relatively little research has been carried out on its dynamics over time and its evolution within households. Using household surveys in Indonesia, Roemling and Qaim (2013) show that DBM is transitory at the household level as many DBM households in one period become overweight in the subsequent period. However, almost as many households remain double-burdened as there are households that experience a change in nutritional status. Although this study paves the way to study the dynamics of the DBM, it lacks an empirical model showing the causal relationship between a nutritional status in period $t-1$ and a nutritional status in period t. This lack of empirical analysis on the dynamics of double burden is one of the motivations of our paper. Furthermore, the mechanisms that underpin the dynamics of nutritional status are not fully understood.

To analyze the dynamics of nutritional statuses, we use a dynamic random-effects probit model with unobserved heterogeneity (Rabe-Hesketh and Skrondal, 2013; Grotti and Cutuli, 2018). We draw on data from the five waves of surveys of the National Income Dynamics Study (NIDS) implemented in South Africa. We find that DBM is a transitory phenomenon as most double burden households over one survey period do not remain so in the subsequent waves. These findings are consistent with Roemling and Qaim (2013), who find that DBM is transitory in Indonesia. We also find that OVOB households remain so, implying persistence at the household level, whereas until now, obesity persistence has been observed at the individual level. Our results also show that DBM households in $t-1$ are more likely to become OVOB in the follow-up periods. These dynamics at the household level stem from individual intra-household dynamics. We observe that the persistence of OVOB at the household level is mainly driven by the persistence of overweight or obesity at the individual level. Finally, the intra-household explanation for the transition from DBM to OVOB is that underweight does not persist over time since

the household level but rather to the individual level or overweight and obesity in general.

[4] UNICEF/WHO/World Bank Joint child malnutrition estimates expanded database: stunting, wasting and overweight (May 2021). Available at: https://data.unicef.org/resources/dataset/malnutrition-data. Last accessed: 26 October 2021.

most underweight individuals become normal. On the other hand, these individuals may eventually become overweight or obese.

The contribution of this paper to the literature is threefold. Firstly, the dynamics of nutritional indicators at the household level have not yet been deepened, except for Roemling and Qaim (2013). Their conclusions are drawn from descriptive analysis. In this study, we explore this issue through econometrics inference to determine the transient nature of the DBM and the nutritional fate of these households in South Africa. Therefore, to the best of our knowledge, our paper is one of the first to empirically analyze the dynamics of the DBM at the household level. Analyzing the DBM at the household level is of particular interest since most studies are done at the individual level. The household level analysis allows us to study the differences between households composed solely of at least one overweight or obese individual (OVOB households) as opposed to households composed of at least one underweight and one overweight or obese individual (DBM households). The focus on households is also of particular interest because if we monitor the dynamics of the DBM at the national level by observing the prevalence of undernutrition and obesity, we do not necessarily understand what is happening at the household level and, therefore, at the individual level. For instance, if the prevalence of undernutrition and obesity at the national level increases, one does not know how this translates into household dynamics. Indeed, one may wonder if the same individuals remain overweight/obese over the years or if new individuals become overweight or obese. The same issue arises for undernutrition: when the prevalence of undernutrition is persistent at the national level, one may wonder if this persistence is due to the same individuals remaining underweight or to new individuals becoming underweight while others do not remain so? We believe this deserves a specific analysis at the household level to analyze later the implications of individual shifts at the household level. Secondly, the household level analysis enables us to investigate the implications for intra-household dynamics. Indeed, this additional analysis of individual implications provides insight into the composition of double-burdened and OVOB households and how individual factors translate into household level changes. Thirdly, the literature on DBM often uses cross-section data (Guevara-Romero et al., 2021). We use up-to-date panel data, thus providing an updated vision of the DBM in South Africa. Panel data are particularly essential for studying dynamics and unobservable heterogeneity. In addition, given that South Africa is a middle-income country and at a relatively advanced stage of nutritional transition with a growing obesity epidemic, the case study provides a glimpse of what might happen in other Sub-Saharan African countries in their development process.

The double burden is a growing public health concern in LMICs that requires specific policy interventions, namely double-duty actions. These aim to simultaneously tackle undernutrition, overweight, obesity, and diet-related non-communicable diseases. These initiatives may consist of new policy proposals or existing interventions that address one form of malnutrition but scale up to address the multiple forms of the DBM. Some double-duty actions are meant to be implemented in the agriculture sector since this sector can contribute to the struggle against the double burden.

The rest of the paper is structured as follows. Section 2.2 introduces the data, provides descriptive statistics, and describes how individuals and households are categorized. Section 2.3 presents the empirical strategy. Section 2.4 reviews the results. Section 2.5 implements robustness analysis. A discussion of the results and policy implications are presented in Section 2.6. Section 2.7 draws the main conclusions.

2.2 Data

2.2.1 National Income Dynamics Study (NIDS)

We use five waves of the National Income Dynamics Study (NIDS) data collected by the Southern Africa Labour and Development Research Unit (SALDRU) of the University of Cape Town.[5] NIDS surveys examine the living conditions of South African individuals and cover issues related to economic activity, poverty and well-being, participation in the labor market, education, or health. Since 2008, surveys have been conducted every two to three years and the last available wave was conducted in 2017. The survey is based on a nationally representative sample of more than 28,000 individuals living in 7,305 households. The sample is ensured to be representative by a two-stage cluster sampling design. 400 Primary Sampling Units (PSUs) were selected from a main sample of 3,000 PSUs and randomization was conducted within representative strata of the 53 districts in South Africa.

Surveys track individuals over time but do not track households. To build a panel of households, we first define when a household could be identified as the same unit over five consecutive periods. We categorized a household as the same unit if at least two individuals could be identified as residents in a given dwelling during the five waves. We relied on Harris (2016) to transform the data from a panel of individuals into a panel of households.[6] Based on this representative sample, we kept households with at least two members present from Wave 1 to Wave 5, as single-person households do not allow us to classify households according to a status. We do not consider households dissolved or formed after the first wave. We also removed households for which it was impossible to define a nutritional status in more than two waves. Finally, our sample is made up of 2,711 households spread over five waves.

2.2.2 Nutritional status

We use anthropometric measures to determine individuals' nutritional status. We use the Body Mass Index (BMI) for adults (from 19 years old in NIDS). The BMI is criticized (Ortega et al., 2016). However, it remains the most used tool. In NIDS, interviewers also measured waist circumference, which allowed us to measure the waist-to-height ratio (WHtR: waist circumference (cm)/height (cm)). To support our choice of using the BMI, we studied the correlation between BMI and WHtR, and classified individuals by nutritional status according to the WHtR and the Ashwell Shape Chart (Ashwell, 2011). The correlation matrix in Table 2.A.1 in Appendix 2.A shows that these indicators are not statistically different. Most individuals were categorized in the same status, whether using the BMI or the WHtR. Therefore, we decided to keep the BMI measure. The thresholds are: underweight for BMI ≤ 18.5 kg/m^2; normal if BMI > 18.5 kg/m^2 and < 25 kg/m^2; and overweight or obesity for BMI ≥ 25 kg/m^2. For children, we use BMI-for-age zscores. Underweight occurs when BMI-for-age $<$ -2 Standard Deviation (SD) and overweight occurs when BMI-for-age $>$ 1 SD. Children between the two thresholds are considered normal (WHO, 2006). In NIDS, anthropometric measures are performed three times. We calculate the average of the first two measurements for weight and height since the interviewers did not consider performing a third one when the first two were consistent.

[5] NIDS portal is available at: `https://www.datafirst.uct.ac.za/dataportal/index.php/catalog/NIDS`. Last accessed: 26 October 2021.

[6] For more details on the sampling design and the implementation of the database into a panel of households, the interested reader can refer to Section 4 of Harris (2016), titled *Construction a panel of households - formation and dissolution*.

To reduce the number of missing data, we made replacements as follows. Regarding the height of adults, we considered that individuals do not experience any change in height after the age of 20. According to the WHO growth charts, adolescent growth stabilizes after age 19, although this may occur earlier for girls (De Onis, 2015). We use 20 years of age as the cut-off point since even if growth has stopped before then, it can be assumed that the adult has reached his or her final height. Therefore, we replaced the heights that varied by more than 5 cm between each wave with a missing value. We also substituted the missing values for the size of adults by the average. For the weight of adults, we calculated the average of the waves framed by values and substituted the missing values by the average. We proceed similarly for children. To not bias the categories, we replaced the indicator of nutritional status of pregnant women and children under two years of age with missing values (Doak et al., 2005; Roemling and Qaim, 2013). To ensure that these corrections do not bias our results, we also performed regressions on "raw" data in which no replacement is made (Appendix 2.C, Tables 2.C.3 and 2.C.4). Then, we categorize households according to individuals' nutritional status.

In line with the literature, we consider three household nutritional categories: DBM (household with at least one underweight individual and at least one overweight or obese individual)[7], OVOB (household with at least one overweight or obese individual and a diverse number of normal individuals, but no underweight individuals) and another nutritional status (i) normal: household entirely composed of normal individuals. ii) underweight: household with at least one underweight individual and a diverse number of normal individuals, but no overweight or obese individuals) (Doak et al., 2002, 2005; Roemling and Qaim, 2013).

2.2.3 Descriptive statistics

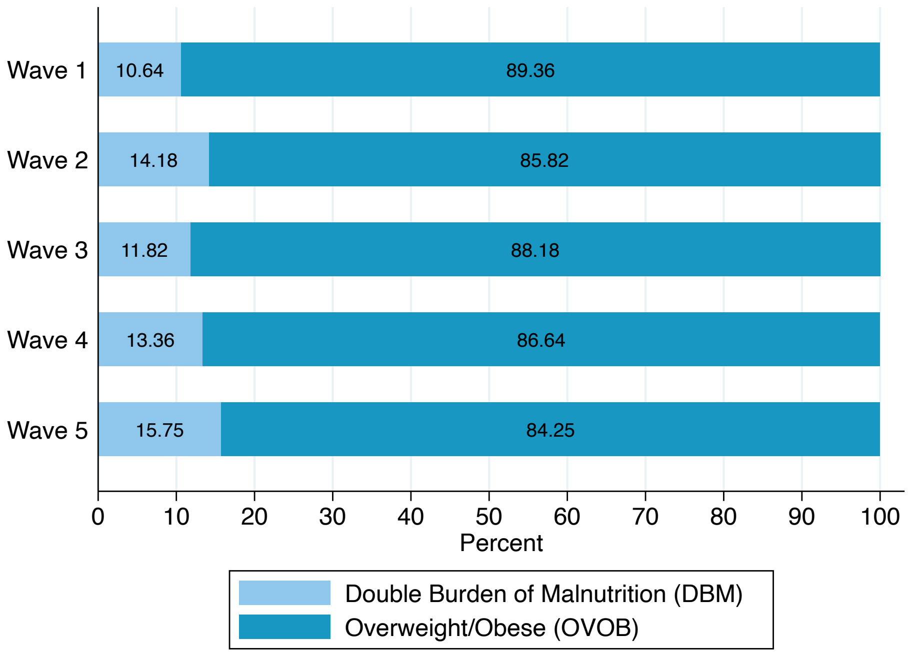
Figure 2.1 shows that about 13.76% of DBM and OVOB households are classified as DBM and 86.24% as OVOB across all waves. The rate of DBM households ranges from 10.64% to 15.75%, while the rate of OVOB households ranges from 84.25% to 89.36%. Most households are therefore OVOB. Table 2.1 displays the proportion of households that changed their nutritional status. Among DBM households, 40.89% remained so in the following survey period and 59.11% moved to the OVOB status. Among OVOB households, 90.55% were still OVOB in the next wave, while 9.45% became DBM households. The first line of Table 2.1 suggests that the DBM is a transitory phenomenon and the second line that OVOB is rather persistent.

Table 2.1: Transition matrix of DBM and OVOB households

Nutritional status of households in one survey period	Nutritional status of the same households in the subsequent period		
	DBM	OVOB	Total
DBM	40.89	59.11	100.00
OVOB	9.45	90.55	100.00

Notes: All values are in percentages.

[7]We use the definition of the double burden (which includes underweight and overweight or obesity) and not the triple burden which includes micronutrient deficiencies. Although this may be considered a narrow definition, it is frequently used in the literature (Doak et al., 2002, 2005; Roemling and Qaim, 2013; Vaezghasemi et al., 2014).



Notes: Across all waves, 13.76% are DBM households and 86.24% are OVOB households.

Figure 2.1: Distribution of DBM and OVOB households

Table 2.2 presents the descriptive statistics. It provides an overview of households' characteristics, their distribution, and the differences between DBM and OVOB households. Almost 87% of household heads have at least a primary education and OVOB households are more likely to have a better-educated household head. Less than half of household heads are inactive, but DBM households are more likely to have an inactive household head than OVOB households (49.89% versus 42.57%). Approximately 76% of household heads are African and the rest are either Coloured[8] (11.80%), Asian/Indian (2.84%), or White (9.50%). DBM households are more likely to be African-headed than OVOB households (78.99% versus 75.19%). On the other hand, White-headed households are more likely to be OVOB than DBM (10.80% versus 1.43%). The average age of the household head is about 50 years old and DBM household heads are older than OVOB household heads. About 54% of household heads are women. Over three-fifths of the households live in urban areas and there are more DBM and OVOB households in urban areas. The average household size is about 5 members and there are more children and adults in DBM households than in OVOB households. Between two waves, more households are concerned about entries of individuals in DBM households than in OVOB households. The percentage of households with exits is about the same between DBM and OVOB households. Finally, members of OVOB households tend to be older. We describe all variables in Appendix 2.B.

[8]"Coloured refers to an individual of mixed-blood that includes children/descendants from Black-White, Black-Asian, White-Asian, and Black-Coloured unions" (Tomita et al., 2015).

Table 2.2: Descriptive statistics

	Total	DBM	OVOB	t-test
Variable	Mean (SD)	Mean (SD)	Mean (SD)	P-value
Expenditures	6912.76 (544.06)	5915.97 (776.36)	7419.99 (610.10)	0.071
Education of the head				
No education*	13.32	19.00	12.27	0.000
Primary education*	17.42	18.58	17.29	0.538
Secondary education*	51.17	53.87	50.46	0.235
Higher education*	18.09	8.55	19.98	0.000
Employment of the head				
Inactive*	43.20	49.89	42.57	0.003
Unemployed*	10.29	10.45	9.55	0.568
Employed*	46.51	39.66	47.88	0.005
Ethnic group of the head				
African*	75.86	78.99	75.19	0.298
Coloured*	11.80	13.33	11.63	0.464
Asian/Indian*	2.84	6.25	2.38	0.141
White*	9.50	1.43	10.80	0.000
Female head*	54.44	54.90	56.18	0.652
Age of the head	50.21 (0.42)	51.93 (0.73)	50.58 (0.40)	0.066
Urban*	61.32	56.02	62.39	0.055
Household size	4.86 (0.09)	6.30 (0.19)	4.85 (0.10)	0.000
Entry*	32.08	45.07	31.55	0.000
Exit*	34.58	33.95	34.33	0.881
Number of children	1.57 (0.04)	2.13 (0.10)	1.57 (0.05)	0.000
Mean age of household	29.95 (0.42)	27.95 (0.56)	30.11 (0.44)	0.000

Notes: * percent; standard deviations are in parentheses. Sampling weights are applied.

2.3 Empirical strategy

We study separately two nutritional statuses (DBM and OVOB) to determine whether they are transitory or persistent. The econometric specifications presented below reflect our estimation strategies.

2.3.1 Transitory and persistent patterns

The dynamic random-effects probit model with unobserved heterogeneity takes the form:

$$y_{it} = \alpha y_{it-1} + \beta \bar{Z}_{it} + c_i + u_{it} \tag{1}$$

for household i (i = 1,...,N), and period t (t = 1,...,T_i), where y_{it} is the nutritional status (DBM_{it} or $OVOB_{it}$) of household i in t, y_{it-1} the nutritional status (DBM_{it-1} or $OVOB_{it-1}$) in $t-1$, $\bar{Z}_{it}$ a set of explanatory variables, c_i the household-specific unobserved effect, and u_{it} the idiosyncratic error term normally distributed with mean 0 and variance σ_u^2. y_{it-1} captures genuine state dependence, i.e., the effect exerted by the nutritional status of the previous period on the nutritional status in t. All explanatory variables $\bar{Z}_{it}$ are listed in Appendix 2.B. The variables are based on the literature and include several demographic and socioeconomic factors (Garrett and Ruel, 2005; Guevara-Romero et al., 2021). Income is also often used as a determinant of the DBM. Some studies report that compared to underweight households, DBM households have higher incomes (Doak et al., 2005; Tzioumis and Adair, 2014), while other studies have found no significant difference between DBM and normal households (Kosaka and Umezaki, 2017). Households whose heads have received a higher level of education are less likely to face DBM (Fongar et al., 2019; Vaezghasemi et al., 2014) and DBM households tend to be male-headed (Roemling and Qaim, 2013; Vaezghasemi et al., 2014). We use the variables related to the socioeconomic status (monthly household expenditures, educational attainment, and employment status of the household head), the household's living area (urban vs. rural), the ethnic group, gender and age of the head, and the household composition (household size, entry or exit of individuals between waves, number of children, and mean age of household members and its square).

The estimation of Eq. (1) requires further exploration. In particular, some specific features inherent to the dynamic probit model are worth discussing, including unobserved heterogeneity and initial conditions. Many unobserved factors cannot be captured by control variables such as genetic differences. Most households are composed of individuals belonging to the same family and share similar genes. There are genetic factors that increase the likelihood of being overweight or obese. Therefore, it is crucial to control for such factors. In addition, omitted variables may be correlated with observed and unobserved factors. For instance, the nutritional status of a household in the first period can be correlated with unobserved factors. Also, a household's location in the first period may be correlated with the proximity of a fast-food restaurant or nearby supermarkets containing highly processed or hyper-caloric food products. These factors influence the nutritional status of individuals and thus the household's status. As a result, unobserved heterogeneity and initial condition problem must be accounted for.

The retained specification is the Wooldridge (2005) model, which addresses these issues. To control for unobserved heterogeneity and the initial condition problem, Wooldridge (2005) proposes to include time-varying explanatory variables at each period (except the initial period). Rabe-Hesketh and Skrondal (2013) controlled for the unobserved effects by including within-unit averages computed on the time-varying explanatory variables and by augmenting the model specification with the initial period of the dependent variable and the initial period of the time-varying explanatory variables. Their model assumes the strict exogeneity of the explanatory variables. We assume that this condition is met since it is unlikely that there is a retroactive effect of the status in t on past or future explanatory variables. We also readily assume the predeterminedness of initial variables. Relying on Rabe-Hesketh and Skrondal (2013) and Grotti and Cutuli (2018), the household-specific unobserved effect takes the form:

$$c_i = \alpha_0 + \alpha_1 y_{i0} + \alpha_2 Z_{i0} + \alpha_3 \bar{Z}_l + \epsilon_i \tag{2}$$

where y_{i0} is the initial value of the dependent variable (DBM_{i0} or $OVOB_{i0}$), Z_{i0} the initial values of the time-varying explanatory variables, $\bar{Z}_l$ the within-unit averages of the time-varying explanatory variables, and ϵ_i a unit-specific time-constant error term (assumed to be normally distributed with mean 0 and variance σ_ϵ^2). The variables used to estimate unobserved heterogeneity are time-varying, which include monthly household expenditures, employment status, gender and age of the household head, and household composition variables. Controlling for unobserved heterogeneity allows us to distinguish the effects of unobserved heterogeneity from genuine state dependence of the nutritional status in $t-1$. Assuming that unobserved heterogeneity is captured by c_i, then the lagged value of the dependent variable can be interpreted as a genuine state dependence. Eq. (1) is estimated using a dynamic random effects probit model and conditional maximum likelihood estimator (Wooldridge, 2005; Rabe-Hesketh and Skrondal, 2013; Grotti and Cutuli, 2018). In this estimation, we employ mean-variance adaptive Gauss-Hermite quadrature with 12 integration points following Grotti and Cutuli (2018). We expect to observe a state dependence for the OVOB status, meaning a positive sign of the lagged variable's coefficient. We also expect to observe the transient nature of the DBM.

The dynamic random effects probit model allows computing expected transition probabilities and persistence. The probabilities are computed from the estimated coefficients and used to determine the probabilities of entering, staying in, or leaving a nutritional status. The estimated entry rate is:

$$Pr(1|0) = Pr(y_{it} = 1|y_{it-1} = 0, X) = \Phi[\gamma X] \tag{3}$$

The estimated persistent rate is:

$$Pr(1|1) = Pr(y_{it} = 1|y_{it-1} = 1, X) = \Phi[\alpha + \gamma X] \tag{4}$$

where y_{it} is the nutritional status (DBM_{it} or $OVOB_{it}$), X includes the time-constant explanatory variables, the time-varying explanatory variables, and all the variables capturing unobserved heterogeneity, γ is a vector of associated coefficients, α is the coefficient associated with y_{it-1}, and Φ represents the standard normal cumulative distribution function. The exit rate is derived from the estimated persistent rate and is computed as $1 - Pr(1|1)$. We expect to observe a high persistence rate for the OVOB status, which would provide evidence that this status is persistent. We also expect to observe a low persistence rate and a high exit rate for the DBM, which would show that the DBM is transitory.

2.3.2 What do DBM households become?

Now we study what happens to households whose status is transitory. Initially, we assume that DBM is a transitory phenomenon. Therefore, this status may be a transition to OVOB. There are two alternatives for a DBM household to change nutritional status to OVOB: the underweight individual(s) become normal or overweight/obese. In both cases, DBM households in $t-1$ will now be considered OVOB in t. However, it is also possible that the individual(s) formerly considered underweight leaves the household. If we take the example of an underweight individual who leaves, then, in t, if other individuals keep the same status, the household will be considered OVOB. However, this change will not be related to a change in the nutritional status of individuals but rather to a change in household size or composition. Therefore, to avoid bias in our estimates, we control for the household composition effect. To observe what happens to DBM households, we run the following regression:

$$OVOB_{it} = \alpha_1 OVOB_{it-1} + \alpha_2 DBM_{it-1} + \beta \bar{Z}_{it} + c_i + u_{it} \tag{5}$$

for household $i = 1,...,N$, and $t = 1,...,T_i$, where $\bar{Z}_{it}$ and c_i are the same variables as in Eqs. (1) and (2). With DBM_{it-1} and $OVOB_{it-1}$ the status of the household i in $t-1$ and, u_{it} the idiosyncratic error term normally distributed with mean 0 and variance σ_u^2. If the double burden is transient to the OVOB status, we should observe a positive sign associated with the coefficient of the variable DBM_{it-1}. As in Eq. (1), we use a dynamic random effects probit model following Grotti and Cutuli (2018) method.

2.4 Results

2.4.1 The dynamics of nutritional statuses

Table 2.3 provides the results of the average marginal effects from the estimation of Eq. (1). Column (1) reports the results for the DBM, while column (2) reports the results for OVOB. In column (1), the coefficient of DBM in $t-1$ is positive and statistically significant, meaning that the likelihood of a household to face the DBM increases by 3.9% if the household was double burdened in the previous period. The coefficient indicates genuine state dependence. It implies that a DBM household in $t-1$ is more likely to stay in DBM in t. We also observe a statistically significant and positive effect of the initial condition DBM in t_0. This indicates that a DBM household in the first period is more likely to remain so in t. Considering household characteristics, the likelihood of being in DBM decreases with the level of education. The likelihood of a Couloured or Asian/Indian household head being in DBM increases compared to an African household. Finally, the likelihood of being in DBM increases when the household lives in an urban area. In this model, we also controlled for household composition. The more individuals and children in the household, the more likely a household will be considered in DBM. Also, the likelihood of being in DBM increases with the mean age of the household up to a threshold (about 50 years old) and reverses beyond, which shows non-linearity. On the other hand, entries and exits do not have a significant impact. This shows that the household composition is probably not endogenous. The entry or exit from the household does not affect the results.
In column (2), the average marginal effect associated with OVOB in $t-1$ is positive and statistically significant. The likelihood of a household being OVOB increases by 10.2% if the household was OVOB in the previous period. Accordingly, once controlled for initial conditions and net of unobserved heterogeneity, the coefficient indicates genuine state dependence. The coefficient of the OVOB variable in t_0 is also positive and significant. The probability of being OVOB in t is lower by 2% for urban households compared to their rural counterparts. As for ethnicity, African households are more likely to be considered OVOB. The likelihood of being OVOB increases with education level. Finally, the likelihood of a household being OVOB decreases with the number of children.

Table 2.3: Average marginal effects from the dynamic probit model

	(1) DBM	(2) OVOB
DBM in $t-1$	0.039***	
	(0.012)	
DBM in t_0	0.340***	
	(0.018)	
OVOB in $t-1$		0.102***
		(0.015)
OVOB in t_0		0.426***
		(0.015)
Expenditures	0.015**	-0.010
	(0.006)	(0.007)
Primary education	0.003	0.008
	(0.010)	(0.013)
Secondary education	0.002	0.018
	(0.010)	(0.013)
Higher education	-0.025*	0.031*
	(0.013)	(0.018)
Unemployed	0.010	-0.023
	(-0.014)	(0.017)
Employed	-0.013	-0.003
	(0.009)	(0.012)
Urban	0.023***	-0.020**
	(0.008)	(0.010)
Coloured	0.027***	-0.045***
	(0.010)	(0.013)
Asian/Indian	0.109***	-0.123***
	(0.036)	(0.040)
White	-0.028	-0.068**
	(0.025)	(0.033)
Female head	-0.000	0.015
	(0.009)	(0.012)
Age of the head	-0.000	0.001
	(0.000)	(0.001)
Household size	0.017***	0.002
	(0.004)	(0.005)
Entry	0.011	-0.006
	(0.008)	(0.010)
Exit	-0.004	0.005
	(0.008)	(0.010)
Number of children	0.017***	-0.019**
	(0.006)	(0.008)
Mean age of household	0.009***	-0.003
	(0.003)	(0.003)
Mean age of household squared	-0.001**	0.000
	(0.001)	(0.000)
Initial values (Z_{i0})	Yes	Yes
Averages ($\bar{Z}_l$)	Yes	Yes
Observations	9,514	9,514
N. households	2,693	2,693
Log-likelihood	-2,743.259	-4,002.224
Wald chi2(42)	1,022.94	1,786.82
Prob > chi2	0.000	0.000

Notes: Standard deviations are in parentheses. The reference categories of the explanatory variables are no education; inactive; rural; and African.
*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table 2.4 presents the predicted probabilities of each nutritional status and the expected duration spent in a status. A household has an average probability of entering DBM of about 10.57%. The average probability of being in DBM at time t conditional of having been in the status at time $t-1$ is about 14.33%. Despite the positive and statistically significant marginal effect in column (1) of Table 2.3, these results suggest that DBM is not a persistent phenomenon as the average probability of exiting DBM is about 85.67%. Finally, the expected average duration spent in DBM is 1.17 waves, which means that a DBM household will stay in this status for an average of just over one wave. Therefore, it is very likely that most DBM households will experience a change in nutritional status over time. The second row of Table 2.4 shows that a household has an average probability of entering the OVOB category of 68.11%. OVOB households have a very high probability of staying in this category (78.47%) and, therefore, not moving out (21.53%). Lastly, the expected mean duration is about 4.65 waves, which means that an OVOB household spends almost all five waves in this status. The predicted probabilities are consistent with the results in Table 2.3: OVOB is a persistent nutritional status since a genuine state dependence is established.

Table 2.4: Predicted probabilities

Nutritional status	Dynamics			
	Entry	Persistence	Exit	Mean duration
DBM	10.57	14.33	85.67	1.17
OVOB	68.11	78.47	21.53	4.65

Notes: Probabilities can be interpreted as entry, persistence and exit rates: $Pr(1|0)$; $Pr(1|1)$; $Pr(0|1)$. Mean duration: $1/Pr(0|1)$.

Table 2.5 reports the predicted probabilities of household characteristics. When the household head has received higher education, the probability that the household exits DBM is about 88.48% compared to 85.53% for households with an uneducated head. Households that have an employed head have the highest probability of exiting the DBM and the lowest persistence. The rural area is also associated with the highest probability of exiting the DBM (86.92% versus 84.14%). White-headed households are less likely to remain in DBM (10.33%), less likely to enter (7.43%), and more likely to leave (89.67%). Finally, female-headed households are almost as likely to remain in DBM. Regarding OVOB, households whose head has a higher level of education, is inactive, an African, or lives in a rural area are more likely to remain OVOB. Whatever the category, persistence rates associated with DBM are relatively low compared to OVOB. This reinforces our initial findings on the persistence of OVOB and the transitory nature of the DBM.

Table 2.6 below summarizes estimates for rural and urban subsamples. Rural households in DBM in $t-1$ are less likely to remain in DBM than urban households. This result is consistent with the positive sign associated with the urban variable and the magnitude of the coefficient associated with DBM in $t-1$ in Table 2.3. For OVOB households, the difference in percentage points between the two areas is greater, but the results remain the same since OVOB households in rural areas are more likely to remain so.

Table 2.5: Predicted probabilities of categorical variables

Category	Dynamics			
	Entry	Persistence	Exit	Mean duration
DBM				
Education of the head				
No education	10.67	14.47	85.53	1.17
Primary education	10.93	14.80	85.20	1.17
Secondary education	10.82	14.66	85.33	1.17
Higher education	8.36	11.52	88.48	1.13
Employment status of the head				
Inactive	10.92	14.78	85.22	1.17
Unemployed	11.86	15.96	84.04	1.19
Employed	9.72	13.26	86.74	1.15
Area				
Rural	9.59	13.08	86.92	1.15
Urban	11.80	15.86	84.14	1.19
Ethnic group				
African	10.01	13.65	86.35	1.16
Coloured	12.55	16.85	83.15	1.20
Asian/Indian	20.61	26.60	73.40	1.36
White	7.43	10.33	89.67	1.11
Gender of the head				
Male	10.59	14.36	85.64	1.17
Female	10.55	14.31	85.69	1.17
OVOB				
Education of the head				
No education	66.41	77.13	22.87	4.37
Primary education	67.46	77.98	22.02	4.54
Secondary education	68.63	78.92	21.08	4.74
Higher education	70.26	80.21	19.79	5.00
Employment status of the head				
Inactive	68.50	78.80	21.20	4.72
Unemployed	65.58	76.43	23.57	4.24
Employed	68.11	78.48	21.52	4.65
Area				
Rural	69.27	79.39	20.61	4.85
Urban	66.74	77.36	22.64	4.42
Ethnic group				
African	69.42	79.55	20.45	4.89
Coloured	63.73	74.94	25.06	3.99
Asian/Indian	54.22	66.80	33.20	3.01
White	60.97	72.63	27.37	3.65
Gender of the head				
Male	66.88	77.48	22.52	4.44
Female	68.75	78.99	21.01	4.76

Notes: Probabilities can be interpreted as entry, persistence and exit rates: $Pr(1|0)$; $Pr(1|1)$; $Pr(0|1)$. Mean duration: $1/Pr(0|1)$.

Table 2.6: Average marginal effects, urban and rural sample

	Urban		Rural	
	DBM	OVOB	DBM	OVOB
	(1)	(2)	(3)	(4)
DBM in $t-1$	0.042**		0.035**	
	(0.017)		(0.017)	
OVOB in $t-1$		0.088***		0.111***
		(0.021)		(0.020)
Explanatory variables ($\bar{Z}_{it}$)[1]	Yes	Yes	Yes	Yes
Initial values (Z_{i0})	Yes	Yes	Yes	Yes
Averages ($\bar{Z}_{l}$)	Yes	Yes	Yes	Yes
Observations	4,581	4,581	4,933	4,933
N. households	1,353	1,353	1,341	1,341
Log-likelihood	-1,241.097	-1,766.870	-1,454.286	-2,196.200
Wald chi2(41)	503.07	886.11	562.32	903.52
Prob > chi2	0.000	0.000	0.000	0.000

Notes: [1] Except urban. Standard deviations are in parentheses.
*** $p < 0.01$; ** $p < 0.05$; $p < 0.1$.

2.4.2 Addressing potential endogeneity issues

We suspect the presence of endogeneity for two variables: household expenditures and the labor market status of the household head. For instance, unobserved variables could determine both nutritional status and expenditures. Also, for the employment status of the household head, individuals may be so malnourished (either under- or overweight) that this may affect their job seeking. To address potential endogeneity bias, we estimate the dynamic random-effects probit model with two alternatives: i) removing the variables suspected of endogeneity (Columns (1) and (2) of Table 2.7) and ii) keeping only the initial values of these variables (Columns (3) and (4) of Table 2.7). By including only the initial values of the presumed endogenous variables, we argue that they are predetermined since they are likely to be uncorrelated to present and future values.

Table 2.7 displays the results of the dynamic random-effects probit model, taking into account the potential endogeneity of some explanatory variables. According to columns (1) and (3), the likelihood of a household to face the double burden increases by 4.1% or 4.4% if the household was double burdened in the previous period, which is roughly similar to the main results (3.9% in Table 2.3). The magnitude of the coefficients associated with the lagged OVOB variable is slightly different, but the findings remain similar. Most coefficients of explanatory variables are also quite similar to those in Table 2.3 for both DBM and OVOB. The predicted probabilities are provided in Table 2.8 and values remain similar. As a result, even when correcting for the assumed endogeneity of some explanatory variables, the conclusions remain consistent.

Table 2.7: Average marginal effects from the dynamic probit model with endogeneity

	(1) DBM	(2) OVOB	(3) DBM	(4) OVOB
DBM in $t-1$	0.041***		0.044***	
	(0.012)		(0.013)	
DBM in t_0	0.337***		0.333***	
	(0.018)		(0.19)	
OVOB in $t-1$		0.113***		0.114***
		(0.015)		(0.016)
OVOB in t_0		0.428***		0.422***
		(0.016)		(0.016)
Primary education	-0.001	0.015	0.003	0.015
	(0.010)	(0.013)	(0.010)	(0.013)
Secondary education	-0.004	0.038***	0.005	0.030**
	(0.010)	(0.013)	(0.010)	(0.014)
Higher education	-0.036***	0.080***	-0.021	0.054***
	(0.012)	(0.016)	(0.014)	(0.018)
Urban	0.018**	0.000	0.019**	-0.012
	(0.007)	(0.010)	(0.008)	(0.010)
Coloured	0.024**	-0.034***	0.021**	-0.037***
	(0.010)	(0.013)	(0.010)	(0.014)
Asian/Indian	0.089***	-0.060*	0.096***	-0.11**
	(0.033)	(0.036)	(0.036)	(0.042)
White	-0.045**	0.011	-0.042*	-0.039
	(0.021)	(0.027)	(0.024)	(0.032)
Female head	-0.000	0.019	-0.002	0.021
	(0.009)	(0.012)	(0.010)	(0.012)
Age of the head	-0.000	0.001***	-0.000	0.001*
	(0.000)	(0.001)	(0.000)	(0.001)
Household size	0.018***	0.002	0.018***	0.001
	(0.004)	(0.005)	(0.004)	(0.005)
Entry	0.014*	-0.010	0.015*	-0.015
	(0.008)	(0.010)	(0.008)	(0.011)
Exit	-0.004	0.006	-0.004	0.006
	(0.008)	(0.010)	(0.008)	(0.010)
Number of children	0.017***	-0.018**	0.017***	-0.017**
	(0.006)	(0.008)	(0.006)	(0.008)
Mean age of household	0.010***	-0.004	0.011***	-0.006
	(0.003)	(0.003)	(0.003)	(0.004)
Mean age of household squared	-0.001**	0.000	-0.001**	0.000
	(0.001)	(0.000)	(0.001)	(0.000)
Expenditures in t_0			-0.007	0.026***
			(0.005)	(0.006)
Unemployed in t_0			-0.010	0.025*
			(0.011)	(0.014)
Employed in t_0			-0.015*	0.025**
			(0.008)	(0.010)
Initial values (Z_{i0})	Yes	Yes	Yes	Yes
Averages ($\bar{Z}_l$)	Yes	Yes	Yes	Yes
Observations	9,648	9,648	8,624	8,624
N. households	2,701	2,701	2,396	2,396
Log-likelihood	-2,801.015	-4,111.871	-2,473.374	-3,650.184
Wald chi2(33) or chi2(36)[1]	1,028.32	1,795.44	952.03	1,650.73
Prob > chi2	0.000	0.000	0.000	0.000

Notes: [1] Wald chi2(33) in Columns (1) and (2) and Wald chi2(36) in Columns (3) and (4). Standard deviations are in parentheses. The reference categories of the explanatory variables are no education; inactive; rural; and African. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table 2.8: Predicted probabilities with endogeneity

Nutritional status	Dynamics			
	Entry	Persistence	Exit	Mean duration
Columns (1) and (2)				
DBM	10.50	14.48	85.52	1.17
OVOB	67.40	78.87	21.13	4.73
Columns (3) and (4)				
DBM	10.50	14.85	85.15	1.17
OVOB	67.17	78.73	21.27	4.70

Notes: Probabilities can be interpreted as entry, persistence and exit rates: $Pr(1|0)$; $Pr(1|1)$; $Pr(0|1)$. Mean duration: $1/Pr(0|1)$.

2.4.3 The fate of DBM households

The transition matrix in Table 2.1 shows that more than half of DBM households in $t-1$ become OVOB in t. To study what happens to DBM households from one period to another, we use model (5). Table 2.9 reports the results. The coefficient of DBM in $t-1$ is positive and statistically significant, meaning that the likelihood of a household to be OVOB in t increases by 9.4% if the household was double burdened in the previous period. We also notice that the average marginal effect of OVOB in $t-1$ is slightly higher than the DBM. In other words, an OVOB household in $t-1$ is more likely to be OVOB in t than a DBM household. Based on these results, the double burden of malnutrition can be considered transient to OVOB since most households in DBM become OVOB over time.

Table 2.9: Average marginal effects

	OVOB
OVOB in $t-1$	0.138***
	(0.013)
DBM in $t-1$	0.094***
	(0.013)
Explanatory variables ($\bar{Z}_i$)	Yes
Initial values (Z_{i0})	Yes
Averages ($\bar{Z}_l$)	Yes
Observations	9,514
N. households	2,693
Log-likelihood	-3,977.368
Wald chi2(42)	1,835.96
Prob > chi2	0.000

Notes: Standard deviations are in parentheses. *** $p < 0.01$; ** $p < 0.05$; $p < 0.1$.

2.5 Robustness checks

In this section, we conducted three types of robustness. We use other anthropometric indicators for children, different household size, and "raw" data.

2.5.1 Alternative anthropometric indicators for children

Formerly, we used BMI-for-age to assess the nutritional status of children. For the purpose of robustness with alternative anthropometric measures, we use weight-for-age. The results are provided in Tables 2.10 and 2.11. The findings are similar to the main results. The likelihood of a household facing the DBM increases by 3.6% (3.9% for the main results) if the household was double burdened in the previous period. On the other hand, an OVOB household in $t-1$ is more likely to remain so in t: the likelihood of a household being OVOB increases by 12.0% (10.2% for the main results) if the household was OVOB in the previous period. Table 2.11 presents the predicted probabilities. A household has an average probability of entering the OVOB category of 65.50%, staying in of 77.71%, and moving of 22.29% (versus 68.11%, 78.47%, and 21.53% for the main results). The predicted probabilities of the DBM category are also similar to the main results. The entry rate is 9.27%, the persistence rate is 12.72%, and the exit rate is 87.28%. To check further robustness, we use nutritional status depending on the age of the children. We distinguished children as follows: for children under five, we use the weight-for-height and the BMI-for-age for children between 5 and 19 years. These results are available in Tables 2.C.1 and 2.C.2 in Appendix 2.C. The conclusions remain unchanged. We also used two additional indicators for child anthropometry: weight-for-height and height-for-age. The results remain similar and are available upon request.

Table 2.10: Average marginal effects from the dynamic probit model, weight-for-age for children

	(1) DBM	(2) OVOB
DBM in $t-1$	0.036***	
	(0.012)	
OVOB in $t-1$		0.120***
		(0.015)
Explanatory variables ($\bar{Z}_{it}$)	Yes	Yes
Initial values (Z_{i0})	Yes	Yes
Averages ($\bar{Z}_{l}$)	Yes	Yes
Observations	9,277	9,277
N. households	2,657	2,657
Log-likelihood	-2,483.362	-3893.840
Wald chi2(42)	908.20	1,845.07
Prob > chi2	0.000	0.000

Notes: Standard deviations are in parentheses. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table 2.11: Predicted probabilities, weight-for-age for children

Nutritional status	Dynamics			
	Entry	Persistence	Exit	Mean duration
DBM	9.27	12.72	87.28	1.15
OVOB	65.50	77.71	22.29	4.49

Notes: Probabilities can be interpreted as entry, persistence and exit rates: $Pr(1|0)$; $Pr(1|1)$; $Pr(0|1)$. Mean duration: $1/Pr(0|1)$.

2.5.2 Households with less than five individuals

In this check, we limit the sample to a number of individuals in each household. Household size may influence the way we categorize a status. The larger the household, the more likely it is to be composed of at least one overweight or obese individual and one underweight individual. According to the descriptive statistics in Table 2.2, the average household size is about five individuals (4.86). We remove households with more than five individuals across all five waves and estimate the dynamic probit model. The results are available in Tables 2.12 and 2.13. We find the same signs and significance. Although the magnitude of the predicted probabilities is slightly different, the results are also similar. The DBM is a transitory phenomenon: persistence rate of 9.52%; entry rate of 5.16%; and exit rate of 90.48%. On the other hand, OVOB is still persistent. The conclusions for this sample remain consistent.

Table 2.12: Average marginal effects from the dynamic probit model, household with less than five individuals

	(1) DBM	(2) OVOB
DBM in $t-1$	0.049**	
	(0.020)	
OVOB in $t-1$		0.104***
		(0.022)
Explanatory variables ($\bar{Z}_{it}$)	Yes	Yes
Initial values (Z_{i0})	Yes	Yes
Averages ($\bar{Z}_l$)	Yes	Yes
Observations	4,151	4,151
N. households	1,222	1,222
Log-likelihood	-775.063	-1,513.820
Wald chi2(42)	303.85	733.56
Prob > chi2	0.000	0.000

Notes: Standard deviations are in parentheses. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table 2.13: Predicted probabilities, households with less than five individuals

Nutritional status	Dynamics			
	Entry	Persistence	Exit	Mean duration
DBM	5.16	9.52	90.48	1.11
OVOB	71.36	81.83	18.17	5.50

Notes: Probabilities can be interpreted as entry, persistence and exit rates: $Pr(1|0)$; $Pr(1|1)$; $Pr(0|1)$. Mean duration: $1/Pr(0|1)$.

2.5.3 Raw anthropometric data

In data processing, we slightly modified the height and weight values for adults and the BMI-for-age values for children. We perform the same estimates with "raw" data to ensure that these changes do not affect the results. In this data, we do not replace missing values. The results are consistent. All results are available in Tables 2.C.3 and 2.C.4 in Appendix 2.C.

2.6 Discussion

2.6.1 OVOB: A persistent status

We find evidence of the persistence of OVOB. Some drivers that explain it are household characteristics. For example, households living in rural areas are more likely to remain OVOB, which is inconsistent with previous findings that often associate obesity with urban areas (Adeboye et al., 2012; Ziraba et al., 2009). This result may be due to the lack of resources or care options facing rural households. These factors disadvantage rural households, which would suffer from overweight or obesity. African-headed households are also the most likely to remain OVOB. In South Africa, being overweight is not perceived in the same way among ethnic groups (Cois and Day, 2015). A larger waist is often seen as a sign of beauty, prosperity, and good health among the African population. Therefore, this perception often presents in women can be a factor explaining overweight and obesity (Micklesfield et al., 2013). In addition, the higher the education of the household head, the more likely OVOB persists. This positive relationship between education and overweight or obesity is often observed in LMICs, where obesity is more prevalent in the better-off socioeconomic groups (Dinsa et al., 2012). At the household level, all these characteristics contribute to explain the persistence of OVOB.
Table 2.14 describes individuals' nutritional statuses dynamics for households that remain OVOB from one period to the next. We argue that the persistence of overweight and obesity at the individual level induces persistence at the household level. We observed that almost 84% of individuals remain overweight or obese. However, the individual dynamics differ depending on age. In households that remain OVOB, 91.78% of adults remain overweight or obese versus 55.69% for children and the remaining children become normal (44.31%). For the other individuals composing an OVOB household, i.e. normal individuals, most of them remain in this category (82.62%). As a result, intra-household dynamics contribute to explain the persistence of the OVOB status at the household level.

Table 2.14: Transition matrix of individuals' nutritional status in households that remain OVOB

Nutritional status of individuals in one survey period	Nutritional status of the same individuals in subsequent period		
	Normal	Overweight/Obese	Total
Adults			
Normal	76.63	23.37	100.00
Overweight/Obese	8.22	91.78	100.00
Children			
Normal	87.16	12.84	100.00
Overweight/Obese	44.31	55.69	100.00
Both			
Normal	82.62	17.38	100.00
Overweight/Obese	16.01	83.99	100.00

Notes: All values are in percentages.

2.6.2 The double burden: A transitory status

We find that a household with at least one overweight or obese and one underweight individual does not stay in this situation for long. For instance, the higher the household head's education, the less likely the household will face DBM. This negative relationship between DBM and education is often identified in the literature and can be explained by differences in health and nutrition knowledge (Fongar et al., 2019; Vaezghasemi et al., 2014; Kosaka and Umezaki, 2017). Like education, the urban area appears to be more conducive to the transitory nature of the DBM (Jones et al., 2016; Kosaka and Umezaki, 2017). DBM has been described as an urban phenomenon associated with a wider variety of food choices, more sedentary lifestyles, the westernization of diets, and eating environments dominated by supermarkets and fast-food restaurants (Doak et al., 2002; Roemling and Qaim, 2013). In terms of ethnicity, White households are the least likely to experience DBM for multiple periods. Finally, female-headed households are less likely to remain in the DBM category, which is consistent with the findings of Roemling and Qaim (2013) and Vaezghasemi et al. (2014).

We also find evidence that a large proportion of DBM households switch to the OVOB category from $t-1$ to t. Indeed, if a household formerly in DBM moves to the OVOB status, this implies that the underweight individual(s) have either become normal or become overweight or obese. Therefore, the change in individual status leads to the transition at the household level. To examine intra-household dynamics, we provide Table 2.15, which details changes in individual status for households transitioning from DBM to OVOB. The transition is due to underweight individuals becoming normal since 82.79% of underweight individuals in $t-1$ are now considered normal in t, whether they are children (81.56%) or adults (84.38%). On the other hand, adults are more likely to have persistent overweight or obesity than children in households that transitioned (89.13% versus 43.36%). Regarding other individuals, most normal individuals remain so (79.18%) and those who become overweight or obese are more likely to be adults (26.73% versus 15.64%). We also find a nutritional improvement for most children as 56.05%

transitioned from overweight or obese to normal.

To summarize, the individual level rationale for the transitional status of the DBM is that underweight does not persist over time as underweight individuals become normal regardless of age. It may be considered an improvement or a recovery from being underweight for some individuals. Nevertheless, it is not for others as some remain trapped in overweight and obesity. Moreover, if we refer to the Barker hypothesis, it is not excluded that underweight children who become normal develop chronic and non-communicable conditions later (Barker, 1990; Edwards, 2017). According to the Barker hypothesis, adverse nutrition in early life can increase susceptibility to the metabolic syndrome, including overweight and obesity later on. Therefore, this can contribute to the growing obesity epidemic over time.

These findings are helpful to compare our results to South Africa's national dynamics where at the macro level, while obesity is spreading, undernutrition persists. Following households over several periods, we have seen that DBM at the household level is a transitory phenomenon. At the individual level, we explain it by the fact that underweight individuals do not remain so over time. Hence, while obesity is persistent at the national level and by following individuals over multiple survey periods, on the contrary undernutrition does not persist for the same individuals. It represents a key result of our study. When undernutrition persists in a country, it is not necessarily the same individuals who are underweight throughout their lives but rather new individuals who might become underweight over time. This contributes to explain the persistence of undernutrition at the national level in South Africa, while at the individual level, it is not the same individuals who remain underweight.

Table 2.15: Transition matrix of individual's nutritional status in households that transitioned from DBM to OVOB

Nutritional status of individuals in one survey period	Nutritional status of the same individuals in subsequent period			
	Underweight	Normal	Overweight/Obese	Total
Adults				
Underweight	0.00	84.38	15.62	100.00
Normal	1.24	72.03	26.73	100.00
Overweight/Obese	0.14	10.73	89.13	100.00
Children				
Underweight	0.00	81.56	18.44	100.00
Normal	0.97	83.39	15.64	100.00
Overweight/Obese	0.59	56.05	43.36	100.00
Both				
Underweight	0.00	82.79	17.21	100.00
Normal	1.04	79.18	19.78	100.00
Overweight/Obese	0.22	20.23	79.55	100.00

Notes: All values are in percentages.

2.6.3 Policy implications

These issues call for appropriate measures to address the detrimental effects of South Africa's double burden. Our results are valid for South Africa. However, the same findings have been observed for Indonesia (Roemling and Qaim, 2013). Therefore, without generality, this issue may not be specific to these countries exclusively, and instead, it could also apply to some LMICs affected by the DBM. Overall, dealing with this issue involves implementing actions that limit malnutrition in all its forms. The concept of double-duty actions was introduced in the 2015 Global Nutrition Report (IFPRI, 2015) and then gradually adopted in the nutrition research field to address this public health issue. These actions include interventions, programs and policies that have the potential to simultaneously reduce the risk of both undernutrition and overweight, obesity, or diet-related NCDs (WHO, 2017). Double-duty actions are not necessarily new initiatives in nature, as adequate measures are already in use to tackle obesity or undernutrition separately. Nevertheless, these measures should be further pursued for DBM households.

Based on the findings stemming from the individual level that overweight/obesity is persisting while most underweight individuals do not remain so, the challenge is to improve the nutritional status of underweight individuals, while simultaneously tackling overweight and obesity. In other words, regarding undernutrition these findings suggest several courses of action to improve the individual's nutritional status while also preventing them from developing chronic and non-communicable conditions later. In addition, these research findings also point to the need to enhance the overweight or obese people's condition, i.e., helping them lose weight. Therefore, it is essential to consider all dimensions of the double burden, especially in DBM households that suffer from intra-household nutritional inequality with overweight and underweight individuals. Moreover, in households that transition from DBM to OVOB, most individuals become normal, which can be considered an improvement, while the others remain overweight or obese. However, it is likely that the previous underweight individuals, especially children, become overweight or obese (Barker, 1990; Edwards, 2017). This is particularly problematic since once overweight or obesity is established, it persists both at the household and individual levels. Policy interventions that address undernutrition or overweight/obesity should thus be careful not to interfere with the other component of the DBM and instead provide coordinated nutrition action. Several proposals of double-duty actions are highlighted by Hawkes et al. (2020), including redesigning guidance for complementary feeding practices, redesigning school feeding programs, devising new nutritional guidelines for food in and around educational institutions, scaling up nutrition-sensitive agriculture programs, and designing new agricultural and food system policies. For example, in the case of a double burden household defined as an overweight or obese mother and an underweight child, while the mother will remain obese, the child will become normal. However, since the child was malnourished, she will be more likely to become obese in the future. Acting through double-duty actions such as maternal nutrition and antenatal care will, therefore, positively affect both the mother and the child.

Regarding the obesity aspect of the double burden, public actions must target individuals who may be at risk of becoming overweight or obese and those who suffer from persistent overweight or obesity. In terms of policies targeting the entire population, some have proven to be effective, such as the sugar-content-based tax called the Health Promotion Levy implemented in 2018, which has reduced sugary drink intake, mainly through reformulation and behavior change (Essman et al., 2021). Public policies and incentives such as the sugar tax must be maintained to limit unhealthy and ultra-processed foods, especially since DBM at the household level may be caused by the fact that individuals are eating ultra-processed foods. Reducing the consumption of ultra-processed foods can be done through fiscal policies and the labeling of unhealthy foods (Reardon et al., 2021).

In South Africa, policies implemented to mitigate the obesity epidemic must be strengthened, especially for low-income households or minority groups. Our results show that households most affected by persistent OVOB are already experiencing socioeconomic inequalities: Africans, the economically inactive, the less educated, female-headed households, and those residing in rural areas. Therefore, public policies should focus on those households that are not directly targeted by the policies already implemented to mitigate the existing socioeconomic inequalities. Regarding agriculture, food systems, and food environments, a few main ideas are to promote interventions that can improve nutrition outcomes, promote diversity in food production, or include approaches to empower women in agricultural programs. The agricultural sector can play a crucial role in addressing the double burden of malnutrition by addressing inadequate access to nutrient-rich food, mainly through nutrition-sensitive agriculture (Ruel and Alderman, 2013). These measures are all the more critical since the COVID-19 pandemic is likely to increase the double burden. Indeed, changes in food choices, food shortages, lockdowns, mobility restrictions, and increased food insecurity are expected to make the situation much worse (Littlejohn and Finlay, 2021).

2.7 Conclusion

This paper studies the dynamics of two nutritional statuses at the household level: the DBM and OVOB in South Africa. Our findings can be summarized in four points. First, this study shows that the DBM is transitory within South African households. We find that a DBM household in $t-1$ has only between 9.52 and 14.85% chances of remaining so in t. These findings are robust to potential endogeneity issues, the use of robustness checks such as other anthropometric indicators, different household size, and the use of "raw" data. Our conclusions are consistent with the results of Roemling and Qaim (2013), who pointed out the DBM transitory nature at the household level. Second, we find that OVOB is a persistent status. The persistence rate is between 77.62 and 81.83%. As the persistence of OVOB at the household level has not yet been studied, this is quite a novel finding. Third, we find that DBM households in $t-1$ are more likely to be OVOB in t, rather than remaining in double burden in t. These findings are also consistent with Roemling and Qaim (2013). Fourth, these dynamics at the household level stem from individual level dynamics. The transition from DBM to OVOB is mainly explained by the fact that individuals who were underweight become normal, which can be considered a recovery from undernutrition and a general improvement. On the other hand, the persistence of OVOB at the household level is explained by a persistence at the individual level and overweight/obese individuals in DBM households remain so. Moreover, in the case of individuals with adverse nutrition in childhood, they may be more likely to become obese later on. It could imply a worsening of the obesity epidemic. These issues highlight the need to address multiple forms of malnutrition through appropriate policies such as double-duty actions. This study sets the stage for future research. Here, we considered the dynamics of the DBM at the household level analysis, which may seem surprising as one might think that an individual level model would make more sense in terms of human biology and behavioral science. However, the purpose of our paper was to examine the dynamics at the household level and then investigate the dynamics at the individual level to explore the implications. Therefore, the field of nutritional status study can also be extended to analyze deeply the dynamics of individual nutritional statuses in the framework of Markov chains. Likewise, one of the possible measures of the DBM is throughout the lifetime of individuals. Notwithstanding the time frame of our data does not allow us to measure DBM over an individual's lifetime, this deserves further study. In addition, we did not exhaustively examine the explanatory factors for the transition from DBM to OVOB and vice versa. This also requires further consideration.

References of Chapter 2

Adeboye, B., G. Bermano, and C. Rolland (2012): "Obesity and its health impact in Africa: a systematic review," *Cardiovascular Journal of Africa*, 23, 512–521.

Ashwell, M. (2011): "Charts Based on Body Mass Index and Waist-to-Height Ratio to Assess the Health Risks of Obesity: A Review," *The Open Obesity Journal*, 3.

Barker, D. J. (1990): "The fetal and infant origins of adult disease." *BMJ : British Medical Journal*, 301, 1111.

Cois, A. and C. Day (2015): "Obesity trends and risk factors in the South African adult population," *BMC obesity*, 2.

Conde, W. L. and C. Monteiro (2014): "Nutrition transition and double burden of undernutrition and excess of weight in Brazil," *The American Journal of Clinical Nutrition*, 100, 1617–1622.

Daouli, J., A. Davillas, M. Demoussis, and N. Giannakopoulos (2014): "Obesity persistence and duration dependence: Evidence from a cohort of US adults (1985–2010)," *Economics & Human Biology*, 12, 30–44.

De Onis, M. (2015): "World Health Organization Reference Curves," .

Dinsa, G., Y. Goryakin, E. Fumagalli, and M. Suhrcke (2012): "Obesity and socioeconomic status in developing countries: a systematic review," *Obesity Reviews*, 13, 1067–1079.

Doak, C., L. Adair, M. Bentley, Z. Fengying, and B. Popkin (2002): "The underweight/overweight household: an exploration of household sociodemographic and dietary factors in China," *Public Health Nutrition*, 5, 215–221.

Doak, C., L. Adair, M. Bentley, C. Monteiro, and B. Popkin (2005): "The dual burden household and the nutrition transition paradox," *International Journal of Obesity (2005)*, 29.

Edwards, M. (2017): "The Barker hypothesis," in *Handbook of Famine, Starvation, and Nutrient Deprivation: From Biology to Policy*, 191–211.

Essman, M., L. S. Taillie, T. Frank, S. W. Ng, B. M. Popkin, and E. C. Swart (2021): "Taxed and untaxed beverage intake by South African young adults after a national sugar-sweetened beverage tax: A before-and-after study," *PLOS Medicine*, 18, e1003574.

Fongar, A., T. Gödecke, and M. Qaim (2019): "Various forms of double burden of malnutrition problems exist in rural Kenya," *BMC Public Health*, 19, 1543.

Garrett, J. and M. Ruel (2005): "Stunted Child–Overweight Mother Pairs: Prevalence and Association with Economic Development and Urbanization," *Food and Nutrition Bulletin*, 26, 209–221.

Gordon-Larsen, P., N. S. The, and L. Adair (2010): "Longitudinal trends in obesity in the United States from adolescence to the third decade of life," *Obesity*, 18, 1801–1804.

Grotti, R. and G. Cutuli (2018): "Xtpdyn: A community-contributed command for fitting dynamic random-effects probit models with unobserved heterogeneity," *The Stata Journal*, 18, 844–862.

Guevara-Romero, E., V. Flórez-García, L. E. Egede, and A. Yan (2021): "Factors associated with the double burden of malnutrition at the household level: A scoping review," *Critical Reviews in Food Science and Nutrition*, 1–12.

Harris, T. (2016): "Household electricity access and household dynamics: insights into the links between electricity access and household dynamics in South Africa between 2008 and 2012," *University of Cape Town*, 37.

Hawkes, C., M. Ruel, L. Salm, B. Sinclair, and F. Branca (2020): "Double-duty actions: seizing programme and policy opportunities to address malnutrition in all its forms," *The Lancet*, 395, 142–155.

IFPRI (2015): "Global Nutrition Report 2015: Actions and accountability to advance nutrition and sustainable development." Tech. rep., International Food Policy Research Institute.

Jehn, M. and A. Brewis (2009): "Paradoxical malnutrition in mother–child pairs: Untangling the phenomenon of over- and under-nutrition in underdeveloped economies," *Economics & Human Biology*, 7, 28–35.

Jones, A. D., Y. Acharya, and L. P. Galway (2016): "Urbanicity gradients are associated with the household-and-individual-level double burden of malnutrition in Sub-Saharan Africa," *The Journal of Nutrition*, 146, 1257–1267.

Kimani-Murage, E. W., K. Kahn, J. M. Pettifor, S. M. Tollman, D. B. Dunger, X. F. Gómez-Olivé, and S. A. Norris (2010): "The prevalence of stunting, overweight and obesity, and metabolic disease risk in rural South African children," *BMC Public Health*, 10, 158.

Kimani-Murage, E. W., S. K. Muthuri, S. O. Oti, M. K. Mutua, S. van de Vijver, and C. Kyobutungi (2015): "Evidence of a double burden of malnutrition in urban poor settings in Nairobi, Kenya," *PLoS ONE*, 10.

Kosaka, S. and M. Umezaki (2017): "A systematic review of the prevalence and predictors of the double burden of malnutrition within households," *British Journal of Nutrition*, 117, 1118–1127.

Littlejohn, P. and B. B. Finlay (2021): "When a pandemic and an epidemic collide: COVID-19, gut microbiota, and the double burden of malnutrition," *BMC Medicine*, 19, 31.

Micklesfield, L. K., E. V. Lambert, D. J. Hume, S. Chantler, P. R. Pienaar, K. Dickie, J. H. Goedecke, and T. Puoane (2013): "Socio-cultural, environmental and behavioural determinants of obesity in black South African women," *Cardiovascular Journal of Africa*, 24, 369–375.

O'Rahilly, S. and I. Farooqi (2006): "Genetics of obesity," *Phil. Trans. R. S. B*, 361, 1095–1105.

Ortega, F. B., X. Sui, C. J. Lavie, and S. N. Blair (2016): "Body Mass Index, the Most Widely Used But Also Widely Criticized Index: Would a Criterion Standard Measure of Total Body Fat Be a Better Predictor of Cardiovascular Disease Mortality?" *Mayo Clinic Proceedings*, 91, 443–455.

Otterbach, S., H. R. Oskorouchi, M. Rogan, and M. Qaim (2021): "Using Google data to measure the role of Big Food and fast food in South Africa's obesity epidemic," *World Development*, 140, 105368.

Popkin, B. (2004): "The nutrition transition: an overview of world patterns of change," *Nutrition Reviews*, 62, 140–143.

Popkin, B., C. Corvalan, and L. M. Grummer-Strawn (2020): "Dynamics of the double burden of malnutrition and the changing nutrition reality," *The Lancet*, 395, 65–74.

Rabe-Hesketh, S. and A. Skrondal (2013): "Avoiding biased versions of Wooldridge's simple solution to the initial conditions problem," *Economics Letters*, 120, 346–349.

Reardon, T., D. Tschirley, L. S. O. Liverpool-Tasie, T. Awokuse, J. Fanzo, B. Minten, R. Vos, M. Dolislager, C. Sauer, R. Dhar, C. Vargas, A. Lartey, A. Raza, and B. M. Popkin (2021): "The processed food revolution in African food systems and the double burden of malnutrition," *Global Food Security*, 28, 100466.

Roemling, C. and M. Qaim (2013): "Dual burden households and intra-household nutritional inequality in Indonesia," *Economics & Human Biology*, 11, 563–573.

Rohde, K., M. Keller, L. la Cour Poulsen, M. Blüher, P. Kovacs, and Y. Böttcher (2019): "Genetics and epigenetics in obesity," *Metabolism*, 92, 37–50.

Rosenbaum, M., H. R. Kissileff, L. E. Mayer, J. Hirsch, and R. L. Leibel (2010): "Energy intake in weight-reduced humans," *Brain research*, 1350, 95–102.

Ruel, M. and H. Alderman (2013): "Nutrition-sensitive interventions and programmes: how can they help to accelerate progress in improving maternal and child nutrition?" *The Lancet*, 382, 536–551.

Shekar, M. and B. Popkin (2020): *Obesity: Health and economic consequences of an impending global challenge*, The World Bank.

Shrimpton, R. and C. Rokx (2012): *The Double Burden of Malnutrition: A Review of Global Evidence*, World Bank.

Tomita, A., C. A. Labys, and J. K. Burns (2015): "A multilevel analysis of the relationship between neighborhood social disorder and depressive symptoms: Evidence from the South African National Income Dynamics Study," *The American journal of orthopsychiatry*, 85, 56–62.

Tzioumis, E. and L. Adair (2014): "Childhood dual burden of under- and overnutrition in low- and middle-income countries: a critical review," *Food and Nutrition Bulletin*, 35, 230–243.

Vaezghasemi, M., A. Öhman, M. Eriksson, M. Hakimi, L. Weinehall, H. Kusnanto, and N. Ng (2014): "The effect of gender and social capital on the dual burden of malnutrition: a multilevel study in Indonesia," *PLoS ONE*, 9, e103849.

WHO (2006): *WHO child growth standards: length/height-for-age, weight-for-age, weight-for-length, weight-for-height and body mass index-for-age ; methods and development*, Geneva: WHO Press.

——— (2017): "Double-duty actions for nutrition: policy brief," Tech. Rep. WHO/NMH/NHD/17.2.

Wooldridge, J. M. (2005): "Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity," *Journal of Applied Econometrics*, 20, 39–54.

Ziraba, A. K., J. C. Fotso, and R. Ochako (2009): "Overweight and obesity in urban Africa: A problem of the rich or the poor?" *BMC public health*, 9, 465.

Appendix to Chapter 2

2.A Correlations between BMI and WtHR

Table 2.A.1: Correlation matrix between BMI and WtHR

	Pearson correlation coefficients		Spearman correlation coefficients	
	BMI	WtHR	BMI	WtHR
All waves				
BMI	1		1	
WtHR	0.753***	1	0.816***	1
Wave 1				
BMI	1		1	
WtHR	0.764***	1	0.818***	1
Wave 2				
BMI	1		1	
WtHR	0.528***	1	0.645***	1
Wave 3				
BMI	1		1	
WtHR	0.809***	1	0.820***	1
Wave 4				
BMI	1		1	
WtHR	0.878***	1	0.888***	1
Wave 5				
BMI	1		1	
WtHR	0.882***	1	0.896***	1

Notes: *** $p < 0.01$

2.B Variables used in the analysis

Table 2.B.1: Definition of variables

Variable	Definition
DBM	DBM (Double Burden of Malnutrition) household: at least one overweight/obese individual (BMI $\geq$ 25 kg/m^2 or BMI-for-age > 1 SD) and one underweight individual (BMI < 18.5 kg/m^2 or BMI-for-age < -2 SD) (1 = yes; 0 otherwise)
OVOB	OVOB (Overweight/Obese) household: at least one overweight or obese individual (BMI $\geq$ 25 kg/m^2 or BMI-for-age > 1 SD) and a varying number of normal individuals (18.5 kg/m^2 $\leq$ BMI < 25 kg/m^2 or -2 SD $\leq$ BMI-for-age $\leq$ 1 SD) (1 = yes; 0 otherwise)
Socioeconomic variables	
Expenditures	Monthly household expenditures (in logarithm)
Education of household head	Educational level of the household head (1 = no education; 2 = primary education; 3 = secondary education; 4 = higher education)
Employment of household head	Labor market status of the household head (1 = inactive; 2 = unemployed; 3 = employed)
Additional variables	
Urban	Household living area (0 = rural; 1 = urban)
Ethnic group of household head	Ethnic origin of household head (1 = African; 2 = Coloured ; 3 = Asian/Indian*; 4 = White)
Female head of household	Household head is a female (1 = yes; 0 otherwise)
Age of household head	Age of the household head (in years)
Household composition	
Household size	Number of individuals in the same dwelling
Entry	At least one entry of an individual in the household between two waves (birth or arrival) (1 = yes; 0 otherwise)
Exit	At least one exit of an individual in the household between two waves (death or exit) (1 = yes; 0 otherwise)
Number of children	Number of children in the household
Mean age of household	Sum of individuals' ages divided by the household size

Notes: * The association of these two ethnic origins is derived from the NIDS surveys.

2.C Additional robustness checks

Table 2.C.1: Average marginal effects from the dynamic probit model, weight-for-height and BMI-for-age for children

	(1) DBM	(2) OVOB
DBM in $t-1$	0.040***	
	(0.012)	
OVOB in $t-1$		0.096***
		(0.014)
Explanatory variables ($\bar{Z}_{it}$)	Yes	Yes
Initial values (Z_{i0})	Yes	Yes
Averages ($\bar{Z}_l$)	Yes	Yes
Observations	9,526	9,526
N. households	2,693	2,693
Log-likelihood	-2,755.459	-3,959.538
Wald chi2(42)	1,040.60	1,757.19
Prob > chi2	0.000	0.000

Notes: Standard deviations are in parentheses. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table 2.C.2: Predicted probabilities, weight-for-height and BMI-for-age for children

Nutritional status	Dynamics			
	Entry	Persistence	Exit	Mean duration
DBM	10.79	14.67	85.33	1.17
OVOB	68.79	78.56	21.44	4.66

Notes: Probabilities can be interpreted as entry, persistence and exit rates: $Pr(1|0)$; $Pr(1|1)$; $Pr(0|1)$. Mean duration: $1/Pr(0|1)$.

Table 2.C.3: Average marginal effects from the dynamic probit model, raw data

	(1) DBM	(2) OVOB
DBM in $t-1$	0.025**	
	(0.012)	
OVOB in $t-1$		0.074***
		(0.014)
Explanatory variables ($\bar{Z}_{it}$)	Yes	Yes
Initial values (Z_{i0})	Yes	Yes
Averages ($\bar{Z}_{l}$)	Yes	Yes
Observations	8,780	8,780
N. households	2,612	2,612
Log-likelihood	-2,612.702	-3,854.904
Wald chi2(42)	849.48	1,414.00
Prob > chi2	0.000	0.000

Notes: Standard deviations are in parentheses. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table 2.C.4: Predicted probabilities, raw data

Nutritional status	Dynamics			
	Entry	Persistence	Exit	Mean duration
DBM	10.75	13.14	86.86	1.15
OVOB	70.06	77.62	22.38	4.47

Notes: Probabilities can be interpreted as entry, persistence and exit rates: $Pr(1|0)$; $Pr(1|1)$; $Pr(0|1)$. Mean duration: $1/Pr(0|1)$.

Chapter 3

Migration and Nutrition of the Left Behind: Evidence from Ghana

This chapter has been published in *World Development*, https://doi.org/10.1016/j.worlddev.2024.106853

3.1 Introduction

Migration is a phenomenon that affects those who migrate, the communities they move to, and those who stay behind. The individuals who remain in the household of origin after one of their members emigrates are often referred to as left-behind individuals (Antman, 2013; Démurger, 2015). While the common narrative usually emphasizes the experiences of those left behind by international migrants, internal migration is more prevalent than international migration worldwide. This is especially true in regions like West Africa, where most migration occurs within national borders (McAuliffe and Oucho, 2024). As a result, internal migration is likely to affect more left-behind individuals than international migration. Therefore, understanding the impact of internal migration on those left behind, particularly regarding their health and nutrition, becomes a critical aspect to consider.

While the primary motivation for migration often revolves around better income prospects (Kennan and Walker, 2011), it also serves as a strategy to diversify risks and elevate household welfare (Stark and Bloom, 1985). Migration is also undertaken to improve the overall living conditions of the household through income and to finance consumption, mainly through remittances (Stark and Lucas, 1988). Therefore, it should enhance the health and nutrition of left-behind individuals, especially with the additional income and remittances that could positively impact the quality and quantity of food consumed (De Brauw and Mu, 2011). However, migration's disruptive nature can also usher in adverse effects. For instance, the absence of a migrant who was a primary caregiver can lead to children experiencing poorer dietary habits (Démurger, 2015). This open-ended issue is also observed in the literature. Indeed, there has been no definitive evidence on the direction of the impact of migration on the nutrition of left-behind individuals (Thow et al., 2016; Fellmeth et al., 2018). Some studies argue that migration has positive effects on nutritional indicators for both children and adults left behind (see, for instance, Hildebrandt et al. (2005), Zezza et al. (2011), Carletto et al. (2011), De Brauw (2011), Böhme et al. (2015), Liu et al. (2021a)), while others find mixed effects (Antón (2010), Gibson et al. (2011), De Brauw and Mu (2015), Sznajder et al. (2021), Wei (2022), Vikram (2023)), or even negative effects (Fellmeth et al., 2018). The uncertainty regarding the direction may stem from various factors, and studies may not be incorrect per se in their findings. Indeed, it could depend on context dependence (Démurger, 2015), the type of migration studied (international or internal), the characteristics of the migrants, the time frame under examination, or the methods employed.

This paper aims to identify the impact of internal migration on the nutritional status of adults and children left behind. However, studying such an effect is a complex issue fraught with multiple threats and challenges. Indeed, migration decisions are typically non-random, introducing concerns of endogeneity and selection bias (Antman, 2013; Gibson et al., 2013). This selection can manifest both inter-household, relating to the household's collective choice to send a migrant, and intra-household, determining which household member migrates (Murard, 2019). To address these challenges, I adopt an approach using panel data from two waves of the Ghana Socioeconomic Panel Survey (GSPS), spanning 2013/2014 to 2017/2018. A combination of kernel matching and difference-in-differences (DID) is employed to establish two comparable groups: individuals who are left behind and those who are not. This approach addresses selection bias on observable factors, while the Oster test is applied to ensure no bias from unobservable factors or omitted variables (Oster, 2019).

Using the context of Ghana, where internal migration is predominant, I find that migration adversely affects the nutritional status of those left behind, particularly children. Indeed, children experience a decrease in their BMI-for-age z-score and are more susceptible to having a detrimental nutritional status. The primary channel underlying these findings is the disruptive effect of migration. This effect may be seen in the perturbation of normal household functioning as the migrant moved (Stephen and Bean, 1992; Davis and Brazil, 2016). Such disruption can appear in various ways: economic, with the loss of a previous income contributor; social, as roles and household structures can shift; or even psychological, as family dynamics are affected. Essentially, I am capturing the short-term effect of migration that can adversely affect the nutritional status of those left behind. The migration process entails significant costs, not only in financing the migrant's journey but also, primarily, in terms of the loss of their contribution to household income. The negative impact likely stems mainly from the loss of the migrant's previous economic contribution to their household, which may result in a negative income shock following the migrant's departure. This adverse income shock, while having a moderate effect on adults' weight, has a profoundly detrimental impact on children's nutritional status. Furthermore, even though I find a positive long-term impact of remittances on children, this short-term negative effect could cast a long-lasting shadow, affecting the enduring nutritional health and growth of children who are less resilient compared to adults.

I make four distinct contributions to the literature. First, I seek to reconcile the mixed results found in the literature by placing a strong emphasis on understanding the mechanisms and the temporality of the effects. I contribute to the existing literature that investigates the transmission channels through which migration can impact the nutrition of left-behind individuals (see, for instance, De Brauw and Mu (2011); De Brauw (2011); Carletto et al. (2011); Zezza et al. (2011); De Brauw and Mu (2015); Davis and Brazil (2016); Viet Nguyen (2016)). In particular, I delve into the influence of remittances, aligning with prior work like Davis and Brazil (2016) and Vikram (2023), among others. Simultaneously, I explore other potential transmission channels, focusing on migration's disruptive effect. Hence, a key contribution is to understand the rationales behind the diverse and inconsistent effects documented in the literature. My findings suggest a short-term disruptive effect primarily through the negative income shock channel. However, in the long term, via the remittances, I also uncover the potential for positive effects on children. This underscores the importance of considering the temporal dimension when studying the impact of migration.

Second, to fill the gaps in the literature and further explain why findings might be mixed, this paper investigates the heterogeneity of the results in detail. To do so, I not only focus on left-behind adults (Gibson et al., 2011; Böhme et al., 2015; Liu et al., 2021a; Sznajder et al., 2021; Wei, 2022) or on left-behind children (Antón, 2010; Gao et al., 2010; Carletto et al., 2011; De Brauw, 2011; De Brauw and Mu, 2011, 2015; Davis and Brazil, 2016; Viet Nguyen, 2016; Vikram, 2023), but on all household members. Additionally, I study the impact of migration by differentiating the effects according to gender, age, and nutritional status, while also investigating the outcomes of parental migration. Furthermore, the profile of the migrants is examined to understand the underlying mechanisms, an aspect often overlooked in the literature. Indeed, what is lacking in current research is an understanding of who migrates, beyond just focusing on parental migration. Therefore, I explore as much heterogeneity as possible to assess what might be lacking in current research and to explain the mechanisms and diverse effects at play.

Third, from a methodological perspective, while many prior studies rely on cross-sectional data (Fellmeth et al., 2018), only a limited number of studies use longitudinal data (see, for instance, Tian et al. (2017), Liu et al. (2021a), and Yi et al. (2019)). Therefore, I capitalize on the advantages of tracking individuals over time to better address household-level selection bias using a combination of kernel matching and

DID, while ensuring no bias from unobservable factors using an Oster test. I also address the question of intra-household level selection bias, a step not commonly undertaken in the literature.

Finally, to the best of my knowledge, this issue has been studied to a limited extent in sub-Saharan Africa and even less so in Ghana.[1] Indeed, many studies focus on Asia and particularly China (Gao et al., 2010; De Brauw and Mu, 2011, 2015; Viet Nguyen, 2016; Lei et al., 2018; Liu et al., 2021a; Sznajder et al., 2021; Vikram, 2023) or Latin America (Antón, 2010; Carletto et al., 2011; De Brauw, 2011; Ponce et al., 2011; Davis and Brazil, 2016).

The remainder of the paper is structured as follows. Section 3.2 reviews the literature. Section 3.3 introduces the data. Section 3.4 presents the empirical strategy and Section 3.5 the results. Section 3.6 explores robustness and heterogeneity checks. A discussion of the transmission channels is presented in Section 3.7. Finally, Section 3.8 draws the main conclusions.

3.2 Literature review

Existing literature has shown that migration can significantly impact the nutritional status of individuals who remain in the household of origin after a member migrates, leading to mixed findings. These mixed results highlight the complexity of migration's effects on the left behind. Therefore, to fully understand these results, it is crucial to examine the underlying mechanisms and transmission channels, as well as how they contribute to explaining the impact of migration on the nutritional status of the non-migrating members.

The most direct channel through which migration can impact the nutrition of individuals left behind is the income channel. When a household decides to send a migrant, it expects the overall household income to increase, mainly through remittances. These transfers can improve the consumption and investment of the remaining household members (Yang, 2011), and by increasing the household's available income, remittances can alleviate budgetary constraints, thereby increasing food expenditures. Thus, the additional income should enhance the quantity and/or quality of food consumed, enabling the household to invest more in the nutrition of those left behind (De Brauw and Mu, 2011; Islam et al., 2019). Empirical studies support this; for example, De Brauw (2011) suggests that the positive correlation between migration and the height-for-age z-scores (HAZ) of children left behind is likely driven by international remittances. Similarly, Vikram (2023) highlights the role of internal transfers in improving HAZ among boys in Indian households, while Quartey (2006) emphasizes the critical importance of international transfers for maintaining consumption and meeting the immediate needs of the poorest households in Ghana. Additionally, Mora-Rivera and van Gameren (2021) reveal that in Mexico, remittances, particularly international ones, significantly reduce food insecurity in rural households. By reviewing the literature, Thow et al. (2016) also report that remittances (both from domestic and international sources) have positive effects on the nutrition of left-behind individuals. Based on this channel, migration should positively impact nutrition, yet evidence also shows mixed or even negative impacts of remittances. For example, while some studies find no impact of international remittances (Ponce et al., 2011), others suggest that remittances can have indirect adverse effects. Higher income, for instance, may lead to a shift in dietary patterns towards higher energy and fat intake, along with increased consumption of meat and processed foods, potentially contributing to obesity (Guo et al., 2000).

[1] To the extent of my knowledge, only Karamba et al. (2011) have studied the impact of migration on household-level food consumption patterns in Ghana, and they do not address the individual (anthropometric) dimension.

Another dimension of the income channel is the disruptive effect that migration can generate in the household of origin of the migrant (Davis and Brazil, 2016). Indeed, after migration, the household may be disorganized for a period. Until the migrant finds a job, the household may lose a working-age individual who supported the family. This loss of income, even temporarily, can harm the nutrition of those left behind, particularly children, during a critical phase of their development. The loss of skilled and working-age labor can be detrimental, particularly if remittance income fails to sufficiently compensate for this loss. Further research is needed to explain why and how the income channel can have dual effects, both through the positive impact of remittances and the potential disruption caused by lost income. This paper aims to address this issue in part.

Nevertheless, the income channel represents one among multiple mechanisms that interact to shape the nutritional outcomes of left-behind individuals. Indeed, several other channels have been identified (for a detailed summary, see Zezza et al. (2011)). Among these, evidence suggests that when the migrant is a parent, their absence may negatively impact child nutrition (Gao et al., 2010; Lei et al., 2018). De Brauw and Mu (2015) also highlight that reduced parental supervision due to the absence of the migrant parent can have detrimental effects on children's nutrition. Using multi-country data, Viet Nguyen (2016) concludes that parental migration results in less frequent contact and care for children, which is detrimental to their nutrition. In addition, Bai et al. (2022) show that maternal migration can have negative effects on a child's diet. This could be gathered into a time allocation effect, which might also imply changes in how household members allocate their time.

Consequently, individuals remaining in the household of origin may have to take on tasks previously handled by the migrant, such as those related to agricultural production (Mu and van de Walle, 2009) or household chores. Specifically, less time may be dedicated to monitoring children's eating habits, and adults might spend less time on food preparation and shopping. Additionally, children may take on more chores, such as cooking and childcare, reducing the care they receive and negatively impacting their nutritional status, especially older children (De Brauw and Mu, 2011).

Other channels may also have unintended effects. For instance, migration can alter intra-household bargaining and cooperation dynamics (De Brauw and Mu, 2015), such as when the departure of a male household head transfers the role to a female member. In line with this shift, evidence suggests that female household heads often prioritize children's nutrition (Kennedy and Peters, 1992), potentially leading to improved nutritional outcomes. Finally, some channels may be intertwined. In this regard, Carletto et al. (2011) argue that international remittances can improve child growth but are also related to a combination of different effects, and in the context of Mexico, Hildebrandt et al. (2005) observe that children in households with migrants who went to the United States are less likely to be underweight than those in non-migrant households, explaining this through non-monetary channels like raising health knowledge of mothers in addition to the direct effect on health of higher wealth after migration.

To summarize, migration can influence the nutrition of the left behind through multiple channels. Some channels may yield positive outcomes, primarily through remittances, while others can have adverse consequences, such as the disruptive income effect or parental absence, especially for children. Additionally, some channels may produce undetermined or mixed effects. Therefore, what remains to be understood in this literature is the reason behind the uncertain direction of the effects. This uncertainty likely stems from multiple factors, including the specific context, migrant profiles, or the interplay between different mechanisms. However, this paper aims to demonstrate that the primary reason is related to the temporality of these impacts.

3.3 Data

3.3.1 Migration in Ghana

Over the last two decades, Ghana has experienced a significant rise in internal migration, with internal moves accounting for more than 90% of all migration (Ghana Statistical Service, 2013). Predominantly, these internal migrations are long-distance, with individuals relocating between regions rather than within them (International Organization for Migration, 2020). In addition, about 35% of the 2010 population census had moved from their place of birth to another location in the country (Ghana Statistical Service, 2013). Therefore, Ghana offers a pertinent context for analyzing the impact of migration, with a specific focus on internal migration.

The patterns captured in the data corroborate these observations. As illustrated in Figure 3.A.1, the vast majority of labor migration occurs within Ghana, with 59% of migrants moving to a region different from their origin. Further, as depicted in the final two figures of Appendix 3.A, most of the work-related migrations for more than six months involve regional relocations. Although migrants originate from diverse regions across Ghana (Figure 3.A.2), a substantial proportion, exceeding 56%, choose the Accra (around 36%) or Ashanti (about 20%) regions as their destinations (Figure 3.A.3).

3.3.2 Ghana Socioeconomic Panel Survey Data (GSPS)

3.3.2.1 Dataset

This study draws data from the last two waves of surveys from the EGC-ISSER Socioeconomic Panel Survey, also known as the Ghana Socioeconomic Panel Survey Data (GSPS), and implemented by Yale University's Economic Growth Center (EGC) and the Institute of Statistical, Social and Economic Research (ISSER).[2] The last two waves were conducted between 2013/2014 and 2017/2018.[3] The survey's main objective is to provide a framework to study the medium- and long-term economic development processes. The survey is based on a nationally representative sample for the ten regions of Ghana, initially covering 5,010 households from 334 Enumeration Areas (EAs) selected from a master sampling frame. The sample was ensured to be representative by a two-stage stratified sample design.

3.3.2.2 Sample

The sample is restricted to individuals present in Wave 2 (2013/2014) and Wave 3 (2017/ 2018), those who did not exit a household, and those in households that did not dissolve. As I compare those left behind to those not, I do not consider new household members in Wave 3.[4] The sample includes individuals

[2] Publicly available data can be accessed at: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi%3A10.7910/DVN/E5QP0F. Last accessed: 29 October 2024.

[3] Not all three waves from the GSPS were used due to methodological considerations. The primary reason is the eight-year gap between waves 1 and 3, which could introduce inconsistencies. Since this study examines the effect of migration on both children and adults, using all three waves would mean that many children interviewed in Wave 1 would have transitioned into adulthood by Wave 3, thereby complicating the comparability of nutritional outcomes. Indeed, as discussed later, anthropometric measures differ for children and adults, and relying on all three waves would restrict the analysis to children under 10 years old in Wave 1, as older children would have different anthropometric indicators by Wave 3. Additionally, I chose to use the two most recent waves to provide the most up-to-date analysis. Nevertheless, data from Wave 1 is still incorporated, both to examine prior trends and in robustness checks. Thus, while Wave 1 is not part of the main analysis, it plays an essential role in supporting the findings.

[4] Regarding individuals who moved into households between the two survey waves, particularly within migrant households, they may differ from those left behind in terms of nutritional status. To investigate this, I estimate a regression where

who remained in the same household and were successfully interviewed between the two survey waves (n = 11,945). Given the use of two different outcomes for adults and children (weight and BMI-for-age z-scores), the analysis focuses on individuals who were already adults in Wave 2 and children who did not become adults between the two waves (556 individuals concerned, n = 11,389). Children under two are omitted as anthropometric measurements are less trustworthy among this age group (518 individuals concerned, n = 10,871) (WHO and UNICEF, 2019). Pregnant women are also excluded considering the weight variation during pregnancy (267 women concerned, n = 10,604). Observations with implausible information on anthropometry during at least one wave have been deleted (888 individuals concerned, n = 9,716). Similarly, observations with missing anthropometric data in at least one wave were removed (1,422 individuals concerned, n = 8,294).[5] Other observations were missing for some variables and deleted (935 individuals concerned, n = 7,359).[6] I opted for a complete case analysis rather than an available case analysis to compare different samples easily. Moreover, given the different heterogeneities I seek to observe, it is more appropriate to use comparable samples. Finally, the resulting balanced sample includes 7,359 individuals (4,579 adults and 2,780 children) spread over two waves, corresponding to 14,718 observations (9,158 observations for adults and 5,560 for children). Additionally, the use of longitudinal data necessitates addressing concerns related to attrition. An examination of potential attrition bias is presented in Appendix 3.D. In summary, the findings are not threatened by attrition bias.[7]

3.3.3 Variables

3.3.3.1 Anthropometric measures

The nutritional status of individuals is determined using anthropometric measurements. For children aged 2 to 18 years, BMI-for-age z-scores are utilized. This indicator is used as it reflects medium- to long-term effects of nutritional changes and provides insights into dietary or lifestyle alterations. For robustness checks, I also use the height-for-age z-scores, which are commonly used to capture the long-term effects of undernutrition and can indicate chronic malnutrition. As Ghana faces the double burden of malnutrition—despite significant reductions in hunger over the past decades, undernutrition remains prevalent and now coexists with overweight and obesity (Casu et al., 2021)—I categorized children into different nutritional statuses to assess the heterogeneity of the results. The children's nutritional statuses

the outcome variable is the nutritional status of individuals, with the key independent variable being a binary indicator: 1 if the individual moved into a migrant household and 0 if the individual was left behind. The regression also includes controls at both the individual and household levels, as well as household fixed effects. The goal is to assess whether those who moved into households differ from those left behind in terms of their nutritional status, following the approach outlined by Murard (2019). More details on this methodology are provided in Section 3.4.4, which addresses potential within-household selection bias. In summary, there is no significant difference in nutritional indicators between left-behind individuals and those who moved into the household between waves 2 and 3. These results are available upon request.

[5]In most cases, the missing data were due to the individual being away from home during the interview or not wanting their data to be collected.

[6]In Appendix 3.D.1, I assess whether excluding individuals with missing values for either outcomes or control variables introduces any bias in the analysis and conclude that it does not.

[7]This analysis demonstrates that attrition is not a significant issue. However, within migrant households, left-behind individuals and those who moved out of the household may differ. Although most of those who moved out are attritors, this alone might not be sufficient to conclude that they are not different, particularly in terms of nutrition. To address this, I apply the same strategy mentioned in footnote 3 for those who moved out of migrant households. Specifically, I estimate a regression where the outcome variable is the nutritional indicator, with the main independent variable being a binary indicator: 1 if the individual moved out of a migrant household and 0 if the individual was left behind. In the results available upon request, I show no significant difference in nutritional indicators between left-behind individuals and those who moved out of migrant households between the two survey waves.

are as follows: underweight when BMI-for-age z-score $<$ -2 SD, healthy if BMI-for-age z-score $\geq$ -2 SD and $\leq$ 1 SD, and overweight/obese when BMI-for-age z-score $>$ 1 SD (WHO, 2006).
For adults (from 19 years old), I use the body weight (in kilograms). Weight is preferred over BMI, as weight changes are easier to interpret. However, BMI is utilized to classify individuals according to their nutritional status, i.e., to also assess the weight changes for individuals who are considered underweight, healthy, or overweight/obese. Indeed, a variation in an individual's weight can have different implications depending on their nutritional status. The thresholds are: underweight for BMI $<$18.5 kg/m^2; healthy if BMI $\geq$ 18.5 kg/m^2 and $<$ 25 kg/m^2; and overweight/obese for BMI $\geq$ 25 kg/m^2 (WHO, 1995).

3.3.3.2 Migration and left-behind individuals

The literature lacks a universal definition for left-behind individuals. Indeed, migration occurs in varied contexts, driven by diverse motivations, with differing durations, destinations, etc. Consequently, as the parameters of migration change, so does the conceptualization of who is considered left behind. However, it commonly revolves around those who continue to reside in their original household after one of its members has migrated, i.e., based on past co-residence with a migrant. In the literature studying the impact of migration on the left behind, migration is often defined as a movement for labor or employment purposes (see Gao et al. (2010), Nguyen and Winters (2011), Lei et al. (2018), Fellmeth et al. (2018), Lu et al. (2020), among others). This study focuses on labor migration for several reasons. First, labor migrants have a strong propensity to remit, which helps sustain connections with their households of origin. This contrasts with other forms of migration, where such ongoing relationships may not endure. Indeed, the salient point is that labor migrants are likely to maintain ties with their households. Moreover, labor migration offers a more relevant context than other types, such as family-driven migration, by facilitating a clearer analysis of underlying mechanisms. For instance, separations due to divorce or other family issues can complicate the analysis, as specific motivations may not always be explicitly identified, and ties with the household of origin may be less clear.
Using the longitudinal nature of the data, I can identify individuals who left their households between the two survey waves and the primary reason for their departure. In the questionnaire, I use the following questions: "What is the main reason [Name] is no longer considered a member of this household?" and "For how long has [Name] not been a member of this household?".[8] Using these questions, two groups are defined:

- Left-behind individuals (the treatment group): those residing in households where, between the two survey waves, one or more members[9] have migrated outside the household looking for work and have been away for more than six months.[10]

- Non-left-behind individuals (the control group): those who remained in the same household between the two survey waves without any household member leaving for work for more than six months.

[8]These questions appear in Section 1: Household Background, Part B2: Household Roster.

[9]The average number of migrants per household is approximately one individual (1.2 migrants). In Appendix 3.J.11, I also analyze the heterogeneities based on whether the household sends one migrant or more than one migrant.

[10]Regarding the age of the migrants, more than 89.75% are over 15 at the time of their migration, but migrants under 15 are also included. Results excluding left-behind individuals from migrants under 15 are available in Appendix 3.J.12. The findings remain consistent.

The focus on migrants absent for more than six months is due to two considerations: first, it is improbable that migration of a lesser duration would substantially influence the nutritional status of those left behind, and second, with this time frame, I can partially overlook seasonal migration, which typically lasts less than six months.[11] Additionally, as indicated by Figure 3.A.1, the findings will reflect the impact of internal migration.[12]

3.3.4 Descriptive statistics

Tables 3.1 and 3.2 show descriptive statistics at baseline (in Wave 2), i.e., before a migration occurs in Wave 3.[13] The statistics are stratified by migration status, which means splitting the sample between left-behind and non-left-behind individuals.[14] Left-behind adults represent 10.30% of all adults and left-behind children represent 11.30% of all children. Left-behind adults are slightly older and are mostly women. There are no significant differences between left-behind and non-left-behind children for BMI-for-age z-scores, age, and gender. The household heads of left-behind individuals are older and less educated than non-left-behinds. Migrant households have more individuals and working-age members but also fewer dependent individuals and are more likely to live in rural areas. According to the wealth index, they also tend to be less wealthy. Consequently, there are already differences at baseline between left-behind and non-left-behind individuals, both for adults and children. It must, therefore, be considered in the empirical strategy.

[11] Some individuals are in households with migrants who moved for less than six months. In the main specification, these individuals are included in the control group. In a robustness check in Appendix 3.J.5, I examine whether this inclusion biases the results. 73.75% of migrants have been away for more than twelve months, while 26.25% migrated between six and twelve months ago. In Appendix 3.J.10, I also examine the differences when defining the treatment group with different migration temporalities.

[12] The robustness is ensured by checking the results for internal migration within Ghana only (Section 3.J.4).

[13] In Appendix 3.C, two tables display the descriptive statistics for Wave 3 (Tables 3.C.2 and 3.C.3).

[14] Additionally, a table of descriptive statistics for the total sample at baseline, without stratification by migration status, is available in Appendix 3.C (Table 3.C.1). This appendix also includes a table presenting the distribution of individuals by nutritional status, by wave, and according to their migration status (Table 3.C.4).

Table 3.1: Descriptive statistics of adults by migration status in wave 2

	Adults				
	Non Left Behind		Left Behind		t-test
	Mean	(SD)	Mean	(SD)	P-value
Individual variables					
Weight	62.951	(13.236)	61.136	(14.008)	0.005
Age	45.544	(15.675)	48.417	(15.418)	0.000
Male	0.463	(0.499)	0.394	(0.489)	0.004
Relationship to the head					
Head	0.623	(0.485)	0.464	(0.499)	0.000
Spouse	0.268	(0.443)	0.360	(0.480)	0.000
Child	0.082	(0.275)	0.132	(0.339)	0.000
Grandchild	0.007	(0.081)	0.011	(0.103)	0.316
Other relationship	0.020	(0.141)	0.034	(0.182)	0.051
Household head variables					
Age of the head	50.825	(15.256)	56.834	(12.882)	0.000
Male head	0.732	(0.443)	0.772	(0.420)	0.058
Education of the head					
Head, none or preschool	0.350	(0.477)	0.504	(0.501)	0.000
Head, primary education	0.154	(0.361)	0.102	(0.303)	0.003
Head, post-primary education	0.350	(0.477)	0.304	(0.461)	0.047
Head, secondary education	0.084	(0.278)	0.028	(0.164)	0.000
Head, tertiary education	0.062	(0.241)	0.062	(0.241)	0.992
Household variables					
Household size	4.214	(2.588)	6.504	(2.920)	0.000
Working-age members	2.330	(1.431)	3.877	(1.758)	0.000
Dependency ratio	1.121	(1.415)	0.824	(0.875)	0.000
Rural	0.624	(0.484)	0.747	(0.435)	0.000
Wealth index					
Wealth index, 1st quintile	0.260	(0.439)	0.364	(0.482)	0.000
Wealth index, 2nd quintile	0.168	(0.374)	0.177	(0.382)	0.635
Wealth index, 3rd quintile	0.193	(0.395)	0.174	(0.380)	0.333
Wealth index, 4th quintile	0.208	(0.406)	0.147	(0.354)	0.002
Wealth index, 5th quintile	0.172	(0.377)	0.138	(0.346)	0.068
Region					
Western Region	0.076	(0.266)	0.049	(0.216)	0.031
Central Region	0.078	(0.268)	0.051	(0.220)	0.038
Greater Accra Region	0.108	(0.310)	0.017	(0.129)	0.000
Volta Region	0.083	(0.275)	0.106	(0.309)	0.079
Eastern Region	0.102	(0.303)	0.085	(0.279)	0.236
Ashanti Region	0.166	(0.372)	0.123	(0.329)	0.017
Brong-Ahafo Region	0.102	(0.302)	0.115	(0.319)	0.374
Northern Region	0.183	(0.387)	0.300	(0.459)	0.000
Upper East Region	0.071	(0.257)	0.134	(0.341)	0.000
Upper West Region	0.031	(0.174)	0.019	(0.137)	0.141
Observations	4,109		470		

Table 3.2: Descriptive statistics of children by migration status in wave 2

	Children				
	Non Left Behind		Left Behind		*t*-test
	Mean	(SD)	Mean	(SD)	*P*-value
Individual variables					
Zbmi	-0.102	(1.793)	-0.048	(1.801)	0.611
Age	8.626	(3.732)	8.882	(3.654)	0.251
Male	0.560	(0.496)	0.561	(0.497)	0.987
Relationship to the head					
Head	0.000	(0.000)	0.000	(0.000)	0.000
Spouse	0.002	(0.045)	0.003	(0.056)	0.677
Child	0.880	(0.325)	0.850	(0.357)	0.127
Grandchild	0.085	(0.279)	0.118	(0.323)	0.052
Other relationship	0.033	(0.178)	0.029	(0.167)	0.693
Household head variables					
Age of the head	46.357	(12.272)	52.459	(11.100)	0.000
Male head	0.751	(0.432)	0.822	(0.383)	0.006
Education of the head					
Head, none or preschool	0.404	(0.491)	0.580	(0.494)	0.000
Head, primary education	0.153	(0.360)	0.124	(0.330)	0.174
Head, post-primary education	0.319	(0.466)	0.239	(0.427)	0.004
Head, secondary education	0.072	(0.259)	0.013	(0.112)	0.000
Head, tertiary education	0.051	(0.220)	0.045	(0.207)	0.620
Household variables					
Household size	6.013	(2.475)	7.997	(2.838)	0.000
Working-age members	2.668	(1.315)	3.987	(1.564)	0.000
Dependency ratio	1.502	(1.022)	1.172	(0.935)	0.000
Rural	0.690	(0.463)	0.885	(0.319)	0.000
Wealth index					
Wealth index, 1st quintile	0.316	(0.465)	0.490	(0.501)	0.000
Wealth index, 2nd quintile	0.156	(0.363)	0.178	(0.383)	0.310
Wealth index, 3rd quintile	0.180	(0.385)	0.102	(0.303)	0.000
Wealth index, 4th quintile	0.194	(0.395)	0.137	(0.344)	0.015
Wealth index, 5th quintile	0.153	(0.360)	0.092	(0.290)	0.004
Region					
Western Region	0.080	(0.271)	0.032	(0.176)	0.002
Central Region	0.070	(0.255)	0.051	(0.220)	0.203
Greater Accra Region	0.067	(0.251)	0.013	(0.112)	0.000
Volta Region	0.065	(0.246)	0.089	(0.285)	0.107
Eastern Region	0.090	(0.286)	0.041	(0.200)	0.004
Ashanti Region	0.162	(0.369)	0.089	(0.285)	0.001
Brong-Ahafo Region	0.108	(0.311)	0.131	(0.337)	0.236
Northern Region	0.259	(0.438)	0.395	(0.490)	0.000
Upper East Region	0.069	(0.253)	0.146	(0.354)	0.000
Upper West Region	0.030	(0.170)	0.013	(0.112)	0.086
Observations	2,466		314		

Notes: Zbmi refers to the BMI-for-age z-score.

3.4 Empirical strategy

3.4.1 Threats to identification

The objective of this paper is to assess the effect of migration on the nutrition of left-behind individuals. However, identifying a causal effect is challenging due to various threats to identification, particularly selection bias and reverse causality (Démurger, 2015). A key aspect of migration is that migrants are not randomly drawn from the general population. Indeed, the decision to migrate is typically a deliberate choice, making self-selection a major issue (Gibson et al., 2013). This self-selection occurs at two levels. At the household level, households with migrants may have different observable and unobservable characteristics that influence their likelihood of migration. Within the household, self-selection may also occur regarding who will migrate, as those chosen to migrate may have distinct characteristics compared to those who remain, potentially leading to intra-household selection bias (Chort and Senne, 2015, 2018; Murard, 2019).

Additionally, migration might be correlated with the same factors that influence the nutrition of the left behind. These factors may interfere in estimating whether migration affects nutrition or whether it is an omitted variable correlated with migration and nutrition that explains the results. In studying the effects of migration, one may capture a wrong effect because of reverse causality (Antman, 2013). A migrant may choose to migrate in response to the poor health of individuals, while conversely, having individuals in poor health may also reduce the likelihood that someone will leave. In any case, the individuals' nutritional status may drive the decision to send a migrant rather than the opposite.

In summary, individuals who are left behind may differ from those who are not in observable or unobservable characteristics. This pattern is also reflected in the data presented in Tables 3.1 and 3.2: at baseline, there are differences between left-behind and non-left-behind individuals regarding observable variables, which can also suggest potential differences in unobservables. Consequently, a rigorous identification strategy that addresses the potential biases arising from self-selection and endogeneity is necessary to estimate the impact of migration on the nutritional status of the left behind.

3.4.2 Addressing self-selection into migration

Given the longitudinal nature of the data, individuals can be tracked and two groups can be compared over two survey periods: the left behind and the non-left behind. To assess the impact of migration, one can compare their trends in anthropometric indicators using a difference-in-differences (DID) approach. By analyzing the before-and-after periods, it is possible to evaluate pre-period differences in nutrition between the treatment and control groups, which helps control for pre-existing differences and mitigate a part of the selection bias. Such a method could partially control for unobserved characteristics common to both groups that might be correlated with both migration and nutrition.

Using DID relies on the key assumption of parallel trends. However, verifying this assumption in this study is challenging, as the GSPS includes only three waves of data. Robust verification typically requires more than two pre-treatment periods. Moreover, testing parallel trends using data from waves 1 and 2 could exclude many individuals from the main sample analyzed between waves 2 and 3. This exclusion includes new household members not present in Wave 1, such as those born between waves 1 and 2 or not interviewed in Wave 1. As a result, the parallel trends assumption can only be rigorously tested for individuals interviewed across all three waves. Despite these limitations, the parallel trends assumption is tested in Appendix 3.E. In summary, for adults, data from waves 1 and 2 show an increase in average body weight in migrant households, whereas a decrease is observed in non-migrant households. This divergent trend mitigates concerns, as the opposite pattern is observed between waves 2 and 3 post-treatment. Indeed, if the trend had remained the same, it would have posed a greater problem. For children, a similar decline in z-scores in both groups before treatment is reassuring, indicating almost similar initial conditions. Moreover, in Wave 2, z-scores are not statistically different between left-behind and non-left-behind children. After Wave 2, trends diverge following the treatment, further supporting this approach. Overall, although the parallel trend assumption is not fully met, these patterns suggest that the DID method remains relatively robust for both adults and children.

Nevertheless, even though DID can yield insightful results, I employ a more robust strategy that combines propensity score matching (PSM) with DID. By integrating these two methods, I create more comparable treatment and control groups, achieving a better balance of observable covariates and reducing initial differences between groups. PSM effectively controls for selection bias on observable factors, assuming that selection into migration depends on observable characteristics, while DID mitigates bias from unobservable factors, provided their influence remains constant over time. This approach offers a more comprehensive response to selection bias.

One potential threat remains to be addressed: even when combining matching with DID, issues related to time-varying unobservables or potential bias due to omitted variables may still arise. To ensure that my strategy is not biased by selection on unobservables, I employ the methodology proposed by Oster (2019). This approach leverages selection on observables to estimate the potential severity of selection on unobservables. By applying Oster bounds, I estimate the likely degree of omitted variable bias and determine whether unobserved factors could significantly influence the findings. Specifically, the principle is to compare coefficients and R-squared values from the baseline model (without controls) to those from the fully controlled model. The bias-corrected coefficients β^* are derived under the assumption that unobserved and observed covariates are equally important ($\delta = 1$), and using a maximum R-squared equal to 1.3 times the R-squared from the saturated specification, as suggested by Oster (2019). If the estimated coefficient bounds interval does not include zero, the estimates are robust to unobservables. The findings from the Oster test are detailed in Section 3.6.1. In short, the results indicate that the outcomes are unlikely to be influenced by selection on unobservables, reinforcing the credibility and validity of the matching DID method.

3.4.3 Kernel matching difference-in-differences

The model retained is a kernel-based propensity score matching difference-in-differences. Kernel matching is more suitable than other matching methods, as it retains more observations within the common support, achieves greater bias reduction, and uses more observations for matching, thereby reducing variance (Liu et al., 2021b). Combining DID with matching is widely considered more reliable than DID alone (Rosenbaum and Rubin, 1983; Khandker et al., 2009). Indeed, employing matching followed by DID on the matched sample with the inclusion of weights generally provides a more credible method for estimating causal effects compared to regression on an unmatched sample. This combination is also recognized as robust when randomization is not feasible and in the context of non-experimental study designs, as it provides more credible estimates (Stuart et al., 2014).

Matching is based on a set of covariates, selected according to the analytical framework of established literature exploring the effects of migration on various outcomes (e.g., Tian et al. (2017); Bai et al. (2018); Yi et al. (2019); Marchetta and Sim (2021)). In this literature, this combined method has already been used. For instance, Tian et al. (2017) applied it to examine the impact of parental migration on children's growth, Lu et al. (2020) used it to study the effect of migration on self-reported health, and Bai et al. (2018) investigated the impact of parental migration on children's academic performance. This process also takes into consideration the inclusion of variables that influence both the likelihood of an individual being left behind and their nutritional status (Caliendo and Kopeinig, 2008). The covariates include individual-level variables such as age and gender (whether the individual is male); household head variables such as the head's age, gender, and education level (represented by dummy variables); and household-level variables including household size, number of working-age members, dependency ratio, wealth quintiles based on a Principal Component Analysis (PCA),[15] rural residency, and region dummies.[16] All variables are measured before the treatment and are described in Appendix 3.B. Among the variables, some might be considered endogenous, especially those related to household headship (as discussed by Bertoli and Marchetta (2014)), household size, or wealth index. This issue is addressed in Appendix 3.J, concluding that the endogeneity of these variables is not a concern.[17]

Conceptually, the strategy starts with a probit model to estimate the likelihood of being treated, yielding propensity scores. These scores are then used to calculate kernel weights. The matching process balances the treatment and control groups, making them more similar in baseline characteristics. Finally, these kernel weights are incorporated into the DID. Overall, this approach ensures that the treatment and control groups are comparable, thereby reducing selection bias and enhancing the credibility of the causal inferences.

The treatment effect of the kernel propensity-score matching DID takes the following form:

$$\begin{aligned} DID = \{E(Y_{i,t=1}|D_{i,t=1} = 1, Z_i = 1) - w_i \times E(Y_{i,t=1}|D_{i,t=1} = 0, Z_i = 0)\} \\ - \{E(Y_{i,t=0}|D_{i,t=0} = 0, Z_i = 1) - w_i \times E(Y_{i,t=0}|D_{i,t=1} = 0, Z_i = 0)\} \end{aligned} \tag{3.1}$$

in which $Y_{i,t}$ is the outcome variable (weight or BMI-for-age z-score), $Z_i = 1$ is the treatment group, and $Z_i = 0$ is the control group. The treatment indicator requires the absence of any intervention in the baseline for either group ($D_{i,t=0} = 0|Z_i = 1, 0$) and it requires the intervention to be positive for the treated group in the follow-up ($D_{i,t=1} = 1|Z_i = 1$). Finally, w_i are the kernel weights that take the form: $w_i = \frac{K\left(\frac{p_i - p_k}{h_n}\right)}{\sum K\left(\frac{p_i - p_k}{h_n}\right)}$, with $K(.)$ being the kernel function, p the propensity scores, and h_n the selected bandwidth parameter, which is set to the default value (0.06).

The analysis is restricted to the common support, ensuring that only individuals with suitable control cases within the common support region are considered to have reasonable matching. The matching

[15] To build the wealth index, I used housing characteristics and durable goods owned by the household. The index was standardized to fit into a 0 to 1 index. Households were then categorized into quintiles.

[16] The treatment is at the household level. However, individual-level variables are included in the matching, leading to different propensity scores among household members. To address this, I also estimate the propensity score at the household level, using only household or household head variables. This analysis is available in Appendix 3.J.7, and the results are similar. It also demonstrates that changing the variables used in matching does not significantly alter the findings, thereby reinforcing the robustness of the strategy.

[17] In addition, I considered including variables related to labor market participation. I constructed variables capturing individual and household-level labor market participation. However, I decided not to incorporate them. This decision was due to the imperfect nature of the measures and missing values for a portion of the sample, which compromised their use. The rationale for their exclusion is elaborated in Appendix 3.J.9. Nonetheless, I use these variables in an alternative model, which is also provided in Appendix 3.J.9.

quality is ensured by ascertaining that both treated and control units share the same support. Additionally, balance tests confirmed that, post-matching, the variables have the same distribution between treated and untreated individuals. Details of the analysis to assess the matching quality are provided in Appendix 3.F. In summary, the matching is of quality and robust to multiple changes.

3.4.4 Within-household selection bias

Within migrant households, there may also be self-selection regarding who is chosen to migrate (Chort and Senne, 2015, 2018; Murard, 2019). Indeed, the individuals who migrate often exhibit different characteristics from those left behind, which can result in intra-household selection bias. The longitudinal data enable us to check whether, within these households, migrants are systematically different from non-migrants in terms of their nutritional status. To show that our results are not biased, I need to provide evidence that there is no intra-household selection bias before migration occurs. Following Murard (2019), the following regression is estimated on a sample composed of the left behind and upcoming migrants:

$$Y_{i,w2} = \alpha + \gamma D_{i,w2-w3} + \beta X_{i,w2} + \mu_h + \epsilon_i \tag{3.2}$$

where $Y_{i,w2}$ is the outcome variable (weight, BMI-for-age z-score, or nutritional status) at baseline for individual i (in Wave 2, i.e., before migration occurs), $D_{i,w2-w3}$ is a binary variable which is equal to 1 if the individual i has migrated between waves 2 and 3 and equal to 0 if the individual is left behind, $X_{i,w2}$ a set of individual, household head, and household level variables at baseline, [18] μ_h represents the fixed effects for each household, and ϵ_i is the error term. The results of the regressions are available in Appendix 3.D.4. To sum up, there is no intra-household selection bias, at least related to the nutritional outcomes.

3.5 Results

3.5.1 Adults' results

Table 3.3 presents the average treatment effects of migration on left-behind adults' weight. For all adults (Panel A), the treatment effect of migration on adults' weight ranges from -1.024 to -1.132 kilograms (kg). In other words, everything else held constant, after at least one individual in a household out-migrated looking for a job, left-behind adults experienced a weight decline. The rest of the results (Panels B to F) outline the differences by gender and baseline nutritional status. The results indicate a negative and statistically significant impact of migration on the weight of left-behind men, with coefficients exceeding those for the whole sample. However, there is no significant effect on left-behind women. Finally, when adults are split based on their baseline nutritional status, the nutritional impact of migration predominantly affects healthy left-behind adults.
In relative terms, the weight loss is not substantial. To assess whether this decline can be interpreted as detrimental to adult health, I also investigate the impact of migration on nutritional status rather than body weight. These results are displayed in Table 3.I.1 in the Appendix. In summary, being left behind implies a higher probability of being underweight, but a lower probability of being overweight or obese.

[18] These variables are the same variables used in the kernel-based PSM-DID.

Additionally, I examine transitions between nutritional statuses across waves. Table 3.G.1 in the Appendix presents the dynamics in a transition matrix. Among individuals who were healthy in Wave 2, a greater proportion of the left behind became underweight compared to the non-left behind, which is consistent with the results of Table 3.I.1. However, fewer healthy individuals became overweight or obese when left behind compared to when they were not left behind, and slightly more individuals remained healthy when left behind. Moreover, a higher proportion of overweight or obese individuals transitioned to a healthy status when they were left behind. Overall, even though we find a decrease in adult weight, it does not necessarily translate into an adverse impact on adult health.

3.5.2 Children's results

Table 3.4 displays the results from the kernel PSM-DID for all children, by gender and baseline nutritional status. For all children left behind, we observe a negative and statistically significant coefficient, indicating lower BMI-for-age z-scores. While weight loss for adults is not necessarily detrimental to their health, for children, a decrease in the BMI-for-age z-score can be interpreted as harmful. Indeed, a decrease indicates that the child is losing or not gaining enough weight compared to other children. By splitting the sample between boys and girls, the results differ from those of adults. Indeed, the coefficients of Panel B suggest that the adverse effect is more pronounced for girls. In Appendix 3.J.1, I also examine the results for children under and over 10 years old to distinguish the effects between childhood and adolescence. The results suggest that the detrimental effects of migration are nearly three times higher for younger children, especially young girls, who suffer more compared to their older, potentially more resilient peers.

In line with the adult analysis, I also examine the effect on nutritional statuses and the transition matrices over the two waves. The results on the impact of migration on the probability of being in a particular nutritional status are presented in Table 3.I.2 of the Appendix. Overall, migration increases the probability for left-behind children to be underweight and reduces their probability of being healthy. With 12.9% of children underweight at baseline regardless of migration status, the probability of becoming underweight has increased by 4.4 percentage points for left-behind children, representing a significant impact.

The dynamics in nutritional status are presented in Table 3.G.2. Although fewer children remain overweight/obese among the left-behind compared to non-left-behind children, many more children transition from overweight/obese to underweight status among the left-behind. Furthermore, a significantly higher percentage of left-behind children remain underweight compared to non-left-behind children, who are more likely to transition to a healthy status. In conclusion, migration has a detrimental impact on the health of children left behind.

Table 3.3: Results from PSM-DID for adults, by gender and nutritional status

	Weight			
	(1)	(2)	(3)	(4)
Panel A - All adults				
Left behind*Post	-1.132**	-1.024*	-1.030**	-1.050**
	(0.559)	(0.529)	(0.525)	(0.528)
Mean weight at baseline	62.765	62.765	62.765	62.765
R^2	0.001	0.122	0.136	0.136
Observations	8,754	8,754	8,754	8,754
Panel B - Males				
Left behind*Post	-1.888**	-1.826**	-1.834**	-1.842**
	(0.773)	(0.746)	(0.734)	(0.735)
Mean weight at baseline	63.633	63.633	63.633	63.633
R^2	0.004	0.094	0.127	0.128
Observations	3,826	3,826	3,826	3,826
Panel C - Females				
Left behind*Post	-0.362	-0.109	-0.117	-0.085
	(0.819)	(0.764)	(0.755)	(0.761)
Mean weight at baseline	62.036	62.036	62.036	62.036
R^2	0.004	0.152	0.174	0.174
Observations	4,560	4,560	4,560	4,560
Panel D - Underweight adults				
Left behind*Post	0.708	0.531	0.538	0.967
	(1.157)	(1.054)	(1.053)	(1.072)
Mean weight at baseline	46.819	46.819	46.819	46.819
R^2	0.143	0.332	0.346	0.351
Observations	572	572	572	572
Panel E - Healthy adults				
Left behind*Post	-1.154**	-1.116**	-1.133**	-1.210***
	(0.498)	(0.458)	(0.456)	(0.458)
Mean weight at baseline	58.998	58.998	58.998	58.998
R^2	0.009	0.178	0.188	0.189
Observations	4,764	4,764	4,764	4,764
Panel F - Overweight adults				
Left behind*Post	-1.653	-1.360	-1.351	-1.437
	(1.227)	(1.178)	(1.167)	(1.170)
Mean weight at baseline	75.279	75.279	75.279	75.279
R^2	0.032	0.134	0.156	0.156
Observations	2,506	2,506	2,506	2,506
Individual variables	No	Yes	Yes	Yes
Household head variables	No	Yes	Yes	Yes
Household variables	No	Yes	Yes	Yes
Rural dummy	No	No	Yes	Yes
Region dummies	No	No	Yes	Yes
Number of entries and exits	No	No	No	Yes

Notes: Standard errors are in parentheses. Individual variables include age, gender, and relationship to the head. Household head variables include age, gender, and education level of the head. Household variables include household size, number of working-age members, dependency ratio, and wealth index quintiles. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table 3.4: Results from PSM-DID for children, by gender and nutritional status

	Zbmi			
	(1)	(2)	(3)	(4)
Panel A - All children				
Left behind*Post	-0.193**	-0.242***	-0.228**	-0.259***
	(0.095)	(0.094)	(0.093)	(0.094)
Mean zbmi at baseline	-0.096	-0.096	-0.096	-0.096
R^2	0.002	0.053	0.074	0.075
Observations	5,054	5,054	5,054	5,054
Panel B - Males				
Left behind*Post	-0.198	-0.287**	-0.266**	-0.294**
	(0.134)	(0.131)	(0.130)	(0.131)
Mean zbmi at baseline	-0.123	-0.123	-0.123	-0.123
R^2	0.001	0.076	0.091	0.093
Observations	2,888	2,888	2,888	2,888
Panel C - Females				
Left behind*Post	-0.337**	-0.367***	-0.357***	-0.402***
	(0.139)	(0.138)	(0.136)	(0.138)
Mean zbmi at baseline	-0.063	-0.063	-0.063	-0.063
R^2	0.005	0.040	0.081	0.083
Observations	2,072	2,072	2,072	2,072
Panel D - Underweight children				
Left behind*Post	-0.255	-0.514**	-0.459*	-0.348
	(0.274)	(0.247)	(0.243)	(0.242)
Mean zbmi at baseline	-2.962	-2.962	-2.962	-2.962
R^2	0.431	0.578	0.604	0.620
Observations	450	450	450	450
Panel E - Healthy children				
Left behind*Post	-0.104	-0.130	-0.120	-0.155*
	(0.083)	(0.084)	(0.083)	(0.084)
Mean zbmi at baseline	-0.406	-0.406	-0.406	-0.406
R^2	0.005	0.033	0.044	0.048
Observations	3,240	3,240	3,240	3,240
Panel F - Overweight children				
Left behind*Post	-0.514***	-0.551***	-0.547***	-0.679***
	(0.167)	(0.163)	(0.159)	(0.160)
Mean zbmi at baseline	2.285	2.285	2.285	2.285
R^2	0.398	0.445	0.476	0.486
Observations	1,180	1,180	1,180	1,180
Individual variables	No	Yes	Yes	Yes
Household head variables	No	Yes	Yes	Yes
Household variables	No	Yes	Yes	Yes
Rural dummy	No	No	Yes	Yes
Region dummies	No	No	Yes	Yes
Number of entries and exits	No	No	No	Yes

Notes: Standard errors are in parentheses. Individual variables include age, gender, and relationship to the head. Household head variables include age, gender, and education level of the head. Household variables include household size, number of working-age members, dependency ratio, and wealth index quintiles. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

3.6 Robustness checks

3.6.1 Robustness to potential unobserved variables bias

To ensure that the results are not driven by selection on unobservables, their robustness against omitted variable bias is assessed using the methodology proposed by Oster (2019). The results of the Oster test are provided in Appendix 3.J.2.[19] Overall, the Oster test indicates that the findings are robust to potential omitted variable bias as the coefficient bounds intervals do not contain zero. Additionally, the estimated treatment effects are quite stable as the bias-adjusted β^* (column 3 of Table 3.J.2) are close to the coefficients of the fully controlled model (column 2 of Table 3.J.2). These findings suggest that selection on unobservables is unlikely to drive the results, confirming the robustness of the main identification strategy.

3.6.2 Alternative anthropometric indicator for children

For robustness checks, I investigate the impact of migration on another anthropometric measure for children: height-for-age z-scores. This metric is frequently used to reflect the long-term effects of undernutrition and can indicate chronic malnutrition, thereby capturing the enduring impacts of nutritional changes. The results are available in Table 3.J.3 in the Appendix. In short, migration also has a negative effect of migration on the height-for-age z-scores of left-behind children, but this effect is more pronounced for boys.

3.6.3 Internal migration only

According to Figure 3.A.1, 5.49% of migrants leave their households to find work abroad. The individuals left behind by these international migrants are included in the treatment group, along with those left behind by internal migrants. However, most migrants are internal, which suggests that the results should mostly be representative of internal flows. To confirm this assumption, regressions are performed on a subsample of individuals left behind only by migrants who moved within the country, excluding those left behind by international migrants. The results are available in Tables 3.J.4.1 and 3.J.4.2 in Appendix 3.J.4. Whether for children or adults, the findings are similar to the main results. The main difference is the slightly higher magnitude in the estimates using only internal migration since the adverse effect of migration on the nutrition of the individuals left behind is larger. In summary, these results suggest that the sample is representative of internal migration in Ghana.

3.6.4 Disentangling the potential bias from previous migration

Defining individuals as either left behind or not introduces the potential that among the untreated households, some may have included cases where a former member has already migrated. Since these households are likely the closest regarding covariates used in the matching, there is a risk of matching treated households with other treated households (considered untreated in this case). In essence, there is a concern that within the untreated group, some households might have experienced prior migration, impacting the results beyond the migration between waves 2 and 3. Therefore, I exclude from the sample the untreated households with a migrant for work for more than six months between waves 1 and 2. The

[19] In interpreting the findings, the focus should be on coefficients that are significant in the main results (Tables 3.3 and 3.4), i.e., those highlighted in bold in Table 3.J.2.

aim is to capture the effect of migration between waves 2 and 3, without potential effects from past migrations.

According to Table 3.J.6.2 in Appendix 3.J.6, the results for adults remain consistent with this new specification. The coefficients closely mirror the main findings across healthy adults and those who are overweight or obese. Remarkably, for children, while the coefficient is no longer statistically significant across all columns, it regains significance in column (4), where it encompasses all control variables (Table 3.J.6.3). However, this significance is now only observed at a 10% threshold. The magnitude of the coefficient exhibits a slight decrease for girls but remains statistically significant. Finally, the findings are similarly robust for overweight or obese children. In general, the results are relatively similar; however, some differences persist. Nevertheless, this does not invalidate the main findings. On the contrary, if the concerns were valid and I matched treated individuals with treated individuals (considered untreated between waves 2 and 3), I would not have found results in the main tables. Therefore, the main results represent a lower bound.

I also conduct an additional robustness check in which I exclude all individuals in households with any migration between waves 1 and 2, regardless of their treatment status in Wave 3. In other words, unlike the previous robustness check where I excluded untreated households with a migrant between waves 1 and 2, here I exclude both treated and untreated households in waves 2 and 3 that already had a migrant between waves 1 or 2.[20] This approach ensures that the sample includes only households with no prior migration before Wave 2, allowing for more accurate isolation of the impact of migration between waves 2 and 3. The results are available in Tables 3.J.6.4 and 3.J.6.5 in the Appendix and do not contradict the main results for children.[21]

3.7 Transmission channels

3.7.1 Can remittances offset the negative effects?

I identified a negative effect of migration on the nutritional outcomes of the left behind. I now investigate the mechanisms. The main channel the literature has identified is the direct effect on the income of the household of origin through remittances (Carletto et al., 2011; De Brauw, 2011; Thow et al., 2016). I further investigate this channel by studying the simultaneous impact of an individual's migration outside the household and the receipt of remittances. It raises the question of whether remittances can offset the detrimental impact of the absence of an individual.

In the GSPS, data on received remittances were collected.[22] Using information on transfers, I refined both the control and treatment groups. To study the effect of remittances in addition to the departure of a migrant, the control group is defined as individuals in households where no one migrated between waves 2 and 3 and where no remittances were received. Three treatment groups are also defined: treat-

[20] In the adult sample, individuals who were already in households with at least one migrant between waves 1 and 2 account for 429 adults (9.37% of adults). In the children's sample, 263 children (9.46% of children). Table 3.J.6.1 describes the distribution of individuals already in migrant households between waves 1 and 2 according to their status in waves 2 and 3.

[21] While one could argue that earlier migrations before Wave 1 might still have an impact, the fact that the robustness check using Wave 1 data, with an eight-year gap from Wave 3, does not contradict the main results suggests that it is unlikely these earlier migrations would have a different effect.

[22] The data does not permit the attribution of remittances to specific individuals, leaving open the possibility of other sources than migrants. Indeed, remittances may be received from a previous household member but also from a migrant who has been outside the household for longer than the baseline wave. Moreover, the sender may be a friend or relative who never belonged to the household.

ment (A) includes individuals from households with a migrant but no received remittances; treatment (B) consists of those from households with remittance receipts but no migrant; treatment (C) incorporates individuals from households experiencing both migration and remittance inflows. The different treatments are meant to capture different effects. Treatment (A) captures the effect of the change in household composition, similar to the previous regressions, although the sample is different. Treatment (B) captures the sole effect of remittances. Finally, treatment (C) reflects the simultaneous effect of migration and remittances.

Table 3.K.1 in the Appendix provides the results. In column (1), adults from migrant households not receiving remittances are compared to adults from households without migrants and not receiving remittances. The magnitudes reveal a more pronounced negative impact on weight across all adults, healthy adults, and men compared to prior results. The results for children are displayed in Table 3.K.2. The coefficient size for girls remains consistent with previous findings. However, for overweight/obese children, the coefficient exhibits an increase, although the signs and significance levels conform to the prior results of Table 3.4.

Results from column (2) of Table 3.K.1 highlight that the receipt of remittances without migration has no impact on the weight of all adults and only seems to have a negative impact on healthy adults. However, from Table 3.K.2, only receiving remittances without having a migrant seems to impact the nutrition of all children and underweight children positively. It is the first observed positive impact and means that the sole receipt of remittances without having a migrant has a positive impact on vulnerable individuals, namely underweight children. However, it negatively impacts the nutrition of overweight and obese children.

Lastly, in column (3) of Table 3.K.1, for men, the weight decline induced by the simultaneous effect of having a migrant and receiving remittances is around the same as having a migrant only but is significantly greater for healthy adults. Among children (Table 3.K.2), the combined effect of remittance receipt and migration also has a negative impact on the z-scores of all children. In contrast, migration only (column 1) had no significant effect. The same is true for boys and healthy children. Interestingly, the combination of migration and remittance receipt no longer exerts a statistically significant effect on the BMI-for-age z-scores of overweight/obese children and girls.

In summary, remittances have no protective effect on adults and accentuate the negative effect on healthy adults' weight. This finding contradicts the view that migration, often undertaken to alleviate the financial burden on the household, would yield positive effects. Conversely, for children, especially the most vulnerable, remittances appear to partially mitigate the adverse effects stemming from the absence of migrants. Intriguingly, the isolated impact of remittances alone yields a positive impact in the z-scores for both all children and those who are underweight. Therefore, I probably capture the impact of the onset of migration. Indeed, while I find a harmful impact, there may be positive effects that might counterbalance in the long run. Migration is a lengthy process as it takes time for migrants to find work, settle down, and send remittances. The positive effect is likely due to remittances from individuals who have been migrating before Wave 2 and are already settled at their destination. These results may represent the long-term effects.

3.7.2 The disruptive effect of migration

Since remittances do not offset the adverse effects, I investigate alternative mechanisms that could explain how migration impacts the nutrition of those left behind, focusing on migrant characteristics. Table 3.L.1 in the Appendix shows that nearly 80% of migrants were the household head's children, predominantly young males (aged 15 to 24) who are often more educated. This pattern suggests a natural transition: as these young men reach adulthood, many migrate to pursue better economic opportunities, with around a third relocating to Accra. Thus, these individuals were already predisposed to migrate. The migration of these young males represents a typical progression, as they grow up and leave their families to establish themselves elsewhere. However, importantly, many of these migrants previously contributed to the household income. Indeed, nearly 40% of them worked on a farm plot, presumably generating income. Additionally, although half of the migrants were students, about a third of them were also working, either as owners of a non-farm enterprise or farm plot, or as workers in non-farm enterprises or farm plots. In summary, most migrants were significant contributors to the income of their household of origin.

The departure of these individuals likely represents a loss of income for their households, inducing a negative income shock. Coupled with the fact that the migration itself may require an investment, the households experience a double-edged financial strain. Consequently, this financial strain manifests most palpably in the nutritional status of the remaining children. This disruptive effect of migration, primarily characterized by an adverse income shock for the household of origin, is the main driver.

Regarding the heterogeneous effects, while adults may display greater resilience, with relatively small impacts on their nutritional status, children endure the harshest repercussions, potentially affecting their long-term growth. Notably, gendered differences are significant, with a harsher impact on girls; however, the literature lacks consistent evidence of a systematically more negative effect by gender (Fellmeth et al., 2018). Given the disruptive effect and potential negative income shocks, one explanation for this greater impact on girls could be a preference for boys, possibly due to their perceived current or future economic contributions. However, there is limited literature to support such claims about preferences for boys in terms of nutrition, and evidence suggests no clear gender preferences in Ghana, or more broadly in sub-Saharan Africa (Rossi and Rouanet, 2015). In such cases, context matters, including household economic structures and kinship systems.

Additionally, in defining migration, I focus on migrants seeking employment who have been absent for at least six months. Given this time frame, some might have left precisely six months prior. As a result, some migrants might be in the initial stages of their journey, possibly struggling to find employment and to settle in properly. During this period, their families back home might suffer income loss as they wait for financial support. This immediate economic strain can adversely affect nutritional outcomes, especially for children at critical developmental stages. Over time, remittances may have a beneficial effect on nutrition, but the effect is likely a short- to medium-term effect arising from the destabilizing influence of migration on the individuals left behind.

3.7.3 Other transmission channels

The time effect may also shed light on our findings. Specifically, this effect captures how migration alters the time allocated to remaining household members, particularly children (De Brauw and Mu, 2015). Should the migrant be a parent, there could be implications for child health through reduced parental attention as they may spend less time with them (De Brauw and Mu, 2011). Among the potential consequences, parental migration, especially maternal migration, can have negative effects on a child's diet (Bai et al., 2022).

To explore the impact of parental migration, I leverage data on individuals' past and present co-residence.[23] The analysis focuses on children who were co-resident with at least one parent in Wave 2 but experienced the departure of a parent between waves 2 and 3, irrespective of the reason. Table 3.M.2 in the Appendix reports the average treatment effects of parental migration on children's BMI-for-age z-score. The results indicate a negative and significant effect of parental migration on children's BMI-for-age z-scores. Therefore, regardless of the motive, and consistent with the literature, parental migration can contribute to explain the adverse impact on the nutritional status of children left behind.

Additionally, as shown in Table 3.L.1, migrants were typically the older children in their households. Their departure leaves younger siblings behind. The absence of these older siblings may be particularly detrimental given their potential active involvement in essential household chores such as meal preparation, which directly affects the nutrition of their younger siblings. Consequently, their departure could also contribute to the negative impact observed on the nutritional status of the remaining siblings, perhaps because the migrating siblings had previously been attentive to their younger siblings' nutritional needs. Also, it has been shown that parental migration may lead to increased time spent on farm and domestic work for left-behind children (Chang et al., 2011). This could potentially compromise their overall health and, by extension, their nutritional status.

In Appendix 3.N, I also inquire how and whether the effects can be attributed to changes in food consumption. The survey has questions about the food items households have produced, purchased, and received over the previous 30 days. Therefore, I examine if individual changes may also stem from household-level changes in dietary consumption. I examine the impact of migration on household food consumption, covering food acquired through purchases, self-production, and received as gifts, alongside total household consumption, using the kernel PSM-DID method. Additionally, I assess the effect of migration on per capita consumption among various food groups and evaluate how migration influences the composition of the household's total food consumption, focusing on the proportion of different food groups. Finally, I also analyze the impact on household food diversity with the Simpson and Shannon indices. In summary, migration significantly reduces food purchases while increasing self-produced food consumption at the household level. Migration also negatively affects the consumption of fruits, vegetables, and eggs, suggesting a decline in dietary quality. Finally, the Simpson index shows a negative impact, indicating reduced dietary diversity.[24]

3.8 Conclusion

Being left behind by an internal migrant in Ghana leads to worsening nutrition, at least in the short term. Particularly, when an individual migrates for work outside the household, it negatively impacts the nutritional status of the children left behind. By further exploring the mechanisms, I did not find that remittances had an offsetting effect for all individuals. However, when the household does not have a migrant between the survey waves and solely receives remittances, probably reflecting the long-term effects of migration, remittances have positive effects on the nutrition of children. Most likely, the

[23] The details of the questions used in the questionnaires, as well as the definitions of the control and treatment groups, are provided in Appendix 3.M.

[24] The interested reader can refer to Appendix 3.N for the results and their implications for understanding the mechanisms of food consumption.

mechanism driving the negative effects is related to the disruptive impact of migration, which often leads to the disorganization of the household of origin after the migrant leaves. A member's migration can incur costs, whether due to the investment in migration or the loss of income resulting from the migrant's departure. Consequently, this short-term negative shock could explain the adverse impact on children's nutritional status, potentially affecting their long-term growth prospects, while adults may recover more readily.

One limitation is the decision not to capture individuals who might change households following migration. Indeed, migration can lead to the mobility of individuals and the dissolution of households (Bertoli and Murard, 2020). Changes in living arrangements following migration can induce individuals to join a new housing unit. For instance, children left behind can start co-residing with their grandparents (Bertoli et al., 2023). In this paper, I exclude these children as they do not fit the definition of left behind. It is also the case for individuals who may enter or leave the household. This analysis, therefore, neglects these individuals who adjusted their living arrangements. Future research could examine the effect of migration on these relocated individuals and the consequences of different living arrangements.

References of Chapter 3

Antman, F. M. (2013): "The impact of migration on family left behind," *International Handbook on the Economics of Migration*, 293–308.

Antón, J.-I. (2010): "The Impact of Remittances on Nutritional Status of Children in Ecuador," *International Migration Review*, 44, 269–299.

Bai, Y., N. Yang, L. Wang, and S. Zhang (2022): "The impacts of maternal migration on the cognitive development of preschool-aged children left behind in rural China," *World Development*, 158, 106007.

Bai, Y., L. Zhang, C. Liu, Y. Shi, D. Mo, and S. Rozelle (2018): "Effect of Parental Migration on the Academic Performance of Left Behind Children in North Western China," *The Journal of Development Studies*, 54, 1154–1170.

Baulch, B. and A. Quisumbing (2011): "Testing and adjusting for attrition in household panel data. CPRC Toolkit Note," .

Bertoli, S., E. Gautrain, and E. Murard (2023): "Left behind, but not immobile: Living arrangements of Mexican transnational households," *Economic Development and Cultural Change*.

Bertoli, S. and F. Marchetta (2014): "Migration, Remittances and Poverty in Ecuador," *The Journal of Development Studies*, 50, 1067–1089.

Bertoli, S. and E. Murard (2020): "Migration and co-residence choices: Evidence from Mexico," *Journal of Development Economics*, 142, 102330.

Böhme, M. H., R. Persian, and T. Stöhr (2015): "Alone but better off? Adult child migration and health of elderly parents in Moldova," *Journal of Health Economics*, 39, 211–227.

Caliendo, M. and S. Kopeinig (2008): "Some Practical Guidance for the Implementation of Propensity Score Matching," *Journal of Economic Surveys*, 22, 31–72.

Carletto, C., K. Covarrubias, and J. A. Maluccio (2011): "Migration and child growth in rural Guatemala," *Food Policy*, 36, 16–27.

Casu, L., M. van den Bold, L. Diop, R. N. O. Aryeetey, and R. Verstraeten (2021): "Double burden of malnutrition in Ghana: a holistic perspective," TNWA Evidence, dakar, Senegal: International Food Research Institute (IFPRI).

Chang, H., X.-y. Dong, and F. MacPhail (2011): "Labor Migration and Time Use Patterns of the Left-behind Children and Elderly in Rural China," *World Development*, 39, 2199–2210.

Chort, I. and J.-N. Senne (2015): "Selection into Migration within a Household Model: Evidence from Senegal," *The World Bank Economic Review*, 29, S247–S256.

——— (2018): "You'll Be a Migrant, My Son: Accounting for Migrant Selection within the Household," *Economic Development and Cultural Change*, 66, 217–263.

Davis, J. and N. Brazil (2016): "Migration, Remittances and Nutrition Outcomes of Left-Behind Children: A National-Level Quantitative Assessment of Guatemala," *PLOS ONE*, 11.

De Brauw, A. (2011): "Migration and child development during the food price crisis in El Salvador," *Food Policy*, 36, 28–40.

De Brauw, A. and R. Mu (2011): "Migration and the overweight and underweight status of children in rural China," *Food Policy*, 36, 88–100.

——— (2015): "Migration and young child nutrition: evidence from rural China," *Journal of Population Economics*, 28, 631–657.

Démurger, S. (2015): "Migration and families left behind," *IZA World of Labor*.

Fellmeth, G., K. Rose-Clarke, C. Zhao, L. K. Busert, Y. Zheng, A. Massazza, H. Sonmez, B. Eder, A. Blewitt, W. Lertgrai, M. Orcutt, K. Ricci, O. Mohamed-Ahmed, R. Burns, D. Knipe, S. Hargreaves, T. Hesketh, C. Opondo, and D. Devakumar (2018): "Health impacts of parental migration on left-behind children and adolescents: a systematic review and meta-analysis," *The Lancet*, 392, 2567–2582.

Fitzgerald, J., P. Gottschalk, and R. Moffitt (1998): "An Analysis of Sample Attrition in Panel Data: The Michigan Panel Study of Income Dynamics," *The Journal of Human Resources*, 33, 251–299.

Gao, Y., L. P. Li, J. H. Kim, N. Congdon, J. Lau, and S. Griffiths (2010): "The impact of parental migration on health status and health behaviours among left behind adolescent school children in China," *BMC Public Health*, 10.

Ghana Statistical Service (2013): "2010 Population and Housing Census: National Analytical Report." Tech. rep., Ghana Statistical Service.

Gibson, J., D. McKenzie, and S. Stillman (2011): "The Impacts of International Migration on Remaining Household Members: Omnibus Results from a Migration Lottery Program," *The Review of Economics and Statistics*, 93, 1297–1318.

——— (2013): "Accounting for Selectivity and Duration-Dependent Heterogeneity When Estimating the Impact of Emigration on Incomes and Poverty in Sending Areas," *Economic Development and Cultural Change*, 61, 247–280.

Guo, X., T. A. Mroz, B. M. Popkin, and F. Zhai (2000): "Structural Change in the Impact of Income on Food Consumption in China, 1989–1993," *Economic Development and Cultural Change*, 48, 737–760.

Hildebrandt, N., D. J. McKenzie, G. Esquivel, and E. Schargrodsky (2005): "The Effects of Migration on Child Health in Mexico [with Comments]," *Economía*, 6, 257–289.

International Organization for Migration (2020): "Migration in Ghana: A Country Profile 2019," Tech. rep.

Islam, M. M., M. N. Khan, and M. N. I. Mondal (2019): "Does parental migration have any impact on nutritional disorders among left-behind children in Bangladesh?" *Public Health Nutrition*, 22, 95–103.

Karamba, W. R., E. J. Quiñones, and P. Winters (2011): "Migration and food consumption patterns in Ghana," *Food Policy*, 36, 41–53.

Kennan, J. and J. R. Walker (2011): "The Effect of Expected Income on Individual Migration Decisions," *Econometrica*, 79, 211–251.

Kennedy, E. and P. Peters (1992): "Household food security and child nutrition: the interaction of income and gender of household head," *World Development*, 20, 1077–1085.

Khandker, S. R., G. B. Koolwal, and H. A. Samad (2009): *Handbook on Impact Evaluation : Quantitative Methods and Practices. World Bank.*

Lei, L., F. Liu, and E. Hill (2018): "Labour Migration and Health of Left-Behind Children in China," *The Journal of Development Studies*, 54, 93–110.

Liu, C., T. Eriksson, and F. Yi (2021a): "Offspring migration and nutritional status of left-behind older adults in rural China," *Economics & Human Biology*, 41.

Liu, M., X.-l. Feng, S.-g. Wang, and Y. Zhong (2021b): "Does poverty-alleviation-based industry development improve farmers' livelihood capital?" *Journal of Integrative Agriculture*, 20, 915–926.

Lu, H., I. T. Kandilov, and R. Zhu (2020): "The Impact of Internal Migration on the Health of Rural Migrants: Evidence from Longitudinal Data in China," *The Journal of Development Studies*, 56, 840–855.

Marchetta, F. and S. Sim (2021): "The effect of parental migration on the schooling of children left behind in rural Cambodia," *World Development*, 146, 105593.

McAuliffe, M. and L. A. Oucho (2024): "World Migration Report 2024," Tech. rep., Geneva, International Organization for Migration (IOM), iSBN: 1561-5502.

Mora-Rivera, J. and E. van Gameren (2021): "The impact of remittances on food insecurity: Evidence from Mexico," *World Development*, 140, 105349.

Mu, R. and D. P. van de Walle (2009): "Left Behind to Farm? Women's Labor Re-Allocation in Rural China," .

Murard, E. (2019): "The Impact of Migration on Family Left Behind: Estimation in Presence of Intra-Household Selection of Migrants," *IZA Discussion Paper*.

Nguyen, M. C. and P. Winters (2011): "The impact of migration on food consumption patterns: The case of Vietnam," *Food Policy*, 36, 71–87.

Oster, E. (2019): "Unobservable Selection and Coefficient Stability: Theory and Evidence," *Journal of Business & Economic Statistics*, 37, 187–204.

Ponce, J., I. Olivié, and M. Onofa (2011): "The role of international remittances in health outcomes in Ecuador: prevention and response to shocks," *The International Migration Review*, 45, 727–745.

Quartey, P. (2006): "The impact of migrant remittances on household welfare in Ghana," *African Economic Research Consortium*, AERC research paper.

Rosenbaum, P. R. and D. B. Rubin (1983): "The central role of the propensity score in observational studies for causal effects," *Biometrika*, 70, 41–55.

Rossi, P. and L. Rouanet (2015): "Gender Preferences in Africa: A Comparative Analysis of Fertility Choices," *World Development*, 72, 326–345.

Stark, O. and D. E. Bloom (1985): "The New Economics of Labor Migration," *American Economic Review*, 75, 173–178.

Stark, O. and R. E. B. Lucas (1988): "Migration, Remittances, and the Family," *Economic Development and Cultural Change*, 36, 465–481.

Stephen, E. H. and F. D. Bean (1992): "Assimilation, Disruption and the Fertility of Mexican-Origin Women in the United States," *International Migration Review*, 26, 67–88.

Stuart, E. A., H. A. Huskamp, K. Duckworth, J. Simmons, Z. Song, M. E. Chernew, and C. L. Barry (2014): "Using propensity scores in difference-in-differences models to estimate the effects of a policy change," 14, 166–182.

Sznajder, K. K., K. Wander, S. Mattison, E. Medina-Romero, N. Alam, R. Raqib, A. Kumar, F. Haque, T. Blumenfield, and M. K. Shenk (2021): "Labor migration is associated with lower rates of underweight and higher rates of obesity among left-behind wives in rural Bangladesh: a cross-sectional study," *Globalization and Health*, 17, 81.

Thow, A. M., J. Fanzo, and J. Negin (2016): "A Systematic Review of the Effect of Remittances on Diet and Nutrition," *Food and Nutrition Bulletin*, 37, 42–64.

Tian, X., C. Ding, C. Shen, and H. Wang (2017): "Does Parental Migration Have Negative Impact on the Growth of Left-Behind Children? New Evidence from Longitudinal Data in Rural China," *International Journal of Environmental Research and Public Health*, 14.

Viet Nguyen, C. (2016): "Does parental migration really benefit left-behind children? Comparative evidence from Ethiopia, India, Peru and Vietnam," *Social Science & Medicine*, 153, 230–239.

Vikram, K. (2023): "Timing and Frequency of Fathers' Migration and Nutritional Status of Left-Behind Children in India: A Life Course Approach," *Population Research and Policy Review*, 42, 7.

Wei, M.-F. (2022): "Does Migration Matter? The Impacts on Senior Food and Nutrition Security in Rural China," *International Journal of Food and Agricultural Economics*, 10.

WHO (1995): "Physical Status: The Use and Interpretation of Anthropometry," World Health Organization Technical Report Series 854.

——— (2006): *WHO child growth standards: length/height-for-age, weight-for-age, weight-for-length, weight-for-height and body mass index-for-age ; methods and development*, World Health Organization.

WHO and UNICEF (2019): *Recommendations for data collection and analysis and reporting on anthropometric indicators in children under 5 years old*, Geneva.

Yang, D. (2011): "Migrant Remittances," *Journal of Economic Perspectives*, 25, 129–152.

Yi, F., C. Liu, and Z. Xu (2019): "Identifying the effects of migration on parental health: Evidence from left-behind elders in China," *China Economic Review*, 54, 218–236.

Zezza, A., C. Carletto, B. Davis, and P. Winters (2011): "Assessing the impact of migration on food and nutrition security," *Food Policy*, 36, 1–6.

Appendix to Chapter 3

3.A Migrant destinations

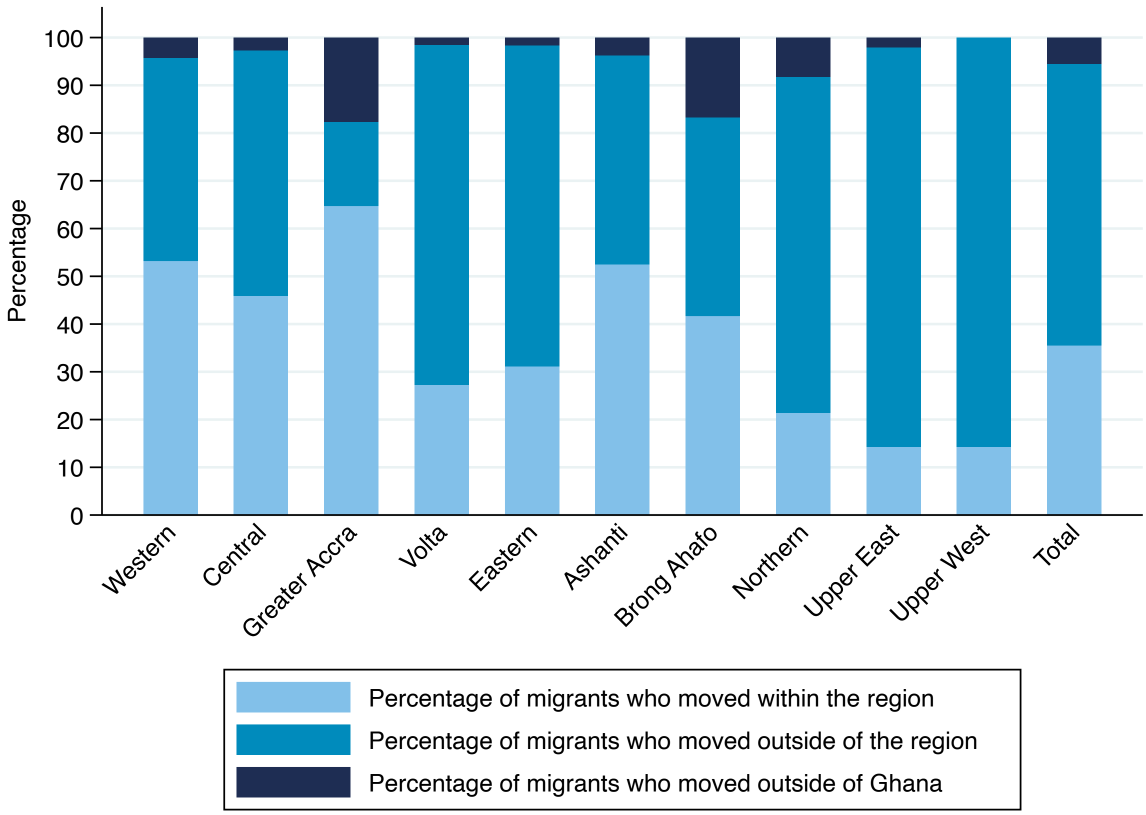


Figure 3.A.1: Destination of migrants according to their region of origin

Notes: Construction by the author using the EGC-ISSER Socioeconomic Panel Survey. Migrants are here defined as individuals who moved out from their households looking for work for more than six months between waves 2 and 3.

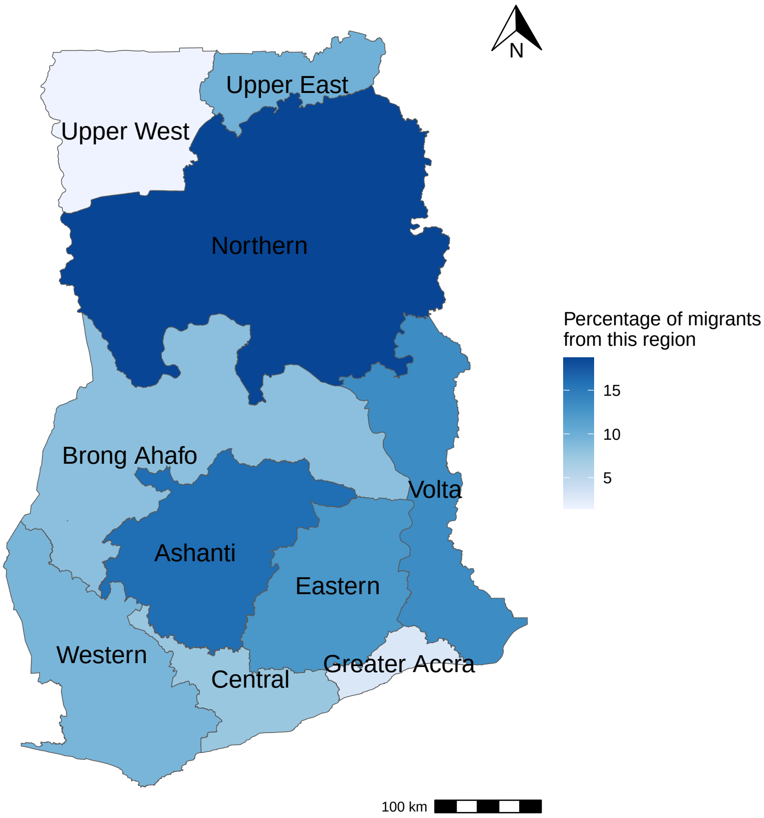


Figure 3.A.2: Region of origin of migrants

Notes: Construction by the author using the EGC-ISSER Socioeconomic Panel Survey. Migrants are here defined as individuals who moved out from their households looking for work for more than six months between waves 2 and 3.

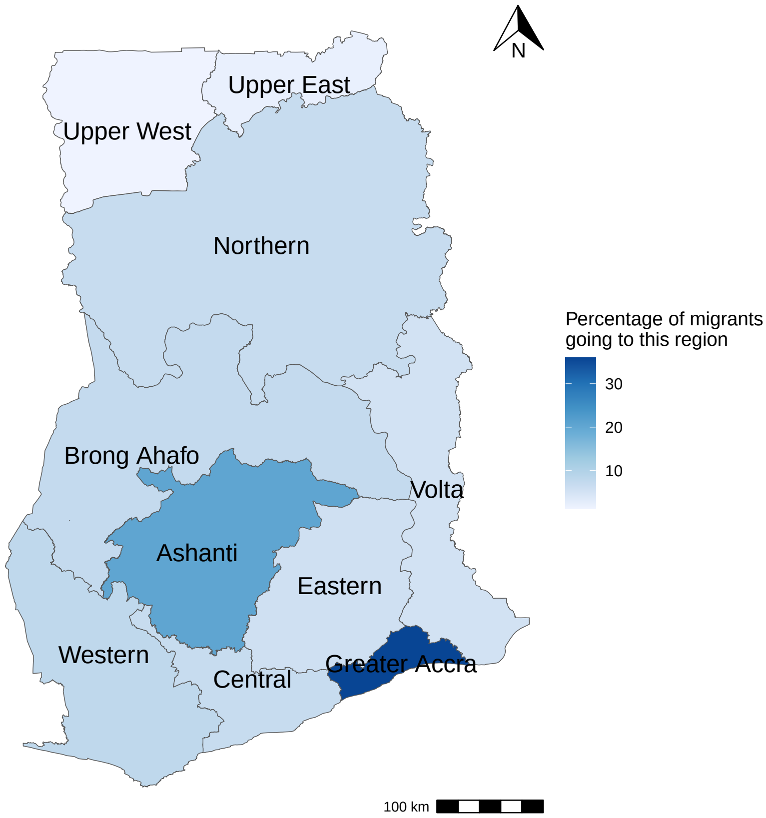


Figure 3.A.3: Region of destination of migrants

Notes: Construction by the author using the EGC-ISSER Socioeconomic Panel Survey. Migrants are here defined as individuals who moved out from their households looking for work for more than six months between waves 2 and 3. Migrants moving within the region are included.

3.B Variables used in the analysis

Table 3.B.1: Definition of the variables

Variable	Definition
Anthropometric variables	
Zbmi	BMI-for-age z-score of children between 2 and 18 years old (in standard deviation)
Weight	Body weight of adults (in kilograms)
Individual level variables	
Age	Age of the individual (in years)
Male	Individual is a male (1 = yes; 0 otherwise)
Relationship to the head	Dummies of the relationship to the household head (Head; Spouse; Child; Grandchild; Other)
Household head level variables	
Head age	Age of the household head (in years)
Male head	Household head is a male (1 = yes; 0 otherwise)
Education of the head	Dummies of the education level of the household head (None or preschool; Primary education; Post-primary education, Secondary education; Tertiary education)
Household level variables	
Household size	Number of individuals in the same dwelling
Working-age members	Number of working-age members (aged 15 to 64) in the household
Dependency ratio	Number of dependents (aged 0 to 14 and over the age of 65) in the household divided by the number of working-age members. When the household is only composed of dependent individuals, the missing value is replaced by the maximum value of the sample.
Rural	Household living area (0 = urban; 1 = rural)
Wealth index	Quintiles of a wealth index based on a Principal Component Analysis (PCA) using housing characteristics and durable goods owned by the household
Region	Dummies of the household living region (Western Region; Central Region; Greater Accra Region; Volta Region; Eastern Region; Ashanti Region; Brong-Ahafo Region; Northern Region; Upper East Region; Upper West Region)
Number of entries	Number of entries of individuals in the household in Wave 3
Number of exits	Number of exits of individuals from the household in Wave 3 (excluding labor migrants)

3.C Descriptive statistics

Table 3.C.1: Descriptive statistics of all sample in Wave 2

	Adults		Children	
	Mean	(SD)	Mean	(SD)
Individual variables				
Weight	62.765	(13.327)		
Zbmi			-0.096	(1.794)
Age	45.839	(15.671)	8.655	(3.724)
Male	0.456	(0.498)	0.560	(0.496)
Relationship to the head				
Head	0.606	(0.489)	0.000	(0.000)
Spouse	0.278	(0.448)	0.002	(0.046)
Child	0.087	(0.282)	0.877	(0.329)
Grandchild	0.007	(0.083)	0.088	(0.284)
Other relationship	0.022	(0.145)	0.032	(0.177)
Household head variables				
Age of the head	51.442	(15.139)	47.046	(12.296)
Male head	0.736	(0.441)	0.759	(0.428)
Education of the head				
Head, none or preschool	0.366	(0.482)	0.424	(0.494)
Head, primary education	0.148	(0.355)	0.150	(0.357)
Head, post-primary education	0.345	(0.476)	0.310	(0.463)
Head, secondary education	0.078	(0.269)	0.065	(0.247)
Head, tertiary education	0.062	(0.241)	0.050	(0.219)
Household variables				
Household size	4.449	(2.715)	6.237	(2.595)
Working-age members	2.489	(1.541)	2.817	(1.409)
Dependency ratio	1.090	(1.373)	1.465	(1.018)
Rural	0.637	(0.481)	0.712	(0.453)
Wealth index				
Wealth index, 1st quintile	0.271	(0.444)	0.336	(0.472)
Wealth index, 2nd quintile	0.169	(0.375)	0.159	(0.365)
Wealth index, 3rd quintile	0.191	(0.393)	0.172	(0.377)
Wealth index, 4th quintile	0.201	(0.401)	0.187	(0.390)
Wealth index, 5th quintile	0.168	(0.374)	0.146	(0.354)
Region				
Western Region	0.074	(0.261)	0.074	(0.263)
Central Region	0.075	(0.263)	0.068	(0.252)
Greater Accra Region	0.098	(0.298)	0.061	(0.240)
Volta Region	0.085	(0.279)	0.068	(0.251)
Eastern Region	0.101	(0.301)	0.085	(0.278)
Ashanti Region	0.162	(0.368)	0.154	(0.361)
Brong-Ahafo Region	0.103	(0.304)	0.111	(0.314)
Northern Region	0.195	(0.396)	0.274	(0.446)
Upper East Region	0.077	(0.267)	0.077	(0.267)
Upper West Region	0.030	(0.171)	0.028	(0.164)
Observations	4,579		2,780	

Notes: Zbmi refers to the BMI-for-age z-score.

Table 3.C.2: Descriptive statistics of adults by migration status in wave 3

	Adults				
	Non Left Behind		Left Behind		t-test
	Mean	(SD)	Mean	(SD)	P-value
Individual variables					
Weight	63.638	(12.908)	60.506	(12.381)	0.000
Age	49.545	(15.715)	52.891	(15.480)	0.000
Male	0.463	(0.499)	0.394	(0.489)	0.004
Relationship to the head					
Head	0.623	(0.485)	0.464	(0.499)	0.000
Spouse	0.268	(0.443)	0.360	(0.480)	0.000
Child	0.082	(0.275)	0.132	(0.339)	0.000
Grandchild	0.007	(0.081)	0.011	(0.103)	0.316
Other relationship	0.020	(0.141)	0.034	(0.182)	0.051
Household head variables					
Age of the head	54.236	(14.979)	60.223	(13.428)	0.000
Male head	0.720	(0.449)	0.713	(0.453)	0.754
Education of the head					
Head, none or preschool	0.344	(0.475)	0.509	(0.500)	0.000
Head, primary education	0.149	(0.357)	0.115	(0.319)	0.045
Head, post-primary education	0.364	(0.481)	0.274	(0.447)	0.000
Head, secondary education	0.075	(0.263)	0.034	(0.182)	0.001
Head, tertiary education	0.067	(0.250)	0.068	(0.252)	0.940
Household variables					
Household size	4.197	(2.582)	4.740	(2.385)	0.000
Working-age members	2.356	(1.523)	2.719	(1.659)	0.000
Dependency ratio	1.317	(1.865)	1.151	(1.468)	0.061
Rural	0.618	(0.486)	0.743	(0.438)	0.000
Wealth index					
Wealth index, 1st quintile	0.244	(0.430)	0.353	(0.478)	0.000
Wealth index, 2nd quintile	0.183	(0.387)	0.215	(0.411)	0.098
Wealth index, 3rd quintile	0.193	(0.395)	0.138	(0.346)	0.004
Wealth index, 4th quintile	0.194	(0.395)	0.187	(0.391)	0.736
Wealth index, 5th quintile	0.185	(0.388)	0.106	(0.309)	0.000
Region					
Western Region	0.076	(0.265)	0.049	(0.216)	0.035
Central Region	0.077	(0.267)	0.051	(0.220)	0.041
Greater Accra Region	0.110	(0.314)	0.017	(0.129)	0.000
Volta Region	0.082	(0.274)	0.106	(0.309)	0.072
Eastern Region	0.101	(0.302)	0.083	(0.276)	0.210
Ashanti Region	0.168	(0.374)	0.126	(0.332)	0.019
Brong-Ahafo Region	0.101	(0.301)	0.115	(0.319)	0.329
Northern Region	0.183	(0.387)	0.300	(0.459)	0.000
Upper East Region	0.070	(0.256)	0.134	(0.341)	0.001
Upper West Region	0.031	(0.174)	0.019	(0.137)	0.141
Observations	4,109		470		

Table 3.C.3: Descriptive statistics of children by migration status in wave 3

	Children				
	Non Left Behind		Left Behind		*t*-test
	Mean	(SD)	Mean	(SD)	*P*-value
Individual variables					
Zbmi	-0.009	(1.590)	-0.256	(1.657)	0.010
Age	12.409	(3.723)	12.618	(3.641)	0.349
Male	0.560	(0.496)	0.561	(0.497)	0.987
Relationship to the head					
Head	0.000	(0.000)	0.000	(0.000)	0.000
Spouse	0.002	(0.045)	0.003	(0.056)	0.677
Child	0.880	(0.325)	0.850	(0.357)	0.127
Grandchild	0.085	(0.279)	0.118	(0.323)	0.052
Other relationship	0.033	(0.178)	0.029	(0.167)	0.693
Household head variables					
Age of the head	49.733	(12.021)	56.146	(11.411)	0.000
Male head	0.735	(0.442)	0.761	(0.427)	0.318
Education of the head					
Head, none or preschool	0.402	(0.490)	0.586	(0.493)	0.000
Head, primary education	0.150	(0.357)	0.131	(0.337)	0.370
Head, post-primary education	0.336	(0.472)	0.223	(0.417)	0.000
Head, secondary education	0.068	(0.252)	0.013	(0.112)	0.000
Head, tertiary education	0.044	(0.206)	0.048	(0.214)	0.773
Household variables					
Household size	6.085	(2.540)	6.137	(2.550)	0.734
Working-age members	3.087	(1.469)	3.194	(1.582)	0.226
Dependency ratio	1.226	(1.057)	1.203	(1.131)	0.718
Rural	0.689	(0.463)	0.882	(0.323)	0.000
Wealth index					
Wealth index, 1st quintile	0.293	(0.455)	0.398	(0.490)	0.000
Wealth index, 2nd quintile	0.173	(0.378)	0.248	(0.433)	0.001
Wealth index, 3rd quintile	0.172	(0.377)	0.131	(0.337)	0.067
Wealth index, 4th quintile	0.187	(0.390)	0.143	(0.351)	0.057
Wealth index, 5th quintile	0.176	(0.381)	0.080	(0.271)	0.000
Region					
Western Region	0.080	(0.271)	0.032	(0.176)	0.002
Central Region	0.070	(0.255)	0.051	(0.220)	0.203
Greater Accra Region	0.068	(0.252)	0.013	(0.112)	0.000
Volta Region	0.065	(0.246)	0.089	(0.285)	0.107
Eastern Region	0.089	(0.285)	0.041	(0.200)	0.004
Ashanti Region	0.163	(0.369)	0.089	(0.285)	0.001
Brong-Ahafo Region	0.108	(0.310)	0.131	(0.337)	0.227
Northern Region	0.259	(0.438)	0.395	(0.490)	0.000
Upper East Region	0.069	(0.253)	0.146	(0.354)	0.001
Upper West Region	0.030	(0.170)	0.013	(0.112)	0.086
Observations	2,466		314		

Notes: Zbmi refers to the BMI-for-age z-score.

Table 3.C.4: Distribution of nutritional status by wave and migration status

Nutritional Status	Wave 2		Wave 3	
	Non Left Behind	Left Behind	Non Left Behind	Left Behind
Adults				
Underweight	9.66	10.43	9.00	11.91
Healthy	59.45	63.19	53.01	62.98
Overweight/Obese	30.88	26.38	37.99	25.11
Children				
Underweight	12.81	13.38	8.52	11.78
Healthy	63.18	64.97	69.79	70.38
Overweight/Obese	24.01	21.66	21.70	17.83

Notes: All values are in percentages.

3.D Attrition, missing values, and selection bias

3.D.1 Attrition and missing values

Attrition Between the two survey waves, 24.28% of individuals dropped out from the sample, i.e., they were interviewed in Wave 2 but were not re-interviewed in Wave 3 and thus are no longer included in the panel.[25] Consequently, we need to check for potential attrition bias. Indeed it can introduce bias if those who drop out differ systematically from those who remain in the sample.[26] To investigate the extent of attrition bias, I estimate an attrition probit model in which I explain attrition between waves 2 and 3 with a set of characteristics. The variables are the same as in the main analysis, except that I add two additional variables at the individual level (whether the individual has been married and his education level). The results of the probits to check whether attrition is random or driven by observable characteristics are displayed in Appendix 3.D.2.

According to the tables from Appendix 3.D.2, the adult weight and child BMI-for-age z-score are not significant predictors of attrition. In contrast, other variables are significant for both adults and children. As a result, although the outcomes and some variables are not significant predictors of attrition, the results may still raise concerns that our analysis suffers from attrition bias. Thus, as a precaution against attrition bias, we ensure that our main results are not biased by reweighting our observations using the inverse probability weighting procedure (Baulch and Quisumbing, 2011; Fitzgerald et al., 1998). Further details of this reweighting procedure are available in Appendix 3.D.3. Overall, the results with the inverse probability weights are not different from the main results.

Additionally, attrition may be linked to migration. To assess any potential link between migration and attrition, I refer to Table 3.D.1.1. This table shows no significant difference in the likelihood of belonging to a migrant household between attritors and non-attritors, suggesting that attrition is not linked to migration. Here, it is possible to identify migrants among attritors because at least one individual remains in the household in Wave 3. Thus, this analysis focuses on individual-level attrition.

However, attrition can also occur at the household level, where the entire household drops out of the sample between the two waves. The vast majority of attrition (around 89%) is individual-level, while approximately 11% corresponds to household-level attrition. For these households, migration status cannot be determined, as the entire household exits the panel, making it impossible to identify whether attrition is linked to migration. Nevertheless, household-level attrition represents less than 11% of total attrition, suggesting a low probability that any difference would substantially affect the results.

Table 3.D.1.1: Comparison of mean values of being in a migrant household for attritors and non-attritors

	(1) Non-attritor	(2) Attritor	(1)-(2)
Being in a migrant household	0.095	0.097	-0.003
	(0.002)	(0.006)	(0.006)
Observations	11,945	2,769	14,714

Notes: Standard errors are in parentheses. This table does not include attritors who migrate with their entire household, nor migrants. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

[25] The 24.28% figure results from 3,820 out of 15,775 individuals, leaving 11,945 successfully re-interviewed across the two waves. This figure excludes individuals lost from the survey due to death or misclassification in the second wave.

[26] Migrants account for approximately 11.40% of the total attrition. They are excluded from our analysis of attrition bias, though including them does not affect the findings.

Missing values There are 11,945 individuals who were successfully re-interviewed across the two waves and included in the sample before the final data processing. After the data processing described in Section 3.3.2.2, the final sample is composed of 7,359 individuals. To go from 11,945 to 7,359 individuals, some observations were excluded. Many of the 11,945 re-interviewed individuals are not included for reasons specific to the analysis itself, which do not introduce bias but simply alter what the sample represents; for example, children under two, or pregnant women were excluded. However, I also removed individuals with missing data for some variables in either Wave 2 or Wave 3. Indeed, I wanted to have a balanced panel. Individuals were excluded due to missing values for two reasons: (1) missing values for outcomes and (2) missing values for control variables. In total, 2,357 (935 + 1,422) individuals are affected by these missing values. Importantly, individuals with missing values as early as Wave 2 cannot create bias since the treatment does not occur until between waves 2 and 3. Therefore, I focus on the issue of individuals who have missing values in Wave 3 and remain in the sample; what I need to justify is that excluding these observations does not introduce bias. Table 3.D.1.2 shows that individuals with missing values in Wave 3 do not differ in terms of outcomes from those without missing values (i.e., those included in the final sample).[27] Furthermore, while anthropometric measures do not show significant differences, it is essential to consider whether these missing values might be systematically linked to migration (being left behind). This is examined in Table 3.D.1.3, which demonstrates that individuals with missing data are neither more nor less likely to be left behind, whether they are adults or children. In summary, regarding missing data, individuals excluded due to missing values do not differ from those retained, and no bias related to the treatment is introduced. Additionally, the mean values of all variables were compared between the balanced and unbalanced samples and were found to be similar.

Table 3.D.1.2: Comparison of mean values of anthropometric indicators in Wave 2 for individuals with missing and non-missing data in Wave 3

	(1) Non-missing data in Wave 3	(2) Missing data in Wave 3	(1)-(2)
Body weight in Wave 2	62.70	63.24	-0.54
	(0.19)	(0.45)	(0.50)
Observations	4,794	812	5,606
Z-score BMI-for-age in Wave 2	-0.087	-0.181	0.094
	(0.034)	(0.079)	(0.088)
Observations	2,826	484	3,310

Notes: Standard errors are in parentheses. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table 3.D.1.3: Comparison of mean values of left-behind status for individuals with missing and non-missing data in Wave 3

	(1) Non-missing data in Wave 3	(2) Missing data in Wave 3	(1)-(2)
Left-behind (adults)	0.095	0.086	0.009
	(0.004)	(0.010)	(0.011)
Observations	5,278	812	6,090
Left-behind (children)	0.105	0.085	0.020
	(0.005)	(0.013)	(0.015)
Observations	3,140	484	3,624

Notes: Standard errors are in parentheses. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

[27] If all cases of missing values (Wave 2 alone, Wave 3 alone, or both) were considered, and we compared the samples before and after excluding the missing data, differences in the means of the anthropometric variables would still not be significant. These results are available upon request.

3.D.2 Attrition probits

Table 3.D.2.1: Probit for adults

	(1) Attrition probit
Weight	-0.001
	(0.002)
Age	-0.019***
	(0.003)
Male	-0.154**
	(0.073)
Spouse	0.978***
	(0.110)
Child	1.751***
	(0.125)
Grandchild	1.532***
	(0.225)
Other	2.263***
	(0.129)
Primary education	0.125
	(0.089)
Post-primary education	0.085
	(0.085)
Secondary education	0.079
	(0.105)
Tertiary education	0.506***
	(0.149)
Married	-0.027
	(0.081)
Age of the head	0.005**
	(0.002)
Male head	0.199**
	(0.078)
Head, primary education	-0.034
	(0.090)
Head, post-primary education	0.000
	(0.083)
Head, secondary education	-0.003
	(0.126)
Head, tertiary education	-0.027
	(0.132)
Household size	0.021
	(0.025)
Number of working-age members	0.027
	(0.042)
Dependency ratio	-0.026
	(0.054)
Rural	0.129*
	(0.069)
Wealth index, 2nd quintile	0.059
	(0.089)
Wealth index, 3rd quintile	0.104
	(0.084)
Wealth index, 4th quintile	0.126
	(0.090)
Wealth index, 5th quintile	-0.163
	(0.115)
Constant	-2.546***
	(0.271)
Region dummies	Yes
Log-likelihood	-1482.386
Pseudo R^2	0.344
Observations	6,430

Notes: Standard errors are in parentheses. The reference categories of the explanatory variables are: Head (relationship to the head); None or preschool (education level); Head, none or preschool and Wealth index, 1st quintile. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table 3.D.2.2: Probit for children

	(1) Attrition probit
Zbmi	0.020
	(0.012)
Age	0.068***
	(0.007)
Male	-0.244***
	(0.039)
Spouse	-1.027**
	(0.519)
Child	-0.650***
	(0.076)
Grandchild	-0.144
	(0.102)
Primary education	-0.161***
	(0.056)
Post-primary education	-0.099
	(0.090)
Secondary education	-0.106
	(0.159)
Tertiary education	-0.324
	(0.627)
Married	-0.064
	(0.563)
Age of the head	0.000
	(0.002)
Male head	0.208***
	(0.053)
Head, primary education	0.080
	(0.063)
Head, post-primary education	0.054
	(0.059)
Head, secondary education	0.181**
	(0.092)
Head, tertiary education	0.193**
	(0.097)
Household size	-0.021
	(0.018)
Number of working-age members	0.061*
	(0.034)
Dependency ratio	0.061*
	(0.034)
Rural	0.190***
	(0.053)
Wealth index, 2nd quintile	0.102*
	(0.061)
Wealth index, 3rd quintile	-0.029
	(0.062)
Wealth index, 4th quintile	-0.135**
	(0.068)
Wealth index, 5th quintile	-0.224***
	(0.084)
Constant	-1.660***
	(0.197)
Region dummies	Yes
Log-likelihood	-2792.462
Pseudo R^2	0.080
Observations	5,289

Notes: Standard errors are in parentheses. The reference categories of the explanatory variables are: Head (relationship to the head); None or preschool (education level); Head, none or preschool and Wealth index, 1st quintile. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

3.D.3 Results with inverse probability weights

From the results of the attrition probit in Tables 3.D.2.1 and 3.D.2.2, I cannot confirm that attrition is random since not all of the variables are non-significant. Therefore, I implement a procedure to address the attrition bias: inverse probability weighting (Baulch and Quisumbing, 2011; Fitzgerald et al., 1998). Following the procedure described in Baulch and Quisumbing (2011), I first define a variable that determines who remains in the sample. This variable is the inverse of attrition, i.e., it takes the value 1 if the individual remains in the sample between waves 2 and 3 and zero if the individual drops out of the sample. Second, using a probit, I regress this participation variable on the variables used in Tables 3.D.2.1 and 3.D.2.2. This first model is the unrestricted model. From this model, predicted probabilities are generated, which predict the probability of remaining in the sample.

Then, based on the results from the unrestricted model, I estimate another probit (the restricted model) explaining participation, in which I only include the variables that do not affect attrition, i.e., the non-significant variables from the unrestricted model. It is equivalent to excluding variables that have a significant impact on attrition. Predicted probabilities are also derived from the restricted model. The ratio of the predicted values of the restricted model to the predicted values of the unrestricted model gives the inverse probability weights. These weights are then incorporated into the kernel-based PSM-DID model. The rationale is to give more weight to individuals with initial characteristics similar to those who subsequently attrit than to individuals whose characteristics make them more likely to remain in the sample (Baulch and Quisumbing, 2011). The findings are similar to the main results.

Table 3.D.3.1: Results from PSM-DID for adults, with inverse probability weights

	Weight			
	(1)	(2)	(3)	(4)
Panel A - All adults				
Left behind*Post	-1.066*	-1.020*	-1.048**	-1.079**
	(0.560)	(0.530)	(0.526)	(0.529)
Mean weight at baseline	62.765	62.765	62.765	62.765
R^2	0.001	0.121	0.135	0.135
Observations	8,712	8,712	8,712	8,712
Panel B - Males				
Left behind*Post	-1.899**	-1.848**	-1.850**	-1.858**
	(0.781)	(0.754)	(0.742)	(0.744)
Mean weight at baseline	63.633	63.633	63.633	63.633
R^2	0.004	0.094	0.125	0.125
Observations	3,746	3,746	3,746	3,746
Panel C - Females				
Left behind*Post	-0.346	-0.046	-0.072	-0.047
	(0.817)	(0.762)	(0.753)	(0.758)
Mean weight at baseline	62.036	62.036	62.036	62.036
R^2	0.004	0.152	0.175	0.175
Observations	4,574	4,574	4,574	4,574
Panel D - Underweight adults				
Left behind*Post	1.000	0.683	0.662	1.167
	(1.210)	(1.098)	(1.095)	(1.116)
Mean weight at baseline	46.819	46.819	46.819	46.819
R^2	0.137	0.335	0.352	0.359
Observations	534	534	534	534
Panel E - Healthy adults				
Left behind*Post	-1.169**	-1.127**	-1.145**	-1.204***
	(0.502)	(0.462)	(0.459)	(0.461)
Mean weight at baseline	58.998	58.998	58.998	58.998
R^2	0.009	0.179	0.189	0.190
Observations	4,696	4,696	4,696	4,696
Panel F - Overweight adults				
Left behind*Post	-1.167	-0.740	-0.765	-0.814
	(1.249)	(1.206)	(1.196)	(1.198)
Mean weight at baseline	75.279	75.279	75.279	75.279
R^2	0.031	0.126	0.146	0.146
Observations	2,486	2,486	2,486	2,486
Individual variables	No	Yes	Yes	Yes
Household head variables	No	Yes	Yes	Yes
Household variables	No	Yes	Yes	Yes
Rural dummy	No	No	Yes	Yes
Region dummies	No	No	Yes	Yes
Number of entries and exits	No	No	No	Yes

Notes: Standard errors are in parentheses. The variables listed at the bottom of the table are included in the DID, i.e., after matching. Individual variables include age, gender, and relationship to the head. Household head variables include age, gender, and education level of the head. Household variables include household size, number of working-age members, dependency ratio, and wealth index quintiles. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table 3.D.3.2: Results from PSM-DID for children, with inverse probability weights

	Zbmi			
	(1)	(2)	(3)	(4)
Panel A - All children				
Left behind*Post	-0.201**	-0.249***	-0.236**	-0.268***
	(0.095)	(0.094)	(0.093)	(0.094)
Mean zbmi at baseline	-0.096	-0.096	-0.096	-0.096
R^2	0.002	0.053	0.074	0.075
Observations	5,056	5,056	5,056	5,056
Panel B - Males				
Left behind*Post	-0.215	-0.296**	-0.273**	-0.301**
	(0.134)	(0.131)	(0.130)	(0.131)
Mean zbmi at baseline	-0.123	-0.123	-0.123	-0.123
R^2	0.001	0.077	0.092	0.094
Observations	2,916	2,916	2,916	2,916
Panel C - Females				
Left behind*Post	-0.331**	-0.359***	-0.347**	-0.399***
	(0.139)	(0.139)	(0.136)	(0.138)
Mean zbmi at baseline	-0.063	-0.063	-0.063	-0.063
R^2	0.005	0.038	0.081	0.083
Observations	2,080	2,080	2,080	2,080
Panel D - Underweight children				
Left behind*Post	-0.250	-0.551**	-0.512**	-0.431*
	(0.257)	(0.242)	(0.240)	(0.239)
Mean zbmi at baseline	-2.962	-2.962	-2.962	-2.962
R^2	0.468	0.572	0.593	0.607
Observations	456	456	456	456
Panel E - Healthy children				
Left behind*Post	-0.110	-0.135	-0.126	-0.160*
	(0.083)	(0.084)	(0.084)	(0.084)
Mean zbmi at baseline	-0.406	-0.406	-0.406	-0.406
R^2	0.006	0.034	0.044	0.049
Observations	3,240	3,240	3,240	3,240
Panel F - Overweight children				
Left behind*Post	-0.453***	-0.490***	-0.479***	-0.624***
	(0.168)	(0.164)	(0.160)	(0.161)
Mean zbmi at baseline	2.285	2.285	2.285	2.285
R^2	0.399	0.446	0.478	0.489
Observations	1,166	1,166	1,166	1,166
Individual variables	No	Yes	Yes	Yes
Household head variables	No	Yes	Yes	Yes
Household variables	No	Yes	Yes	Yes
Rural dummy	No	No	Yes	Yes
Region dummies	No	No	Yes	Yes
Number of entries and exits	No	No	No	Yes

Notes: Standard errors are in parentheses. The variables listed at the bottom of the table are included in the DID, i.e., after matching. Individual variables include age, gender, and relationship to the head. Household head variables include age, gender, and education level of the head. Household variables include household size, number of working-age members, dependency ratio, and wealth index quintiles. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

3.D.4 Intra-household selection bias

The coefficients of interest pertaining to the migrant variables for adults are generally not significant (Table 3.D.4.1). Notable exception includes the female subgroup in regression including household fixed effects, as indicated in column (7) of Table 3.D.4.2. However when household fixed effects are combined with additional control variables, the coefficient is not significant anymore (column (8) of Table 3.D.4.2). Similarly, for the children's sample, coefficients are almost all not significant. Nevertheless, a marginal significance at the 10% level is observed in column (4) of Table 3.D.4.4 for the boys' subgroup. This level of significance, while notable, is considered negligible. In the girls' subgroup of the same table, significance emerges in column (5) for the model without further control variables or household fixed effects. However, this significance dissipates with the inclusion of additional variables and household fixed effects, as demonstrated in columns (6) to (8) of Table 3.D.4.4. It suggests that migrants and left-behind individuals are not different regarding anthropometric indicators.

Table 3.D.4.1: OLS results for adults testing for intra-household selection bias

	Weight			
	All adults			
	(1)	(2)	(3)	(4)
Migrant	-0.362	-0.092	-1.046	-0.676
	(1.183)	(1.556)	(1.145)	(1.617)
Individual variables	No	Yes	No	Yes
Household head variables	No	Yes	No	Yes
Household variables	No	Yes	No	Yes
Rural dummy	No	No	No	Yes
Region dummies	No	No	No	Yes
Household fixed effects	No	No	Yes	Yes
R^2	0.000	0.144	0.570	0.607
Observations	663	662	663	662

Notes: Standard errors are in parentheses. Individual variables include age, gender, and relationship to the head. Household head variables include age, gender, and education level of the head. Household variables include household size, number of working-age members, dependency ratio, and wealth index quintiles. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table 3.D.4.2: OLS results for males and females testing for intra-household selection bias

	Weight							
	Males				Females			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Migrant	-1.511	-0.000	-1.996	-0.194	-0.693	0.047	-4.234**	0.826
	(1.615)	(2.210)	(1.967)	(2.799)	(1.823)	(2.308)	(1.903)	(2.719)
Individual variables	No	Yes	No	Yes	No	Yes	No	Yes
Household head variables	No	Yes	No	Yes	No	Yes	No	Yes
Household variables	No	Yes	No	Yes	No	Yes	No	Yes
Rural dummy	No	No	No	Yes	No	No	No	Yes
Region dummies	No	No	No	Yes	No	No	No	Yes
Household fixed effects	No	No	Yes	Yes	No	No	Yes	Yes
R^2	0.003	0.113	0.796	0.836	0.000	0.238	0.813	0.842
Observations	304	303	304	303	359	359	359	359

Notes: Standard errors are in parentheses. Individual variables include age, gender, and relationship to the head. Household head variables include age, gender, and education level of the head. Household variables include household size, number of working-age members, dependency ratio, and wealth index quintiles. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table 3.D.4.3: OLS results for children testing for intra-household selection bias

	All children			
	(1)	(2)	(3)	(4)
Migrant	0.242	0.282	0.177	0.274
	(0.149)	(0.172)	(0.190)	(0.235)
Individual variables	No	Yes	No	Yes
Household head variables	No	Yes	No	Yes
Household variables	No	Yes	No	Yes
Rural dummy	No	No	No	Yes
Region dummies	No	No	No	Yes
Household fixed effects	No	No	Yes	Yes
R^2	0.010	0.156	0.735	0.750
Observations	274	272	274	272

Notes: Standard errors are in parentheses. Individual variables include age, gender, and relationship to the head. Household head variables include age, gender, and education level of the head. Household variables include household size, number of working-age members, dependency ratio, and wealth index quintiles. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table 3.D.4.4: OLS results for boys and girls testing for intra-household selection bias

	Zbmi							
	Males				Females			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Migrant	0.070	0.116	0.425	0.700*	0.451**	0.412	-0.261	-0.795
	(0.195)	(0.243)	(0.284)	(0.413)	(0.222)	(0.249)	(0.571)	(0.872)
Individual variables	No	Yes	No	Yes	No	Yes	No	Yes
Household head variables	No	Yes	No	Yes	No	Yes	No	Yes
Household variables	No	Yes	No	Yes	No	Yes	No	Yes
Rural dummy	No	No	No	Yes	No	No	No	Yes
Region dummies	No	No	No	Yes	No	No	No	Yes
Household fixed effects	No	No	Yes	Yes	No	No	Yes	Yes
R^2	0.001	0.183	0.839	0.860	0.037	0.308	0.935	0.970
Observations	164	163	164	163	110	109	110	109

Notes: Standard errors are in parentheses. Individual variables include age, gender, and relationship to the head. Household head variables include age, gender, and education level of the head. Household variables include household size, number of working-age members, dependency ratio, and wealth index quintiles. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

3.E Parallel trend assumption

I test for the parallel trend assumption on a sub-sample of the permanent survey members who were successfully interviewed in all three waves. This subset excludes pregnant women, those transitioning from childhood to adulthood, and individuals with missing or implausible anthropometric data. Given these constraints, we have anthropometric indicator values in all three waves for approximately 69% of the sample. Figures 3.E.1 and 3.E.2 display trends of anthropometric indicators for the sample of matched individuals.[28]

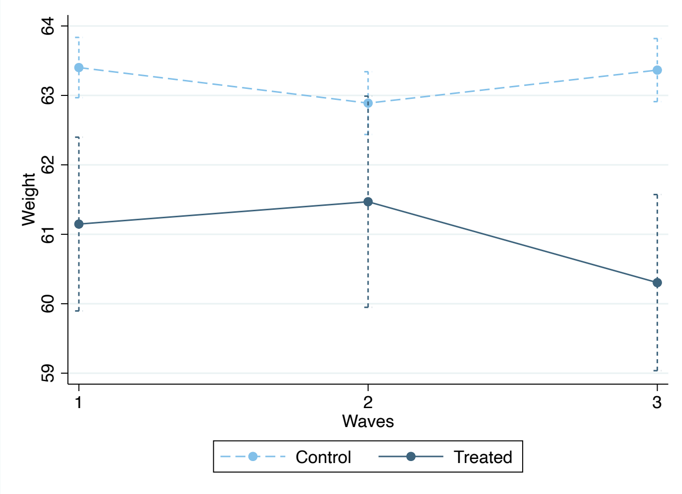


Figure 3.E.1: Trends in adults' body weight over the waves

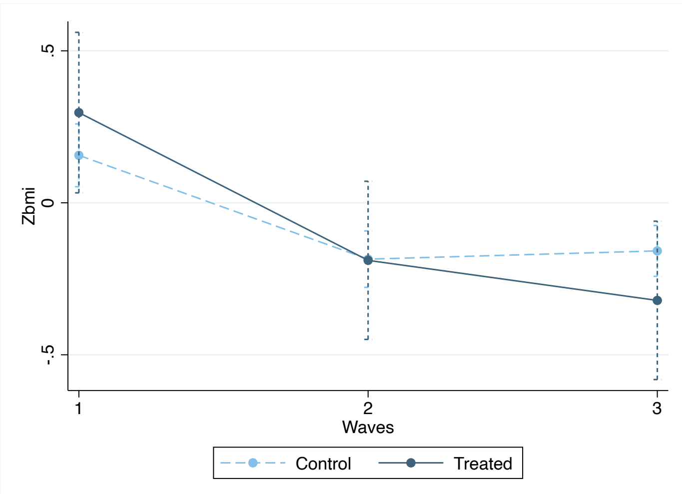


Figure 3.E.2: Trends in children's BMI-for-age z-score over the waves

[28] The graphs are almost identical using the sample not used for matching.

3.F Matching process quality

Table 3.F.1 displays the results of the model that estimates the probability of being left behind for adults (used to create the propensity scores). The results show that the head's age and education, the number of working-age members, and the dependency ratio are determinants for an adult to be left behind. Table 3.F.2 shows the results for children. Like adults, the head's age is a determinant of migration. However, its education is no longer significant. The household size, number of working-age members, and dependency ratio also significantly impact the probability of being left behind. Finally, wealthier households are less likely to be migrant households.

To ensure the quality of the matching process, it is essential to determine whether the treated and control units share the same support. Figures 3.F.1 and 3.F.2 display the kernel density functions of the treated and control groups based on before and post-matching. For adults and children, the kernel density functions of the two groups are different before matching. However, after matching, the right sides of these figures indicate that the kernel density functions are much more similar. The characteristics of the variables are roughly equivalent between the two groups after matching. Although the matching is supposed to overcome the selection bias, it is also necessary to check if, after matching, the variables have the same distribution between the individuals left behind and those not left behind. The results from the balance test after matching are displayed in Tables 3.F.3 and 3.F.4. According to these tables, almost all mean differences between the treatment and the control groups of the variables used for matching are equal to 0 and not statistically significant. Only two mean differences are significant for the adults (household size and working-age members), but only at the 5% level. Whereas for children, there are no differences between control and treated means. As explained earlier, to satisfy the common support hypothesis, I restrict the analysis to the common support, excluding the individuals' outsides of it.[29]

[29]The results are robust to non-restriction to the common support. These are available upon request.

Table 3.F.1: Results of the probit model to estimate the probability of being left behind for adults

	(1)
	Treated
Age	0.011***
	(0.004)
Male	-0.096
	(0.113)
Head	-0.014
	(0.186)
Spouse	0.051
	(0.192)
Child	0.253
	(0.248)
Grandchild	0.332
	(0.404)
Age of the head	0.007*
	(0.004)
Male head	-0.220*
	(0.116)
Head, none or preschool	-0.117
	(0.143)
Head, primary education	-0.371**
	(0.156)
Head, post-primary education	-0.084
	(0.136)
Head, secondary education	-0.367**
	(0.183)
Household size	0.062***
	(0.022)
Number of working-age members	0.210***
	(0.039)
Dependency ratio	-0.101**
	(0.041)
Rural	0.117
	(0.077)
Wealth index, 1st quintile	0.025
	(0.120)
Wealth index, 2nd quintile	-0.052
	(0.118)
Wealth index, 3rd quintile	-0.116
	(0.111)
Wealth index, 4th quintile	-0.284***
	(0.108)
Constant	-3.094***
	(0.319)
Region dummies	Yes
Log-likelihood	-1246.559
Pseudo R^2	0.177
Observations	4,579

Notes: Standard errors are in parentheses. The reference categories of the explanatory variables are: Other (relationship to the head); Head, tertiary education and Wealth index, 5th quintile.
*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table 3.F.2: Results of the probit model to estimate the probability of being left behind for children

	(1)
	Treated
Age	-0.007
	(0.010)
Male	-0.011
	(0.072)
Spouse	0.563
	(0.707)
Child	0.009
	(0.208)
Grandchild	0.208
	(0.240)
Age of the head	0.016***
	(0.003)
Male head	-0.256**
	(0.111)
Head, none or preschool	-0.012
	(0.203)
Head, primary education	-0.152
	(0.215)
Head, post-primary education	0.073
	(0.199)
Head, secondary education	-0.499*
	(0.297)
Household size	0.067**
	(0.031)
Number of working-age members	0.200***
	(0.064)
Dependency ratio	-0.157**
	(0.077)
Rural	0.598***
	(0.119)
Wealth index, 1st quintile	-0.091
	(0.164)
Wealth index, 2nd quintile	-0.172
	(0.166)
Wealth index, 3rd quintile	-0.514***
	(0.167)
Wealth index, 4th quintile	-0.339**
	(0.153)
Constant	-3.550***
	(0.449)
Region dummies	Yes
Log-likelihood	-787.849
Pseudo R^2	0.196
Observations	2,780

Notes: Standard errors are in parentheses. The reference categories of the explanatory variables are: Other (relationship to the head); Head, tertiary education and Wealth index, 5th quintile.
*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table 3.F.3: Balance test of the matched adults sample using kernel matching

Variable	Mean control	Mean treated	Difference
Age	48.191	48.425	0.234
Male	0.394	0.392	-0.002
Head	0.463	0.466	0.003
Spouse	0.353	0.360	0.007
Child	0.142	0.129	-0.013
Grandchild	0.012	0.011	-0.001
Other	0.031	0.034	0.003
Head age	56.836	56.772	-0.065
Male head	0.764	0.769	0.005
Head, none or preschool	0.520	0.509	-0.012
Head, primary education	0.096	0.103	0.007
Head, post-primary education	0.300	0.297	-0.003
Head, secondary education	0.031	0.028	-0.003
Head, tertiary education	0.052	0.063	0.010
Household size	6.272	6.459	0.187**
Working-age members	3.681	3.815	0.134**
Dependency ratio	0.833	0.833	-0.000
Rural	0.755	0.754	-0.001
Wealth index, 1st quintile	0.380	0.369	-0.012
Wealth index, 2nd quintile	0.169	0.179	0.010
Wealth index, 3rd quintile	0.180	0.177	-0.003
Wealth index, 4th quintile	0.144	0.144	0.000
Wealth index, 5th quintile	0.128	0.131	0.004
Western Region	0.057	0.050	-0.007
Central Region	0.054	0.052	-0.002
Greater Accra Region	0.020	0.015	-0.005
Volta Region	0.100	0.099	-0.001
Eastern Region	0.082	0.086	0.005
Ashanti Region	0.126	0.125	-0.001
Brong-Ahafo Region	0.104	0.116	0.013
Northern Region	0.307	0.304	-0.003
Upper East Region	0.128	0.134	0.006
Upper West Region	0.023	0.019	-0.004

Notes: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table 3.F.4: Balance test of the matched children sample using kernel matching

Variable	Mean control	Mean treated	Difference
Age	8.869	8.882	0.013
Male	0.572	0.561	-0.012
Spouse	0.003	0.003	0.001
Child	0.836	0.850	0.014
Grandchild	0.132	0.118	-0.014
Other	0.029	0.029	-0.001
Head age	52.416	52.459	0.043
Male head	0.815	0.822	0.007
Head, none or preschool	0.605	0.580	-0.025
Head, primary education	0.112	0.124	0.013
Head, post-primary education	0.233	0.239	0.006
Head, secondary education	0.012	0.013	0.001
Head, tertiary education	0.039	0.045	0.005
Household size	7.944	7.997	0.053
Working-age members	3.889	3.987	0.098
Dependency ratio	1.201	1.172	-0.029
Rural	0.895	0.885	-0.010
Wealth index, 1st quintile	0.495	0.490	-0.004
Wealth index, 2nd quintile	0.165	0.178	0.013
Wealth index, 3rd quintile	0.165	0.178	0.013
Wealth index, 4th quintile	0.136	0.137	0.001
Wealth index, 5th quintile	0.086	0.092	0.006
Western Region	0.036	0.032	-0.004
Central Region	0.053	0.051	-0.002
Greater Accra Region	0.011	0.013	0.001
Volta Region	0.089	0.089	0.000
Eastern Region	0.043	0.041	-0.001
Ashanti Region	0.088	0.089	0.001
Brong-Ahafo Region	0.111	0.131	0.019
Northern Region	0.425	0.395	-0.030
Upper East Region	0.131	0.146	0.016
Upper West Region	0.014	0.013	-0.001

Notes: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

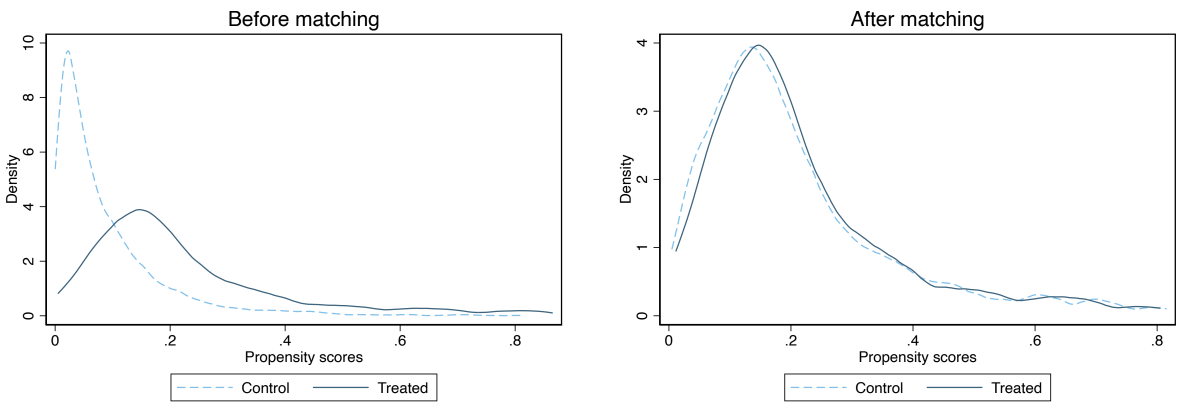


Figure 3.F.1: Kernel density of the treated and control groups for adults

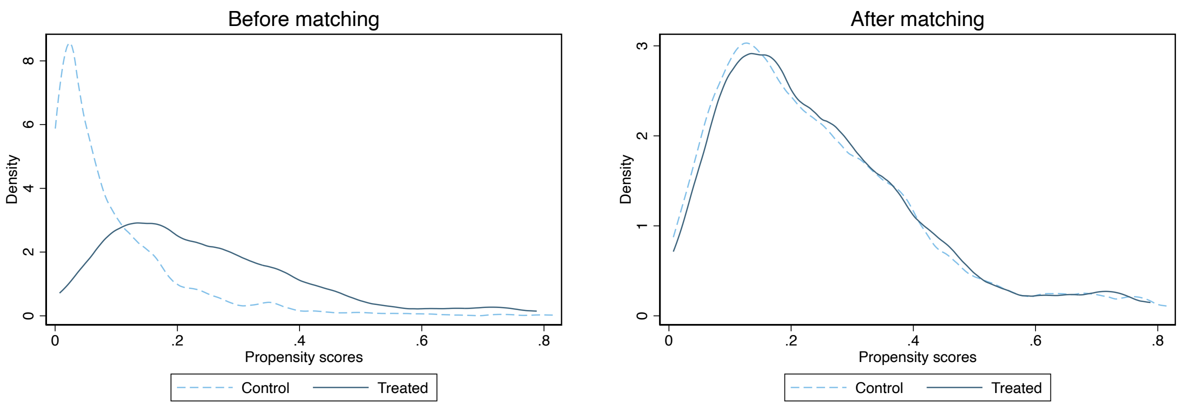


Figure 3.F.2: Kernel density of the treated and control groups for children

3.G Transition matrices of individuals' nutritional statuses

Table 3.G.1: Nutritional status transition for left-behind and non-left-behind adults

Nutritional status of adults in the first wave (Wave 2)	Nutritional status in the following wave (Wave 3)			
	Underweight	Healthy	Overweight/Obese	Total
Non Left Behind adults				
Underweight	33.25	57.93	8.82	100.00
Healthy	8.19	64.80	27.02	100.00
Overweight/Obese	2.99	28.76	68.24	100.00
Left Behind adults				
Underweight	28.57	67.35	4.08	100.00
Healthy	12.46	70.37	17.17	100.00
Overweight/Obese	4.03	43.55	52.42	100.00

Notes: All values are in percentages.

Table 3.G.2: Nutritional status transition for left-behind and non-left-behind children

Nutritional status of children in the first wave (Wave 2)	Nutritional status in the following wave (Wave 3)			
	Underweight	Healthy	Overweight/Obese	Total
Non Left Behind children				
Underweight	11.08	72.47	16.46	100.00
Healthy	8.79	72.53	18.68	100.00
Overweight/Obese	6.42	61.15	32.43	100.00
Left Behind children				
Underweight	23.81	59.52	16.67	100.00
Healthy	7.84	74.51	17.65	100.00
Overweight/Obese	16.18	64.71	19.12	100.00

Notes: All values are in percentages.

3.H Confounders of changes in household composition

Usually, the literature only partially addresses the issues of changes in household composition following a migration (Bertoli and Murard, 2020). Most of the time, the variations in co-residence choices potentially generated by migration are ignored. These changes in living arrangements may have implications for the nutrition of the left-behind individuals and may be confounded with the effects of migration. Therefore, I discuss how changes in living arrangements (entries and exits) may have implications for the empirical strategy.

Regarding the entries, in migrant households, 68.26% of the newcomers are children under 19. Moreover, almost 50% of these children entered because they were born. Since births occur in any household and are not specific to migrant households, they can be considered random events, unlikely to introduce bias. For these reasons, it is unlikely that there are any confounding effects of individuals entering the household. Despite these reasons, I included a variable specifying the number of entries into a household as a control variable. In particular, to control for the entry of individuals because they moved to live with relatives, which accounts for 36.59% of those who arrive in migrant households.

Concerning the exits, the main reasons for individuals leaving migrant households are to live with relatives (24.78%), to move for school (19.78%), or because the individual has died (16.09%). In this case, it is difficult to show that the individuals who leave are unrelated to previous migration. Indeed, when individuals move to live with relatives, these individuals may be joining the migrants. In addition, when individuals move for school, I cannot rule out that these people may send remittances and thus indirectly influence the nutrition of the left behind. Therefore, it is challenging to argue that further exits do not affect nutrition either. To address this potential bias, I also included the number of exits in households as a control (excluding migrants for work for more than six months).[30]

[30] Unfortunately, it is also difficult to ascertain that migration does not lead to further exits. Indeed, even if I can determine that there are entries and exits of individuals in households between waves 2 and 3, I cannot determine precisely when individuals left the household between waves 2 and 3.

3.I Results with nutritional statuses as outcomes

Table 3.I.1: Results from kernel PSM-DID for adults, nutritional status as outcome

	Underweight	Healthy	Overweight
	(1)	(2)	(3)
Left behind*Post	0.030**	0.026	-0.055***
	(0.013)	(0.021)	(0.018)
Mean at baseline	0.097	0.598	0.304
R^2	0.045	0.057	0.117
Observations	8,754	8,754	8,754
Individual variables	Yes	Yes	Yes
Household head variables	Yes	Yes	Yes
Household variables	Yes	Yes	Yes
Rural dummy	Yes	Yes	Yes
Region dummies	Yes	Yes	Yes
Number of entries and exits	Yes	Yes	Yes

Notes: Standard errors are in parentheses. The variables listed at the bottom of the table are included in the DID, i.e., after matching. Individual variables include age, gender and relationship to the head. Household head variables include age, gender and education level of the head. Household variables include household size, number of working age members, dependency ratio and wealth index quintiles. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table 3.I.2: Results from kernel PSM-DID for children, nutritional status as outcome

	Underweight	Healthy	Overweight
	(1)	(2)	(3)
Left behind*Post	0.044**	-0.047*	0.003
	(0.018)	(0.026)	(0.022)
Mean at baseline	0.129	0.634	0.237
R^2	0.036	0.066	0.085
Observations	5,054	5,054	5,054
Individual variables	Yes	Yes	Yes
Household head variables	Yes	Yes	Yes
Household variables	Yes	Yes	Yes
Rural dummy	Yes	Yes	Yes
Region dummies	Yes	Yes	Yes
Number of entries and exits	Yes	Yes	Yes

Notes: Standard errors are in parentheses. The variables listed at the bottom of the table are included in the DID, i.e., after matching. Individual variables include age, gender and relationship to the head. Household head variables include age, gender and education level of the head. Household variables include household size, number of working age members, dependency ratio and wealth index quintiles. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

3.J Robustness checks and heterogeneity

3.J.1 Heterogeneity among children's age groups

Table 3.J.1: Results from PSM-DID for children, heterogeneity by age and gender

	Zbmi			
	(1)	(2)	(3)	(4)
Panel A - Children under 10 years				
Left behind*Post	-0.560***	-0.634***	-0.632**	-0.745***
	(0.200)	(0.200)	(0.197)	(0.199)
Mean zbmi at baseline	0.262	0.262	0.262	0.262
R^2	0.017	0.049	0.090	0.097
Observations	1,446	1,446	1,446	1,446
Panel B - Males under 10 years				
Left behind*Post	-0.452	-0.484*	-0.474*	-0.571**
	(0.289)	(0.291)	(0.286)	(0.289)
Mean zbmi at baseline	0.276	0.276	0.276	0.276
R^2	0.017	0.060	0.112	0.118
Observations	728	728	728	728
Panel C - Females under 10 years				
Left behind*Post	-0.803***	-0.948***	-0.954***	-1.027***
	(0.303)	(0.306)	(0.294)	(0.301)
Mean zbmi at baseline	0.246	0.246	0.246	0.246
R^2	0.026	0.079	0.164	0.168
Observations	590	590	590	590
Panel D - Children over 10 years				
Left behind*Post	-0.017	-0.052	-0.042	-0.050
	(0.104)	(0.104)	(0.103)	(0.104)
Mean zbmi at baseline	-0.281	-0.281	-0.281	-0.281
R^2	0.000	0.034	0.057	0.058
Observations	3,390	3,390	3,390	3,390
Panel E - Males over 10 years				
Left behind*Post	-0.041	-0.149	-0.129	-0.144
	(0.152)	(0.151)	(0.150)	(0.150)
Mean zbmi at baseline	-0.312	-0.312	-0.312	-0.312
R^2	0.001	0.062	0.082	0.085
Observations	1,806	1,806	1,806	1,806
Panel F - Females over 10 years				
Left behind*Post	-0.137	-0.157	-0.152	-0.150
	(0.168)	(0.167)	(0.163)	(0.165)
Mean zbmi at baseline	-0.239	-0.239	-0.239	-0.239
R^2	0.045	0.045	0.106	0.108
Observations	1,140	1,140	1,140	1,140
Individual variables	No	Yes	Yes	Yes
Household head variables	No	Yes	Yes	Yes
Household variables	No	Yes	Yes	Yes
Rural dummy	No	No	Yes	Yes
Region dummies	No	No	Yes	Yes
Number of entries and exits	No	No	No	Yes

Notes: Standard errors are in parentheses. Individual variables include age, gender, and relationship to the head. Household head variables include age, gender, and education level of the head. Household variables include household size, number of working-age members, dependency ratio, and wealth index quintiles. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

3.J.2 Robustness to potential unobserved variables bias

Table 3.J.2: Oster test

Dependent variable	(1)	(2)	(3)	(4)
	Baseline effect β (Std. Error) $[R^0]$	Controlled effect β (Std. Error) $[\tilde{R}]$	Bias adjusted β^* ($R_{max} = 1.3\tilde{R}$, $\delta = 1$)	Coefficient bounds
Adults				
All adults	**-1.132** (0.559) [0.001]**	**-1.050** (0.528) [0.136]**	**-1.098**	**[-0.844,-1.050]**
Males	**-1.888** (0.773) [0.004]**	**-1.842** (0.735) [0.128]**	**-2.006**	**[-1.042,-1.842]**
Females	-0.362 (0.819) [0.004]	-0.085 (0.761) [0.174]	0.154	[-0.941,-0.085]
Underweight adults	0.708 (1.157) [0.143]	0.967 (1.072) [0.351]	0.022	[4.477,0.967]
Healthy adults	**-1.154** (0.498) [0.009]**	**-1.210*** (0.458) [0.189]**	**-1.465**	**[-0.287,-1.210]**
Overweight adults	-1.653 (1.227) [0.032]	-1.437 (1.170) [0.156]	-0.702	[-4.642, -1.437]
Children				
All children	**-0.193** (0.095) [0.002]**	**-0.259*** (0.094) [0.075]**	**-0.298**	**[-0.121, -0.259]**
Males	**-0.198 (0.134) [0.001]**	**-0.294** (0.131) [0.093]**	**-0.339**	**[-0.131, -0.294]**
Females	**-0.337** (0.139) [0.005]**	**-0.402*** (0.138) [0.083]**	**-0.454**	**[-0.231, -0.402]**
Underweight children	-0.255 (0.274) [0.431]	-0.348 (0.242) [0.620]	-1.149	[1.425, -0.348]
Healthy children	-0.104 (0.083) [0.005]	-0.155* (0.084) [0.048]	-0.228	[0.111, -0.155]
Overweight children	**-0.514*** (0.167) [0.398]**	**-0.679*** (0.160) [0.486]**	**-0.298**	**[-1.791, -0.679]**

Notes: Adjustment for unobserved variables based on Oster (2019). Following Oster (2019), we use the model β^* = (R_{max} = 1.3$\tilde{R}$, δ = 1) to obtain consistent estimates of the true coefficients. Standard errors and R squared are reported in parentheses and square brackets, respectively. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

3.J.3 Alternative anthropometric indicator for children

Table 3.J.3: Results from PSM-DID for children, alternative anthropometric indicator (zhfa) as outcome

	Zhfa			
	(1)	(2)	(3)	(4)
Panel A - All children				
Left behind*Post	-0.123	-0.192**	-0.188**	-0.166**
	(0.092)	(0.090)	(0.089)	(0.090)
Mean zhfa at baseline	-1.198	-1.198	-1.198	-1.198
R^2	0.011	0.072	0.089	0.091
Observations	4,880	4,880	4,880	4,880
Panel B - Males				
Left behind*Post	-0.180	-0.274**	-0.272**	-0.267**
	(0.129)	(0.125)	(0.124)	(0.124)
Mean zhfa at baseline	-1.275	-1.275	-1.275	-1.275
R^2	0.004	0.102	0.128	0.129
Observations	2,770	2,770	2,770	2,770
Panel C - Females				
Left behind*Post	-0.151	-0.244*	-0.238*	-0.241*
	(0.131)	(0.129)	(0.128)	(0.129)
Mean zhfa at baseline	-1.101	-1.101	-1.101	-1.101
R^2	0.027	0.075	0.100	0.106
Observations	1,972	1,972	1,972	1,972
Individual variables	No	Yes	Yes	Yes
Household head variables	No	Yes	Yes	Yes
Household variables	No	Yes	Yes	Yes
Rural dummy	No	No	Yes	Yes
Region dummies	No	No	Yes	Yes
Number of entries and exits	No	No	No	Yes

Notes: Standard errors are in parentheses. Individual variables include age, gender, and relationship to the head. Household head variables include age, gender, and education level of the head. Household variables include household size, number of working-age members, dependency ratio, and wealth index quintiles. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

3.J.4 Internal migration only

Table 3.J.4.1: Results from PSM-DID for adults, internal migration only

	Weight			
	(1)	(2)	(3)	(4)
Panel A - All adults				
Left behind*Post	-1.162**	-1.054*	-1.051*	-1.061*
	(0.582)	(0.550)	(0.546)	(0.548)
Mean weight at baseline	62.781	62.781	62.781	62.781
R^2	0.001	0.122	0.136	0.136
Observations	8,254	8,254	8,254	8,254
Panel B - Males				
Left behind*Post	-2.099***	-1.996**	-2.000***	-2.008***
	(0.812)	(0.782)	(0.769)	(0.771)
Mean weight at baseline	63.625	63.625	63.625	63.625
R^2	0.005	0.097	0.130	0.131
Observations	3,504	3,504	3,504	3,504
Panel C - Females				
Left behind*Post	-0.415	-0.081	-0.071	0.008
	(0.828)	(0.772)	(0.762)	(0.766)
Mean weight at baseline	62.072	62.072	62.072	62.072
R^2	0.003	0.153	0.177	0.177
Observations	4,510	4,510	4,510	4,510
Panel D - Underweight adults				
Left behind*Post	0.315	0.345	0.309	0.738
	(1.234)	(1.115)	(1.113)	(1.158)
Mean weight at baseline	46.776	46.776	46.776	46.776
R^2	0.141	0.338	0.352	0.356
Observations	502	502	502	502
Panel E - Healthy adults				
Left behind*Post	-1.108**	-1.113**	-1.124**	-1.192**
	(0.527)	(0.485)	(0.482)	(0.484)
Mean weight at baseline	58.991	58.991	58.991	58.991
R^2	0.009	0.176	0.188	0.189
Observations	4,260	4,260	4,260	4,260
Panel F - Overweight adults				
Left behind*Post	-1.316	-1.033	-1.006	-1.072
	(1.238)	(1.191)	(1.179)	(1.182)
Mean weight at baseline	75.279	75.279	75.279	75.279
R^2	0.031	0.130	0.152	0.153
Observations	2,532	2,532	2,532	2,532
Individual variables	No	Yes	Yes	Yes
Household head variables	No	Yes	Yes	Yes
Household variables	No	Yes	Yes	Yes
Rural dummy	No	No	Yes	Yes
Region dummies	No	No	Yes	Yes
Number of entries and exits	No	No	No	Yes

Notes: Standard errors are in parentheses. The variables listed at the bottom of the table are included in the DID, i.e., after matching. Individual variables include age, gender, and relationship to the head. Household head variables include age, gender, and education level of the head. Household variables include household size, number of working-age members, dependency ratio, and wealth index quintiles. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table 3.J.4.2: Results from PSM-DID for children, internal migration only

	Zbmi			
	(1)	(2)	(3)	(4)
Panel A - All children				
Left behind*Post	-0.249**	-0.312***	-0.295***	-0.325***
	(0.098)	(0.097)	(0.096)	(0.097)
Mean zbmi at baseline	-0.094	-0.094	-0.094	-0.094
R^2	0.003	0.053	0.073	0.074
Observations	4,796	4,796	4,796	4,796
Panel B - Males				
Left behind*Post	-0.262*	-0.345***	-0.320**	-0.335**
	(0.136)	(0.133)	(0.132)	(0.133)
Mean zbmi at baseline	-0.121	-0.121	-0.121	-0.121
R^2	0.002	0.082	0.097	0.098
Observations	2,812	2,812	2,812	2,812
Panel C - Females				
Left behind*Post	-0.301**	-0.325**	-0.317**	-0.358**
	(0.143)	(0.143)	(0.140)	(0.142)
Mean zbmi at baseline	-0.059	-0.059	-0.059	-0.059
R^2	0.006	0.036	0.073	0.075
Observations	1,904	1,904	1,904	1,904
Panel D - Underweight children				
Left behind*Post	-0.330	-0.649**	-0.596**	-0.482*
	(0.319)	(0.282)	(0.275)	(0.271)
Mean zbmi at baseline	-2.962	-2.962	-2.962	-2.962
R^2	0.429	0.601	0.635	0.656
Observations	352	352	352	352
Panel E - Healthy children				
Left behind*Post	-0.130	-0.165*	-0.154*	-0.179**
	(0.085)	(0.086)	(0.086)	(0.086)
Mean zbmi at baseline	-0.407	-0.407	-0.407	-0.407
R^2	0.005	0.034	0.044	0.049
Observations	3,132	3,132	3,132	3,132
Panel F - Overweight children				
Left behind*Post	-0.462***	-0.537***	-0.534***	-0.639***
	(0.171)	(0.167)	(0.164)	(0.164)
Mean zbmi at baseline	2.283	2.283	2.283	2.283
R^2	0.392	0.444	0.473	0.483
Observations	1,130	1,130	1,130	1,130
Individual variables	No	Yes	Yes	Yes
Household head variables	No	Yes	Yes	Yes
Household variables	No	Yes	Yes	Yes
Rural dummy	No	No	Yes	Yes
Region dummies	No	No	Yes	Yes
Number of entries and exits	No	No	No	Yes

Notes: Standard errors are in parentheses. The variables listed at the bottom of the table are included in the DID, i.e., after matching. Individual variables include age, gender, and relationship to the head. Household head variables include age, gender, and education level of the head. Household variables include household size, number of working-age members, dependency ratio, and wealth index quintiles. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

3.J.5 Migration for less than six months

The treatment is defined as being in a household where at least one individual migrated for work for more than six months between waves 2 and 3. Nevertheless, some individuals are also in households with migrants who moved for work less than six months ago. In the main specification, these individuals are included in the control group. I investigate whether this inclusion biases the results. I removed from the sample individuals in households where an individual migrated for work less than six months ago. The results are reported in Tables 3.J.5.1 and 3.J.5.2 and are similar to those of Tables 3.3 and 3.4. Therefore, including individuals left behind by migrants who moved less than six months ago in the control group does not bias the results.

Table 3.J.5.1: Results from PSM-DID for adults, without migration for less than six months

	Weight			
	(1)	(2)	(3)	(4)
Panel A - All adults				
Left behind*Post	-1.095*	-1.025*	-1.073**	-1.108**
	(0.572)	(0.541)	(0.537)	(0.539)
Mean weight at baseline	62.778	62.778	62.778	62.778
R^2	0.001	0.123	0.137	0.137
Observations	8,538	8,538	8,538	8,538
Panel B - Males				
Left behind*Post	-2.093***	-1.992***	-1.971***	-1.939***
	(0.783)	(0.755)	(0.745)	(0.746)
Mean weight at baseline	63.702	63.702	63.702	63.702
R^2	0.005	0.097	0.124	0.124
Observations	3,780	3,780	3,780	3,780
Panel C - Females				
Left behind*Post	-0.409	-0.127	-0.175	-0.176
	(0.837)	(0.783)	(0.774)	(0.779)
Mean weight at baseline	62.002	62.002	62.002	62.002
R^2	0.005	0.148	0.171	0.171
Observations	4,412	4,412	4,412	4,412
Panel D - Underweight adults				
Left behind*Post	0.737	0.557	0.543	0.893
	(1.196)	(1.080)	(1.079)	(1.105)
Mean weight at baseline	46.790	46.790	46.790	46.790
R^2	0.143	0.340	0.354	0.357
Observations	536	536	536	536
Panel E - Healthy adults				
Left behind*Post	-1.240**	-1.219***	-1.244***	-1.296***
	(0.506)	(0.466)	(0.464)	(0.465)
Mean weight at baseline	58.998	58.998	58.998	58.998
R^2	0.008	0.179	0.190	0.191
Observations	4,600	4,600	4,600	4,600
Panel F - Overweight adults				
Left behind*Post	-0.957	-0.518	-0.550	-0.616
	(1.262)	(1.219)	(1.208)	(1.212)
Mean weight at baseline	75.336	75.336	75.336	75.336
R^2	0.036	0.129	0.149	0.151
Observations	2,464	2,464	2,464	2,464
Individual variables	No	Yes	Yes	Yes
Household head variables	No	Yes	Yes	Yes
Household variables	No	Yes	Yes	Yes
Rural dummy	No	No	Yes	Yes
Region dummies	No	No	Yes	Yes
Number of entries and exits	No	No	No	Yes

Notes: Standard errors are in parentheses. The variables listed at the bottom of the table are included in the DID, i.e., after matching. Individual variables include age, gender, and relationship to the head. Household head variables include age, gender, and education level of the head. Household variables include household size, number of working-age members, dependency ratio, and wealth index quintiles. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table 3.J.5.2: Results from PSM-DID for children, without migration for less than six months

	Zbmi			
	(1)	(2)	(3)	(4)
Panel A - All children				
Left behind*Post	-0.204**	-0.254***	-0.240**	-0.268***
	(0.097)	(0.096)	(0.095)	(0.096)
Mean zbmi at baseline	-0.091	-0.091	-0.091	-0.091
R^2	0.003	0.054	0.076	0.077
Observations	4,900	4,900	4,900	4,900
Panel B - Males				
Left behind*Post	-0.239*	-0.308**	-0.289**	-0.311**
	(0.135)	(0.133)	(0.132)	(0.133)
Mean zbmi at baseline	-0.115	-0.115	-0.115	-0.115
R^2	0.002	0.076	0.092	0.093
Observations	2,810	2,810	2,810	2,810
Panel C - Females				
Left behind*Post	-0.211	-0.263*	-0.244*	-0.277*
	(0.144)	(0.143)	(0.140)	(0.142)
Mean zbmi at baseline	-0.060	-0.060	-0.060	-0.060
R^2	0.006	0.050	0.095	0.096
Observations	1,996	1,996	1,996	1,996
Panel D - Underweight children				
Left behind*Post	-0.117	-0.474*	-0.464*	-0.417*
	(0.256)	(0.242)	(0.242)	(0.242)
Mean zbmi at baseline	-2.958	-2.958	-2.958	-2.958
R^2	0.452	0.563	0.578	0.590
Observations	440	440	440	440
Panel E - Healthy children				
Left behind*Post	-0.098	-0.121	-0.111	-0.141*
	(0.084)	(0.085)	(0.085)	(0.085)
Mean zbmi at baseline	-0.405	-0.405	-0.405	-0.405
R^2	0.005	0.032	0.042	0.048
Observations	3,154	3,154	3,154	3,154
Panel F - Overweight children				
Left behind*Post	-0.408**	-0.446***	-0.450***	-0.530***
	(0.171)	(0.167)	(0.164)	(0.164)
Mean zbmi at baseline	2.293	2.293	2.293	2.283
R^2	0.405	0.450	0.480	0.487
Observations	1,148	1,148	1,148	1,148
Individual variables	No	Yes	Yes	Yes
Household head variables	No	Yes	Yes	Yes
Household variables	No	Yes	Yes	Yes
Rural dummy	No	No	Yes	Yes
Region dummies	No	No	Yes	Yes
Number of entries and exits	No	No	No	Yes

Notes: Standard errors are in parentheses. The variables listed at the bottom of the table are included in the DID, i.e., after matching. Individual variables include age, gender, and relationship to the head. Household head variables include age, gender, and education level of the head. Household variables include household size, number of working-age members, dependency ratio, and wealth index quintiles. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

3.J.6 Disentangling the potential bias from previous migration

Table 3.J.6.1: Distribution of individuals already in migrant households between waves 1 and 2 according to their status in waves 2 and 3

	Already in migrant households between waves 1 and 2	Not already in migrant households between waves 1 and 2	Total
Adults			
Left Behind between waves 2 and 3	74	396	470
	(15.74%)	(84.26%)	(100.00%)
Non Left Behind between waves 2 and 3	355	3,754	4,109
	(8.64%)	(91.36%)	(100.00%)
Total	429	4,150	4,579
	(9.37%)	(90.63%)	(100.00%)
Children			
Left Behind between waves 2 and 3	53	261	314
	(16.88%)	(83.12%)	(100.00%)
Non Left Behind between waves 2 and 3	210	2,256	2,466
	(8.51%)	(91.49%)	(100.00%)
Total	263	2,517	2,780
	(9.46%)	(90.54%)	(100.00%)

Note: The figures represent the number of individuals, with the percentages in parentheses corresponding to the proportion relative to the total of each row.

Table 3.J.6.2: Results from PSM-DID for adults, without migration between waves 1 and 2

	Weight			
	(1)	(2)	(3)	(4)
Panel A - All adults				
Left behind*Post	-1.016*	-1.009*	-1.023*	-1.075*
	(0.583)	(0.551)	(0.547)	(0.551)
Mean weight at baseline	62.887	62.887	62.887	62.887
R^2	0.001	0.122	0.135	0.135
Observations	8,058	8,058	8,058	8,058
Panel B - Males				
Left behind*Post	-1.863**	-1.837**	-1.837**	-1.874**
	(0.800)	(0.769)	(0.757)	(0.759)
Mean weight at baseline	63.724	63.724	63.724	63.724
R^2	0.004	0.101	0.133	0.134
Observations	3,546	3,546	3,546	3,546
Panel C - Females				
Left behind*Post	-0.361	-0.133	-0.144	-0.113
	(0.856)	(0.801)	(0.793)	(0.799)
Mean weight at baseline	62.174	62.174	62.174	62.174
R^2	0.005	0.148	0.168	0.168
Observations	4,156	4,156	4,156	4,156
Panel D - Underweight adults				
Left behind*Post	0.902	0.784	0.804	1.204
	(1.299)	(1.200)	(1.203)	(1.243)
Mean weight at baseline	46.944	46.944	46.944	46.944
R^2	0.147	0.315	0.327	0.330
Observations	496	496	496	496
Panel E - Healthy adults				
Left behind*Post	-1.163**	-1.120**	-1.128**	-1.244***
	(0.515)	(0.475)	(0.472)	(0.474)
Mean weight at baseline	58.047	58.047	58.047	58.047
R^2	0.009	0.177	0.187	0.189
Observations	4,416	4,416	4,416	4,416
Panel F - Overweight adults				
Left behind*Post	-1.801	-1.521	-1.521	-1.607
	(1.276)	(1.227)	(1.216)	(1.220)
Mean weight at baseline	75.305	75.305	75.305	75.305
R^2	0.032	0.132	0.152	0.153
Observations	2,316	2,316	2,316	2,316
Individual variables	No	Yes	Yes	Yes
Household head variables	No	Yes	Yes	Yes
Household variables	No	Yes	Yes	Yes
Rural dummy	No	No	Yes	Yes
Region dummies	No	No	Yes	Yes
Number of entries and exits	No	No	No	Yes

Notes: Standard errors are in parentheses. The variables listed at the bottom of the table are included in the DID, i.e., after matching. Individual variables include age, gender, and relationship to the head. Household head variables include age, gender, and education level of the head. Household variables include household size, number of working-age members, dependency ratio, and wealth index quintiles. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table 3.J.6.3: Results from PSM-DID for children, without migration between waves 1 and 2

	Zbmi			
	(1)	(2)	(3)	(4)
Panel A - All children				
Left behind*Post	-0.121	-0.164*	-0.149	-0.182*
	(0.099)	(0.098)	(0.096)	(0.097)
Mean zbmi at baseline	-0.096	-0.096	-0.096	-0.096
R^2	0.002	0.049	0.073	0.074
Observations	4,710	4,710	4,710	4,710
Panel B - Males				
Left behind*Post	-0.082	-0.173	-0.149	-0.178
	(0.138)	(0.135)	(0.134)	(0.135)
Mean zbmi at baseline	-0.123	-0.123	-0.123	-0.123
R^2	0.001	0.074	0.091	0.092
Observations	2,692	2,692	2,692	2,692
Panel C - Females				
Left behind*Post	-0.298**	-0.322**	-0.309**	-0.349**
	(0.144)	(0.144)	(0.141)	(0.143)
Mean zbmi at baseline	-0.063	-0.063	-0.063	-0.063
R^2	0.005	0.039	0.085	0.086
Observations	1,924	1,924	1,924	1,924
Panel D - Underweight children				
Left behind*Post	-0.031	-0.333	-0.291	-0.232
	(0.273)	(0.260)	(0.259)	(0.256)
Mean zbmi at baseline	-2.985	-2.985	-2.985	-2.985
R^2	0.463	0.565	0.585	0.599
Observations	398	398	398	398
Panel E - Healthy children				
Left behind*Post	-0.034	-0.052	-0.042	-0.075
	(0.087)	(0.087)	(0.087)	(0.087)
Mean zbmi at baseline	-0.403	-0.403	-0.403	-0.403
R^2	0.005	0.035	0.047	0.050
Observations	2,994	2,994	2,994	2,994
Panel F - Overweight children				
Left behind*Post	-0.546***	-0.585***	-0.581***	-0.690***
	(0.172)	(0.168)	(0.164)	(0.165)
Mean zbmi at baseline	2.288	2.288	2.288	2.288
R^2	0.399	0.447	0.481	0.489
Observations	1,092	1,092	1,092	1,092
Individual variables	No	Yes	Yes	Yes
Household head variables	No	Yes	Yes	Yes
Household variables	No	Yes	Yes	Yes
Rural dummy	No	No	Yes	Yes
Region dummies	No	No	Yes	Yes
Number of entries and exits	No	No	No	Yes

Notes: Standard errors are in parentheses. The variables listed at the bottom of the table are included in the DID, i.e., after matching. Individual variables include age, gender, and relationship to the head. Household head variables include age, gender, and education level of the head. Household variables include household size, number of working-age members, dependency ratio, and wealth index quintiles. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table 3.J.6.4: Results from PSM-DID for adults, excluding individuals in households with prior migrations between waves 1 and 2

	Weight			
	(1)	(2)	(3)	(4)
Panel A - All adults				
Left behind*Post	-0.499	-0.522	-0.503	-0.575
	(0.580)	(0.550)	(0.545)	(0.549)
Mean weight at baseline	62.889	62.889	62.889	62.889
R^2	0.001	0.118	0.133	0.133
Observations	7,918	7,918	7,918	7,918
Panel B - Males				
Left behind*Post	-1.595**	-1.605**	-1.595**	-1.607**
	(0.810)	(0.778)	(0.767)	(0.768)
Mean weight at baseline	63.705	63.705	63.705	63.705
R^2	0.003	0.098	0.127	0.128
Observations	3,486	3,486	3,486	3,486
Panel C - Females				
Left behind*Post	0.480	0.662	0.721	0.821
	(0.901)	(0.843)	(0.833)	(0.842)
Mean weight at baseline	62.189	62.189	62.189	62.189
R^2	0.005	0.148	0.173	0.173
Observations	3,644	3,644	3,644	3,644
Panel D - Underweight adults				
Left behind*Post	1.740	1.471	1.477	1.979
	(1.295)	(1.177)	(1.178)	(1.214)
Mean weight at baseline	46.948	46.948	46.948	46.948
R^2	0.182	0.367	0.379	0.383
Observations	504	504	504	504
Panel E - Healthy adults				
Left behind*Post	-0.600	-0.584	-0.572	-0.693
	(0.523)	(0.482)	(0.478)	(0.480)
Mean weight at baseline	59.063	59.063	59.063	59.063
R^2	0.010	0.178	0.191	0.193
Observations	4,742	4,276	4,276	4,276
Panel F - Overweight adults				
Left behind*Post	-1.683	-1.376	-1.284	-1.344
	(1.332)	(1.278)	(1.264)	(1.270)
Mean weight at baseline	75.276	75.276	75.276	75.276
R^2	0.030	0.133	0.158	0.158
Observations	2,098	2,098	2,098	2,098
Individual variables	No	Yes	Yes	Yes
Household head variables	No	Yes	Yes	Yes
Household variables	No	Yes	Yes	Yes
Rural dummy	No	No	Yes	Yes
Region dummies	No	No	Yes	Yes
Number of entries and exits	No	No	No	Yes

Notes: Standard errors are in parentheses. The variables listed at the bottom of the table are included in the DID, i.e., after matching. Individual variables include age, gender, and relationship to the head. Household head variables include age, gender, and education level of the head. Household variables include household size, number of working-age members, dependency ratio, and wealth index quintiles. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table 3.J.6.5: Results from PSM-DID for children, excluding individuals in households with prior migrations between waves 1 and 2

	Zbmi			
	(1)	(2)	(3)	(4)
Panel A - All children				
Left behind*Post	-0.148	-0.169*	-0.160	-0.213**
	(0.104)	(0.103)	(0.102)	(0.103)
Mean zbmi at baseline	-0.088	-0.088	-0.088	-0.088
R^2	0.001	0.049	0.065	0.068
Observations	4,284	4,284	4,284	4,284
Panel B - Males				
Left behind*Post	-0.102	-0.185	-0.175	-0.257*
	(0.151)	(0.149)	(0.148)	(0.150)
Mean zbmi at baseline	-0.107	-0.107	-0.107	-0.107
R^2	0.000	0.066	0.078	0.084
Observations	2,278	2,278	2,278	2,278
Panel C - Females				
Left behind*Post	-0.386**	-0.418***	-0.413***	-0.450***
	(0.153)	(0.152)	(0.150)	(0.153)
Mean zbmi at baseline	-0.063	-0.063	-0.063	-0.063
R^2	0.005	0.050	0.075	0.077
Observations	1,718	1,718	1,718	1,718
Panel D - Underweight children				
Left behind*Post	-0.496*	-0.597**	-0.560**	-0.465*
	(0.275)	(0.266)	(0.263)	(0.264)
Mean zbmi at baseline	-2.989	-2.989	-2.989	-2.989
R^2	0.460	0.544	0.568	0.579
Observations	396	396	396	396
Panel E - Healthy children				
Left behind*Post	0.029	0.022	0.024	-0.026
	(0.089)	(0.090)	(0.090)	(0.091)
Mean zbmi at baseline	-0.410	-0.410	-0.410	-0.410
R^2	0.007	0.033	0.040	0.047
Observations	2,788	2,788	2,788	2,788
Panel F - Overweight children				
Left behind*Post	-0.624***	-0.659***	-0.634***	-0.700***
	(0.189)	(0.183)	(0.179)	(0.182)
Mean zbmi at baseline	2.286	2.286	2.286	2.286
R^2	0.363	0.430	0.458	0.461
Observations	986	986	986	986
Individual variables	No	Yes	Yes	Yes
Household head variables	No	Yes	Yes	Yes
Household variables	No	Yes	Yes	Yes
Rural dummy	No	No	Yes	Yes
Region dummies	No	No	Yes	Yes
Number of entries and exits	No	No	No	Yes

Notes: Standard errors are in parentheses. The variables listed at the bottom of the table are included in the DID, i.e., after matching. Individual variables include age, gender, and relationship to the head. Household head variables include age, gender, and education level of the head. Household variables include household size, number of working-age members, dependency ratio, and wealth index quintiles. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

3.J.7 Matching without individual-level variables

Table 3.J.7.1: Results from PSM-DID for adults, matching without individual-level variables

	Weight			
	(1)	(2)	(3)	(4)
Panel A - All adults				
Left behind*Post	-1.305**	-1.195**	-1.172**	-1.158**
	(0.559)	(0.532)	(0.528)	(0.530)
Mean weight at baseline	62.765	62.765	62.765	62.765
R^2	0.002	0.116	0.130	0.130
Observations	8,718	8,718	8,718	8,718
Panel B - Males				
Left behind*Post	-2.011***	-1.972***	-1.991***	-1.992***
	(0.777)	(0.751)	(0.739)	(0.741)
Mean weight at baseline	63.633	63.633	63.633	63.633
R^2	0.004	0.093	0.124	0.124
Observations	3,792	3,792	3,792	3,792
Panel C - Females				
Left behind*Post	-0.624	-0.312	-0.250	-0.199
	(0.812)	(0.760)	(0.751)	(0.755)
Mean weight at baseline	62.036	62.036	62.036	62.036
R^2	0.005	0.147	0.172	0.172
Observations	4,626	4,626	4,626	4,626
Panel D - Underweight adults				
Left behind*Post	0.026	0.019	0.030	0.522
	(1.149)	(1.042)	(1.042)	(1.061)
Mean weight at baseline	46.819	46.819	46.819	46.819
R^2	0.145	0.342	0.353	0.359
Observations	604	604	604	604
Panel E - Healthy adults				
Left behind*Post	-1.140**	-1.171**	-1.171***	-1.271***
	(0.496)	(0.457)	(0.454)	(0.455)
Mean weight at baseline	58.998	58.998	58.998	58.998
R^2	0.011	0.180	0.191	0.193
Observations	4,796	4,796	4,796	4,796
Panel F - Overweight adults				
Left behind*Post	-1.684	-1.470	-1.472	-1.538
	(1.208)	(1.163)	(1.152)	(1.155)
Mean weight at baseline	75.279	75.279	75.279	75.279
R^2	0.029	0.128	0.149	0.150
Observations	2,562	2,562	2,562	2,562
Individual variables	No	Yes	Yes	Yes
Household head variables	No	Yes	Yes	Yes
Household variables	No	Yes	Yes	Yes
Rural dummy	No	No	Yes	Yes
Region dummies	No	No	Yes	Yes
Number of entries and exits	No	No	No	Yes

Notes: Standard errors are in parentheses. The variables listed at the bottom of the table are included in the DID, i.e., after matching. Individual variables include age, gender, and relationship to the head. Household head variables include age, gender, and education level of the head. Household variables include household size, number of working-age members, dependency ratio, and wealth index quintiles. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table 3.J.7.2: Results from PSM-DID for children, matching without individual-level variables

	Zbmi			
	(1)	(2)	(3)	(4)
Panel A - All children				
Left behind*Post	-0.184*	-0.227**	-0.213**	-0.239**
	(0.095)	(0.094)	(0.093)	(0.094)
Mean zbmi at baseline	-0.096	-0.096	-0.096	-0.096
R^2	0.002	0.049	0.070	0.071
Observations	5,040	5,040	5,040	5,040
Panel B - Males				
Left behind*Post	-0.163	-0.261**	-0.238*	-0.262**
	(0.130)	(0.128)	(0.127)	(0.128)
Mean zbmi at baseline	-0.123	-0.123	-0.123	-0.123
R^2	0.001	0.075	0.089	0.091
Observations	2,910	2,910	2,910	2,910
Panel C - Females				
Left behind*Post	-0.356**	-0.387***	-0.377***	-0.427***
	(0.138)	(0.138)	(0.135)	(0.138)
Mean zbmi at baseline	-0.063	-0.063	-0.063	-0.063
R^2	0.005	0.041	0.081	0.083
Observations	2,076	2,076	2,076	2,076
Panel D - Underweight children				
Left behind*Post	-0.180	-0.462**	-0.424*	-0.358
	(0.246)	(0.234)	(0.232)	(0.231)
Mean zbmi at baseline	-2.962	-2.962	-2.962	-2.962
R^2	0.476	0.576	0.597	0.609
Observations	466	466	466	466
Panel E - Healthy children				
Left behind*Post	-0.130	-0.161*	-0.152*	-0.183**
	(0.083)	(0.083)	(0.083)	(0.083)
Mean zbmi at baseline	-0.406	-0.406	-0.406	-0.406
R^2	0.007	0.035	0.045	0.049
Observations	3,294	3,294	3,294	3,294
Panel F - Overweight children				
Left behind*Post	-0.653***	-0.687***	-0.677***	-0.797***
	(0.170)	(0.166)	(0.162)	(0.164)
Mean zbmi at baseline	2.285	2.285	2.285	2.285
R^2	0.383	0.436	0.466	0.474
Observations	1,134	1,134	1,134	1,134
Individual variables	No	Yes	Yes	Yes
Household head variables	No	Yes	Yes	Yes
Household variables	No	Yes	Yes	Yes
Rural dummy	No	No	Yes	Yes
Region dummies	No	No	Yes	Yes
Number of entries and exits	No	No	No	Yes

Notes: Standard errors are in parentheses. The variables listed at the bottom of the table are included in the DID, i.e., after matching. Individual variables include age, gender, and relationship to the head. Household head variables include age, gender, and education level of the head. Household variables include household size, number of working-age members, dependency ratio, and wealth index quintiles. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

3.J.8 Potential endogeneity concerns

Some variables used for matching and DID analysis may be considered endogenous. It is particularly applicable to variables related to household headship. Specifically, variables associated with the household headship, such as age, gender, and educational level, could be endogenous with respect to the treatment (Bertoli and Marchetta, 2014). Similarly, the variable accounting for household size may also be endogenous. The potential endogeneity may invalidate their inclusion. As a result, it is imperative to ensure that these variables do not introduce bias to the results. To address this concern, I exclude these four variables (household head age, gender, and education level, as well as household size) from the matching and the DID analysis. The results are presented in Tables 3.J.8.1 and 3.J.8.2.

For most of the individuals, the results remain consistent. However, the coefficients are no longer significant for the regressions related to the adult population, albeit they approach significance in columns (2) to (4). A second noteworthy difference is the newfound significance for the healthy children subsample, suggesting that migration has a negative effect on the z-score of left-behind healthy children. Nonetheless, the other results remain robust. The notion underlying the possibility of some variables related to household headship being endogenous is based on the idea that migration could entail a change in household headship. However, upon thorough data exploration, I observed that less than 4% of migrants were household heads in Wave 2, i.e., prior to their migration. Furthermore, there is very little change in household heads, both among migrant and non-migrant households (less than 1% in both cases). Hence, even though we have examined the results without these variables, I believe that endogeneity related to household headship is not an issue in our case.[31]

[31] Similarly, the variables related to the wealth index, i.e., the wealth index quintiles, can also be considered endogenous. The results obtained by dropping the wealth index quintile variables in both the matching and the difference-in-differences are very similar to the main results. These results can be provided upon request.

Table 3.J.8.1: Results from PSM-DID for adults, without household headship variables and household size

	Weight			
	(1)	(2)	(3)	(4)
Panel A - All adults				
Left behind*Post	-0.938*	-0.853	-0.822	-0.836
	(0.552)	(0.529)	(0.525)	(0.527)
Mean weight at baseline	62.784	62.784	62.784	62.784
R^2	0.002	0.099	0.113	0.114
Observations	9,120	9,120	9,120	9,120
Panel B - Males				
Left behind*Post	-1.956***	-1.943***	-1.863**	-1.856**
	(0.745)	(0.732)	(0.724)	(0.725)
Mean weight at baseline	63.660	63.660	63.660	63.660
R^2	0.003	0.057	0.081	0.081
Observations	4,122	4,122	4,122	4,122
Panel C - Females				
Left behind*Post	-0.362	-0.155	-0.089	-0.054
	(0.799)	(0.753)	(0.742)	(0.747)
Mean weight at baseline	62.055	62.055	62.055	62.055
R^2	0.004	0.129	0.157	0.157
Observations	4,784	4,784	4,784	4,784
Panel D - Underweight adults				
Left behind*Post	0.850	0.800	0.780	1.017
	(1.094)	(1.027)	(1.028)	(1.052)
Mean weight at baseline	46.835	46.835	46.835	46.835
R^2	0.143	0.265	0.273	0.275
Observations	734	734	734	734
Panel E - Healthy adults				
Left behind*Post	-1.225**	-1.235***	-1.251***	-1.376***
	(0.480)	(0.443)	(0.441)	(0.443)
Mean weight at baseline	59.000	59.000	59.000	59.000
R^2	0.009	0.170	0.178	0.180
Observations	5,206	5,206	5,206	5,206
Panel F - Overweight adults				
Left behind*Post	-0.955	-0.621	-0.512	-0.535
	(1.213)	(1.190)	(1.175)	(1.176)
Mean weight at baseline	75.265	75.265	75.265	75.265
R^2	0.032	0.134	0.156	0.156
Observations	2,634	2,634	2,634	2,634
Individual variables	No	Yes	Yes	Yes
Household variables	No	Yes	Yes	Yes
Rural dummy	No	No	Yes	Yes
Region dummies	No	No	Yes	Yes
Number of entries and exits	No	No	No	Yes

Notes: Standard errors are in parentheses. The variables listed at the bottom of the table are included in the DID, i.e., after matching. Individual variables include age and gender. Household variables include number of working-age members, dependency ratio and wealth index quintiles. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table 3.J.8.2: Results from PSM-DID for children, without household headship variables and household size

	Zbmi			
	(1)	(2)	(3)	(4)
Panel A - All children				
Left behind*Post	-0.200**	-0.245***	-0.237**	-0.266***
	(0.095)	(0.094)	(0.094)	(0.095)
Mean zbmi at baseline	-0.090	-0.090	-0.090	-0.090
R^2	0.002	0.041	0.059	0.059
Observations	5,014	5,014	5,014	5,014
Panel B - Males				
Left behind*Post	-0.129	-0.242*	-0.233*	-0.249*
	(0.137)	(0.135)	(0.135)	(0.135)
Mean zbmi at baseline	-0.118	-0.118	-0.118	-0.118
R^2	0.001	0.066	0.079	0.080
Observations	2,712	2,712	2,712	2,712
Panel C - Females				
Left behind*Post	-0.309**	-0.328**	-0.316**	-0.373***
	(0.137)	(0.136)	(0.135)	(0.137)
Mean zbmi at baseline	-0.055	-0.055	-0.055	-0.055
R^2	0.005	0.025	0.054	0.057
Observations	2,142	2,142	2,142	2,142
Panel D - Underweight children				
Left behind*Post	-0.004	-0.049	-0.009	0.008
	(0.242)	(0.240)	(0.237)	(0.237)
Mean zbmi at baseline	-2.962	-2.962	-2.962	-2.962
R^2	0.451	0.489	0.510	0.520
Observations	480	480	480	480
Panel E - Healthy children				
Left behind*Post	-0.150*	-0.189**	-0.182**	-0.208**
	(0.085)	(0.085)	(0.085)	(0.086)
Mean zbmi at baseline	-0.406	-0.406	-0.406	-0.406
R^2	0.006	0.029	0.039	0.042
Observations	3,098	3,098	3,098	3,098
Panel F - Overweight children				
Left behind*Post	-0.572***	-0.611***	-0.611***	-0.714***
	(0.168)	(0.166)	(0.163)	(0.164)
Mean zbmi at baseline	2.288	2.288	2.288	2.288
R^2	0.380	0.412	0.442	0.451
Observations	1,216	1,216	1,216	1,216
Individual variables	No	Yes	Yes	Yes
Household variables	No	Yes	Yes	Yes
Rural dummy	No	No	Yes	Yes
Region dummies	No	No	Yes	Yes
Number of entries and exits	No	No	No	Yes

Notes: Standard errors are in parentheses. The variables listed at the bottom of the table are included in the DID, i.e., after matching. Individual variables include age and gender. Household variables include number of working-age members, dependency ratio and wealth index quintiles. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

3.J.9 Alternative model including variables of labor market participation

As explained in the empirical strategy section, I explore the potential inclusion of labor market participation variables. This consideration stems from the understanding that labor market conditions might differ between left-behind and non-left-behind individuals, influencing migration decisions and, by extension, nutritional outcomes. Also, left-behind individuals might come from households with poorer labor market conditions, potentially affecting the estimated impact on their nutritional status. This consideration also applies to the labor market conditions of other household members, acknowledging the influence of collective income sharing on individual nutritional status.

Based on the employment screener section of the questionnaire,[32] I constructed variables reflecting both individual and household-level labor market participation. These are employed in the matching process and as controls in the DID. At the individual level, I categorized individuals as economically active, inactive, or looking for a job. At the household level, I constructed three variables representing the number of individuals per household in each of the above job categories. The results, incorporating these variables, are presented in Tables 3.J.9.1 and 3.J.9.2 below for adults and children respectively. Notably, for adults, while results are largely consistent, the introduction of these variables rendered the coefficient for overweight or obese individuals significant, suggesting a weight decline among left-behind adults post-migration. For children, the results varied slightly in coefficients but maintained similar significance and signs.

However, these labor market variables were excluded from the main model. My reservations are threefold: 1) The imperfect representation of labor market conditions due to overlapping categories of activities in the survey data. Indeed, it is challenging to distinctly ascertain whether some individuals are economically active or not. For example, individuals may simultaneously report being full-time homemakers or students while owning an enterprise or contributing to a household farm. Additionally, another section of the questionnaire asking for the main paid occupation over the last seven days is only answered by a small subset of all the individuals currently working; 2) There is a temporal mismatch between the observed labor conditions and the actual migration decisions. Indeed, migration occurs between waves 2 and 3, and there may be a disparity between the labor market conditions observed in Wave 2 and those influencing migration decisions between waves 2 and 3[33]; 3) Additionally, the presence of missing data for these variables further complicated their use, significantly reducing the usable sample size.

[32] In Wave 2, this information appears in Section 1: Individual Information, Part D: Background Information, 1EA: Employment Screener, and in Wave 3, in Section 1: Household Background, Part F: Employment, 0: Employment Screener.

[33] Unfortunately, this issue cannot be mitigated by using Wave 1 to demonstrate that labor market conditions are persistent over time, as the relevant questions from waves 2 and 3 were not included in Wave 1

Table 3.J.9.1: Results from PSM-DID for adults, with variables of labor market participation

	Weight			
	(1)	(2)	(3)	(4)
Panel A - All adults				
Left behind*Post	-1.248**	-1.063*	-1.112*	-1.076*
	(0.606)	(0.573)	(0.569)	(0.572)
Mean weight at baseline	62.440	62.440	62.440	62.440
R^2	0.001	0.124	0.140	0.140
Observations	7,204	7,204	7,204	7,204
Panel B - Males				
Left behind*Post	-2.043**	-2.222***	-2.144**	-2.172***
	(0.873)	(0.840)	(0.833)	(0.836)
Mean weight at baseline	63.193	63.193	63.193	63.193
R^2	0.004	0.107	0.127	0.127
Observations	2,926	2,926	2,926	2,926
Panel C - Females				
Left behind*Post	-0.380	0.199	0.081	0.235
	(0.915)	(0.849)	(0.838)	(0.843)
Mean weight at baseline	61.792	61.792	61.792	61.792
R^2	0.003	0.165	0.190	0.191
Observations	3,452	3,452	3,452	3,452
Panel D - Underweight adults				
Left behind*Post	1.699	1.399	1.364	1.657
	(1.395)	(1.285)	(1.276)	(1.331)
Mean weight at baseline	46.835	46.835	46.835	46.835
R^2	0.143	0.346	0.375	0.379
Observations	422	422	422	422
Panel E - Healthy adults				
Left behind*Post	-0.942*	-0.925*	-0.955*	-1.025**
	(0.566)	(0.517)	(0.516)	(0.518)
Mean weight at baseline	58.914	58.914	58.914	58.914
R^2	0.007	0.194	0.199	0.200
Observations	3,724	3,724	3,724	3,724
Panel F - Overweight adults				
Left behind*Post	-3.128**	-2.790**	-2.536**	-2.639**
	(1.326)	(1.280)	(1.262)	(1.269)
Mean weight at baseline	75.009	75.009	75.009	75.009
R^2	0.037	0.140	0.170	0.170
Observations	2,102	2,102	2,102	2,102
Individual variables	No	Yes	Yes	Yes
Household head variables	No	Yes	Yes	Yes
Household variables	No	Yes	Yes	Yes
Rural dummy	No	No	Yes	Yes
Region dummies	No	No	Yes	Yes
Number of entries and exits	No	No	No	Yes

Notes: Standard errors are in parentheses. The variables listed at the bottom of the table are included in the DID, i.e., after matching. Individual variables include age, gender, relationship to the head, and whether the individual is economically active, inactive, and looking for a job. Household head variables include age, gender, and education level of the head. Household variables include household size, number of working-age members, dependency ratio, wealth index quintiles, and the number of household members who are economically active, inactive, and looking for a job. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table 3.J.9.2: Results from PSM-DID for children, with variables of labor market participation

	Zbmi			
	(1)	(2)	(3)	(4)
Panel A - All children				
Left behind*Post	-0.197*	-0.270***	-0.246**	-0.284***
	(0.104)	(0.103)	(0.102)	(0.103)
Mean zbmi at baseline	-0.099	-0.099	-0.099	-0.099
R^2	0.002	0.054	0.080	0.084
Observations	4,286	4,286	4,286	4,286
Panel B - Males				
Left behind*Post	-0.259*	-0.277*	-0.259*	-0.274*
	(0.148)	(0.143)	(0.142)	(0.143)
Mean zbmi at baseline	-0.105	-0.105	-0.105	-0.105
R^2	0.002	0.098	0.118	0.122
Observations	2,460	2,460	2,460	2,460
Panel C - Females				
Left behind*Post	-0.346**	-0.428***	-0.443***	-0.527***
	(0.156)	(0.155)	(0.151)	(0.153)
Mean zbmi at baseline	-0.091	-0.091	-0.091	-0.091
R^2	0.001	0.053	0.107	0.114
Observations	1,644	1,644	1,644	1,644
Panel D - Underweight children				
Left behind*Post	0.122	-0.008	-0.047	-0.093
	(0.330)	(0.294)	(0.298)	(0.293)
Mean zbmi at baseline	-2.958	-2.958	-2.958	-2.958
R^2	0.465	0.659	0.665	0.685
Observations	274	274	274	274
Panel E - Healthy children				
Left behind*Post	-0.042	-0.078	-0.070	-0.114
	(0.095)	(0.096)	(0.096)	(0.096)
Mean zbmi at baseline	-0.405	-0.405	-0.405	-0.405
R^2	0.005	0.036	0.046	0.053
Observations	2,632	2,632	2,632	2,632
Panel F - Overweight children				
Left behind*Post	-0.767***	-0.955***	-0.956***	-1.120***
	(0.230)	(0.225)	(0.221)	(0.222)
Mean zbmi at baseline	2.307	2.307	2.307	2.307
R^2	0.402	0.467	0.497	0.511
Observations	658	658	658	658
Individual variables	No	Yes	Yes	Yes
Household head variables	No	Yes	Yes	Yes
Household variables	No	Yes	Yes	Yes
Rural dummy	No	No	Yes	Yes
Region dummies	No	No	Yes	Yes
Number of entries and exits	No	No	No	Yes

Notes: Standard errors are in parentheses. The variables listed at the bottom of the table are included in the DID, i.e., after matching. Individual variables include age, gender, relationship to the head, and whether the individual is economically active, inactive, and looking for a job. Household head variables include age, gender, and education level of the head. Household variables include household size, number of working-age members, dependency ratio, wealth index quintiles, and the number of household members who are economically active, inactive, and looking for a job. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

3.J.10 Heterogeneity in migration temporalities

I define two new categories of left behind: those left behind by a migrant for six to twelve months and those left behind by a migrant for more than twelve months. Previously, these two categories of individuals were grouped. The results are available in Tables 3.J.10.1 to 3.J.10.4.

For adults, the initial stages of migration have a negative impact on their body weight (Table 3.J.10.1). However, this effect seems to dissipate after twelve months of migration (Table 3.J.10.3). Regarding gender, women are more affected during the early stages of migration (between six and twelve months), but this effect vanishes after twelve months. In contrast, men are impacted more significantly after twelve months. Regarding nutritional status, healthy adults are primarily affected during the first six to twelve months of migration but less so afterward. On the other hand, undernourished adults exhibit a positive effect on weight as time since migration progresses.

Overall, adults are less dependent on the duration of a member's migration compared to children. Indeed, for children, we observe increasingly positive and significant effects as the duration of migration extends, and the magnitude of these effects also rises (Tables 3.J.10.2 and 3.J.10.4). For instance, while the coefficient increases for all children, it becomes distinctly negative for both boys and girls. In summary, while the effects on adults are heterogeneous with the duration of migration, children are more impacted when migration happened more than twelve months ago compared to six months ago.

Table 3.J.10.1: Results from PSM-DID for adults, migration between six and twelve months

	Weight			
	(1)	(2)	(3)	(4)
Panel A - All adults				
Left behind*Post	-1.709***	-1.857***	-1.695***	-1.830***
	(0.591)	(0.555)	(0.547)	(0.550)
Mean weight at baseline	62.875	62.875	62.875	62.875
R^2	0.008	0.142	0.170	0.170
Observations	7,644	7,644	7,644	7,644
Panel B - Males				
Left behind*Post	-1.655	-1.460	-1.294	-1.670*
	(1.061)	(0.976)	(0.963)	(0.971)
Mean weight at baseline	63.620	63.620	63.620	63.620
R^2	0.004	0.189	0.218	0.224
Observations	1,468	1,468	1,468	1,468
Panel C - Females				
Left behind*Post	-1.927**	-1.797**	-1.770**	-1.878**
	(0.867)	(0.797)	(0.772)	(0.778)
Mean weight at baseline	62.241	62.241	62.241	62.241
R^2	0.009	0.186	0.239	0.239
Observations	4,192	4,192	4,192	4,192
Panel D - Underweight adults				
Left behind*Post	0.078	-0.242	-0.253	-0.401
	(1.952)	(1.807)	(1.794)	(1.890)
Mean weight at baseline	46.733	46.733	46.733	46.733
R^2	0.102	0.337	0.364	0.364
Observations	212	212	212	212
Panel E - Healthy adults				
Left behind*Post	-2.193***	-2.500***	-2.415***	-2.759***
	(0.514)	(0.471)	(0.464)	(0.467)
Mean weight at baseline	59.111	59.111	59.111	59.111
R^2	0.013	0.191	0.219	0.225
Observations	4,468	4,468	4,468	4,468
Panel F - Overweight adults				
Left behind*Post	-2.491	-1.239	-1.189	-1.551
	(2.030)	(1.836)	(1.798)	(1.798)
Mean weight at baseline	75.339	75.339	75.339	75.339
R^2	0.039	0.266	0.311	0.317
Observations	888	888	888	888
Individual variables	No	Yes	Yes	Yes
Household head variables	No	Yes	Yes	Yes
Household variables	No	Yes	Yes	Yes
Rural dummy	No	No	Yes	Yes
Region dummies	No	No	Yes	Yes
Number of entries and exits	No	No	No	Yes

Notes: Standard errors are in parentheses. The variables listed at the bottom of the table are included in the DID, i.e., after matching. Individual variables include age, gender, and relationship to the head. Household head variables include age, gender, and education level of the head. Household variables include household size, number of working-age members, dependency ratio, and wealth index quintiles. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table 3.J.10.2: Results from PSM-DID for children, migration between six and twelve months

	Zbmi			
	(1)	(2)	(3)	(4)
Panel A - All children				
Left behind*Post	-0.175*	-0.197**	-0.191**	-0.182*
	(0.095)	(0.094)	(0.093)	(0.094)
Mean zbmi at baseline	-0.104	-0.104	-0.104	-0.104
R^2	0.001	0.055	0.068	0.070
Observations	4,736	4,736	4,736	4,736
Panel B - Males				
Left behind*Post	-0.053	-0.127	-0.100	-0.030
	(0.168)	(0.162)	(0.162)	(0.164)
Mean zbmi at baseline	-0.128	-0.128	-0.128	-0.128
R^2	0.005	0.118	0.126	0.137
Observations	1,662	1,662	1,662	1,662
Panel C - Females				
Left behind*Post	-0.191	-0.219	-0.218	-0.272**
	(0.141)	(0.141)	(0.136)	(0.137)
Mean zbmi at baseline	-0.074	-0.074	-0.074	-0.074
R^2	0.001	0.034	0.113	0.117
Observations	1,970	1,970	1,970	1,970
Panel D - Underweight children				
Left behind*Post	-0.267	-0.333	-0.325	-0.293
	(0.325)	(0.320)	(0.323)	(0.326)
Mean zbmi at baseline	-2.961	-2.961	-2.961	-2.961
R^2	0.433	0.573	0.597	0.600
Observations	246	246	246	246
Panel E - Healthy children				
Left behind*Post	-0.300***	-0.297***	-0.298***	-0.320***
	(0.109)	(0.108)	(0.107)	(0.109)
Mean zbmi at baseline	-0.416	-0.416	-0.416	-0.416
R^2	0.007	0.070	0.089	0.090
Observations	1,928	1,928	1,928	1,928
Panel F - Overweight children				
Left behind*Post	-0.223	-0.249	-0.249	-0.194
	(0.292)	(0.283)	(0.284)	(0.296)
Mean zbmi at baseline	2.276	2.276	2.276	2.276
R^2	0.408	0.498	0.506	0.507
Observations	316	316	316	316
Individual variables	No	Yes	Yes	Yes
Household head variables	No	Yes	Yes	Yes
Household variables	No	Yes	Yes	Yes
Rural dummy	No	No	Yes	Yes
Region dummies	No	No	Yes	Yes
Number of entries and exits	No	No	No	Yes

Notes: Standard errors are in parentheses. The variables listed at the bottom of the table are included in the DID, i.e., after matching. Individual variables include age, gender, and relationship to the head. Household head variables include age, gender, and education level of the head. Household variables include household size, number of working-age members, dependency ratio, and wealth index quintiles. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table 3.J.10.3: Results from PSM-DID for adults, migration for more than twelve months

	Weight			
	(1)	(2)	(3)	(4)
Panel A - All adults				
Left behind*Post	-0.954*	-0.705	-0.713	-0.702
	(0.575)	(0.544)	(0.539)	(0.541)
Mean weight at baseline	62.831	62.831	62.831	62.831
R^2	0.001	0.125	0.142	0.142
Observations	8,422	8,422	8,422	8,422
Panel B - Males				
Left behind*Post	-2.042**	-1.868**	-1.787**	-1.745**
	(0.800)	(0.772)	(0.760)	(0.761)
Mean weight at baseline	63.660	63.660	63.660	63.660
R^2	0.004	0.099	0.130	0.130
Observations	3,798	3,798	3,798	3,798
Panel C - Females				
Left behind*Post	0.016	0.358	0.321	0.338
	(0.845)	(0.784)	(0.775)	(0.779)
Mean weight at baseline	62.127	62.127	62.127	62.127
R^2	0.003	0.163	0.184	0.185
Observations	4,266	4,266	4,266	4,266
Panel D - Underweight adults				
Left behind*Post	2.230	2.262*	2.263*	2.642**
	(1.367)	(1.271)	(1.269)	(1.282)
Mean weight at baseline	46.816	46.816	46.816	46.816
R^2	0.181	0.374	0.390	0.395
Observations	460	460	460	460
Panel E - Healthy adults				
Left behind*Post	-0.887*	-0.716	-0.734	-0.753
	(0.524)	(0.477)	(0.473)	(0.475)
Mean weight at baseline	59.043	59.043	59.043	59.043
R^2	0.009	0.198	0.212	0.213
Observations	4,368	4,368	4,368	4,368
Panel F - Overweight adults				
Left behind*Post	-0.710	-0.544	-0.568	-0.545
	(1.287)	(1.239)	(1.218)	(1.220)
Mean weight at baseline	75.257	75.257	75.257	75.257
R^2	0.030	0.133	0.167	0.168
Observations	2,386	2,386	2,386	2,386
Individual variables	No	Yes	Yes	Yes
Household head variables	No	Yes	Yes	Yes
Household variables	No	Yes	Yes	Yes
Rural dummy	No	No	Yes	Yes
Region dummies	No	No	Yes	Yes
Number of entries and exits	No	No	No	Yes

Notes: Standard errors are in parentheses. The variables listed at the bottom of the table are included in the DID, i.e., after matching. Individual variables include age, gender, and relationship to the head. Household head variables include age, gender, and education level of the head. Household variables include household size, number of working-age members, dependency ratio, and wealth index quintiles. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table 3.J.10.4: Results from PSM-DID for children, migration for more than twelve months

	Zbmi			
	(1)	(2)	(3)	(4)
Panel A - All children				
Left behind*Post	-0.253**	-0.347***	-0.319***	-0.349***
	(0.101)	(0.099)	(0.098)	(0.098)
Mean zbmi at baseline	-0.094	-0.094	-0.094	-0.094
R^2	0.003	0.062	0.091	0.093
Observations	4,662	4,662	4,662	4,662
Panel B - Males				
Left behind*Post	-0.209	-0.338**	-0.309**	-0.330**
	(0.140)	(0.137)	(0.135)	(0.135)
Mean zbmi at baseline	-0.117	-0.117	-0.117	-0.117
R^2	0.001	0.081	0.109	0.113
Observations	2,598	2,598	2,598	2,598
Panel C - Females				
Left behind*Post	-0.548***	-0.599***	-0.594***	-0.627***
	(0.200)	(0.201)	(0.198)	(0.201)
Mean zbmi at baseline	-0.065	-0.065	-0.065	-0.065
R^2	0.010	0.056	0.091	0.092
Observations	1,018	1,018	1,018	1,018
Panel D - Underweight children				
Left behind*Post	0.080	-0.270	-0.217	-0.178
	(0.294)	(0.273)	(0.270)	(0.270)
Mean zbmi at baseline	-2.951	-2.951	-2.951	-2.951
R^2	0.464	0.605	0.637	0.642
Observations	316	316	316	316
Panel E - Healthy children				
Left behind*Post	-0.050	-0.097	-0.078	-0.117
	(0.088)	(0.088)	(0.088)	(0.088)
Mean zbmi at baseline	-0.417	-0.417	-0.417	-0.417
R^2	0.005	0.031	0.048	0.054
Observations	2,920	2,920	2,920	2,920
Panel F - Overweight children				
Left behind*Post	-0.794***	-0.808***	-0.795***	-0.931***
	(0.182)	(0.177)	(0.172)	(0.174)
Mean zbmi at baseline	2.287	2.287	2.287	2.287
R^2	0.369	0.429	0.471	0.478
Observations	1,102	1,102	1,102	1,102
Individual variables	No	Yes	Yes	Yes
Household head variables	No	Yes	Yes	Yes
Household variables	No	Yes	Yes	Yes
Rural dummy	No	No	Yes	Yes
Region dummies	No	No	Yes	Yes
Number of entries and exits	No	No	No	Yes

Notes: Standard errors are in parentheses. The variables listed at the bottom of the table are included in the DID, i.e., after matching. Individual variables include age, gender, and relationship to the head. Household head variables include age, gender, and education level of the head. Household variables include household size, number of working-age members, dependency ratio, and wealth index quintiles. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

3.J.11 Heterogeneity by the number of migrants

The average number of migrants per household is about 1.2 individuals. In this appendix section, I present the results of heterogeneity based on whether the left-behind individuals are in households with one migrant or more than one migrant. Based on the results from Tables 3.J.11.1 to 3.J.11.4, for adults, we still observe a negative and significant effect of an individual's migration on their body weight. However, individuals in households with more than one migrant no longer exhibit a significant effect. In contrast, for children, we observe a more pronounced negative effect on the BMI-for-age z-score when there is more than one migrant in the household.

In our sample of adults, 394 adults are in households with precisely one migrant for work for more than six months and 76 adults are in households with more than one migrant. In the children's sample, the corresponding figures are 251 children and 63 children. Since the number of individuals decreases for households with more than one migrant, I present the results without restricting to the common support to retain the maximum number of observations.

Table 3.J.11.1: Results from PSM-DID for adults, one migrant

	Weight			
	(1)	(2)	(3)	(4)
Left behind*Post	-1.135**	-1.305**	-1.333**	-1.402***
	(0.561)	(0.527)	(0.523)	(0.524)
Mean weight at baseline	62.801	62.801	62.801	62.801
R^2	0.001	0.133	0.149	0.149
Observations	9,006	9,006	9,006	9,006
Individual variables	No	Yes	Yes	Yes
Household head variables	No	Yes	Yes	Yes
Household variables	No	Yes	Yes	Yes
Rural dummy	No	No	Yes	Yes
Region dummies	No	No	Yes	Yes
Number of entries and exits	No	No	No	Yes

Notes: Standard errors are in parentheses. The variables listed at the bottom of the table are included in the DID, i.e., after matching. Individual variables include age, gender, and relationship to the head. Household head variables include age, gender, and education level of the head. Household variables include household size, number of working-age members, dependency ratio, and wealth index quintiles. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table 3.J.11.2: Results from PSM-DID for children, one migrant

	Zbmi			
	(1)	(2)	(3)	(4)
Left behind*Post	-0.243***	-0.294***	-0.289***	-0.297***
	(0.092)	(0.090)	(0.090)	(0.090)
Mean zbmi at baseline	-0.096	-0.096	-0.096	-0.096
R^2	0.002	0.052	0.068	0.068
Observations	5,432	5,432	5,432	5,432
Individual variables	No	Yes	Yes	Yes
Household head variables	No	Yes	Yes	Yes
Household variables	No	Yes	Yes	Yes
Rural dummy	No	No	Yes	Yes
Region dummies	No	No	Yes	Yes
Number of entries and exits	No	No	No	Yes

Notes: Standard errors are in parentheses. The variables listed at the bottom of the table are included in the DID, i.e., after matching. Individual variables include age, gender, and relationship to the head. Household head variables include age, gender, and education level of the head. Household variables include household size, number of working-age members, dependency ratio, and wealth index quintiles. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table 3.J.11.3: Results from PSM-DID for adults, more than one migrant

	Weight			
	(1)	(2)	(3)	(4)
Left behind*Post	-1.425**	-0.512	-0.473	-0.261
	(0.614)	(0.587)	(0.583)	(0.591)
Mean weight at baseline	62.909	62.909	62.909	62.909
R^2	0.002	0.128	0.142	0.143
Observations	6,172	6,172	6,172	6,172
Individual variables	No	Yes	Yes	Yes
Household head variables	No	Yes	Yes	Yes
Household variables	No	Yes	Yes	Yes
Rural dummy	No	No	Yes	Yes
Region dummies	No	No	Yes	Yes
Number of entries and exits	No	No	No	Yes

Notes: Standard errors are in parentheses. The variables listed at the bottom of the table are included in the DID, i.e., after matching. Individual variables include age, gender, and relationship to the head. Household head variables include age, gender, and education level of the head. Household variables include household size, number of working-age members, dependency ratio, and wealth index quintiles. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table 3.J.11.4: Results from PSM-DID for children, more than one migrant

	Zbmi			
	(1)	(2)	(3)	(4)
Left behind*Post	-0.254**	-0.365***	-0.330***	-0.502***
	(0.109)	(0.106)	(0.102)	(0.105)
Mean zbmi at baseline	-0.102	-0.102	-0.102	-0.102
R^2	0.002	0.099	0.169	0.182
Observations	3,748	3,748	3,748	3,748
Individual variables	No	Yes	Yes	Yes
Household head variables	No	Yes	Yes	Yes
Household variables	No	Yes	Yes	Yes
Rural dummy	No	No	Yes	Yes
Region dummies	No	No	Yes	Yes
Number of entries and exits	No	No	No	Yes

Notes: Standard errors are in parentheses. The variables listed at the bottom of the table are included in the DID, i.e., after matching. Individual variables include age, gender, and relationship to the head. Household head variables include age, gender, and education level of the head. Household variables include household size, number of working-age members, dependency ratio, and wealth index quintiles. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

3.J.12 Migrants over 15 years old only

Table 3.J.12.1: Results from PSM-DID for adults, migrants over 15 years old only

	Weight			
	(1)	(2)	(3)	(4)
Panel A - All adults				
Left behind*Post	-1.350**	-1.170**	-1.158**	-1.137**
	(0.591)	(0.558)	(0.554)	(0.557)
Mean weight at baseline	62.810	62.810	62.810	62.810
R^2	0.002	0.125	0.141	0.141
Observations	7,956	7,956	7,956	7,956
Panel B - Males				
Left behind*Post	-1.369*	-1.349*	-1.303*	-1.293*
	(0.805)	(0.770)	(0.757)	(0.760)
Mean weight at baseline	63.617	63.617	63.617	63.617
R^2	0.002	0.105	0.140	0.140
Observations	3,406	3,406	3,406	3,406
Panel C - Females				
Left behind*Post	-1.054	-0.654	-0.645	-0.574
	(0.883)	(0.824)	(0.814)	(0.821)
Mean weight at baseline	62.131	62.131	62.131	62.131
R^2	0.006	0.154	0.177	0.177
Observations	4,072	4,072	4,072	4,072
Panel D - Underweight adults				
Left behind*Post	0.136	0.309	0.295	0.994
	(1.466)	(1.285)	(1.266)	(1.324)
Mean weight at baseline	46.752	46.752	46.752	46.752
R^2	0.145	0.381	0.409	0.414
Observations	464	464	464	464
Panel E - Healthy adults				
Left behind*Post	-1.331***	-1.292***	-1.301***	-1.351***
	(0.501)	(0.455)	(0.452)	(0.454)
Mean weight at baseline	59.027	59.027	59.027	59.027
R^2	0.011	0.196	0.210	0.211
Observations	4,820	4,820	4,820	4,820
Panel F - Overweight adults				
Left behind*Post	-0.598	-0.198	-0.179	-0.194
	(1.289)	(1.239)	(1.223)	(1.228)
Mean weight at baseline	75.257	75.257	75.257	75.257
R^2	0.032	0.139	0.166	0.166
Observations	2,370	2,370	2,370	2,370
Individual variables	No	Yes	Yes	Yes
Household head variables	No	Yes	Yes	Yes
Household variables	No	Yes	Yes	Yes
Rural dummy	No	No	Yes	Yes
Region dummies	No	No	Yes	Yes
Number of entries and exits	No	No	No	Yes

Notes: Standard errors are in parentheses. The variables listed at the bottom of the table are included in the DID, i.e., after matching. Individual variables include age, gender, and relationship to the head. Household head variables include age, gender, and education level of the head. Household variables include household size, number of working-age members, dependency ratio, and wealth index quintiles. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table 3.J.12.2: Results from PSM-DID for children, migrants over 15 years old only

	Zbmi			
	(1)	(2)	(3)	(4)
Panel A - All children				
Left behind*Post	-0.322***	-0.366***	-0.351***	-0.364***
	(0.102)	(0.100)	(0.099)	(0.100)
Mean zbmi at baseline	-0.104	-0.104	-0.104	-0.104
R^2	0.003	0.060	0.078	0.078
Observations	4,346	4,346	4,346	4,346
Panel B - Males				
Left behind*Post	-0.383***	-0.485***	-0.458***	-0.446***
	(0.145)	(0.141)	(0.141)	(0.141)
Mean zbmi at baseline	-0.128	-0.128	-0.128	-0.128
R^2	0.003	0.093	0.106	0.107
Observations	2,348	2,348	2,348	2,348
Panel C - Females				
Left behind*Post	-0.464***	-0.494***	-0.500***	-0.589***
	(0.154)	(0.153)	(0.151)	(0.153)
Mean zbmi at baseline	-0.073	-0.073	-0.073	-0.073
R^2	0.006	0.043	0.080	0.086
Observations	1,652	1,652	1,652	1,652
Panel D - Underweight children				
Left behind*Post	-0.490	-0.834***	-0.769***	-0.572**
	(0.307)	(0.277)	(0.272)	(0.267)
Mean zbmi at baseline	-2.961	-2.961	-2.961	-2.961
R^2	0.436	0.585	0.613	0.638
Observations	376	376	376	376
Panel E - Healthy children				
Left behind*Post	-0.379***	-0.394***	-0.391***	-0.406***
	(0.084)	(0.084)	(0.084)	(0.084)
Mean zbmi at baseline	-0.411	-0.411	-0.411	-0.411
R^2	0.008	0.051	0.059	0.064
Observations	2,878	2,878	2,878	2,878
Panel F - Overweight children				
Left behind*Post	-0.607***	-0.548***	-0.556***	-0.612***
	(0.184)	(0.177)	(0.174)	(0.182)
Mean zbmi at baseline	2.284	2.284	2.284	2.284
R^2	0.324	0.406	0.436	0.437
Observations	1,026	1,026	1,026	1,026
Individual variables	No	Yes	Yes	Yes
Household head variables	No	Yes	Yes	Yes
Household variables	No	Yes	Yes	Yes
Rural dummy	No	No	Yes	Yes
Region dummies	No	No	Yes	Yes
Number of entries and exits	No	No	No	Yes

Notes: Standard errors are in parentheses. The variables listed at the bottom of the table are included in the DID, i.e., after matching. Individual variables include age, gender, and relationship to the head. Household head variables include age, gender, and education level of the head. Household variables include household size, number of working-age members, dependency ratio, and wealth index quintiles. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

3.K Results considering remittances

Table 3.K.1: Results from PSM-DID for adults, by gender and nutritional status with new treatment and control groups

	Weight		
	Migrant	Remittances	Migrant and remittances
	(1)	(2)	(3)
Panel A - All adults			
DID	-1.346*	-0.862	0.990
	(0.699)	(0.543)	(0.804)
Mean weight at baseline	63.333	62.951	63.207
R^2	0.138	0.166	0.186
Observations	4,922	8,188	4,004
Panel B - Males			
DID	-1.929*	-0.658	-1.921*
	(0.992)	(0.687)	(0.987)
Mean weight at baseline	64.225	63.650	63.966
R^2	0.118	0.146	0.205
Observations	2,322	3,796	1,436
Panel C - Females			
DID	-0.492	-0.828	-0.472
	(0.998)	(0.796)	(1.187)
Mean weight at baseline	62.415	62.348	62.442
R^2	0.194	0.196	0.241
Observations	2,392	4,362	2,146
Panel D - Underweight adults			
DID	-0.530	-0.726	-0.606
	(1.610)	(0.989)	(2.860)
Mean weight at baseline	47.545	46.728	47.537
R^2	0.510	0.359	0.516
Observations	260	756	86
Panel E - Healthy adults			
DID	-1.514***	-1.149**	-2.068***
	(0.577)	(0.464)	(0.615)
Mean weight at baseline	59.784	59.165	59.642
R^2	0.245	0.196	0.180
Observations	2,982	4,868	2,672
Panel F - Overweight adults			
DID	1.057	-0.120	-0.615
	(1.689)	(1.065)	(2.074)
Mean weight at baseline	75.557	75.316	75.480
R^2	0.207	0.128	0.260
Observations	1,386	2,506	886
Individual variables	Yes	Yes	Yes
Household head variables	Yes	Yes	Yes
Household variables	Yes	Yes	Yes
Rural dummy	Yes	Yes	Yes
Region dummies	Yes	Yes	Yes
Number of entries and exits	Yes	Yes	Yes

Notes: Standard errors are in parentheses. Individual variables include age, gender, and relationship to the head. Household head variables include age, gender, and education level of the head. Household variables include household size, number of working-age members, dependency ratio, and wealth index quintiles. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table 3.K.2: Results from PSM-DID for children, by gender and nutritional status with new treatment and control groups

	Zbmi		
	Migrant	Remittances	Migrant and remittances
	(1)	(2)	(3)
Panel A - All children			
DID	-0.191	0.153*	-0.453***
	(0.118)	(0.092)	(0.127)
Mean zbmi at baseline	-0.050	-0.102	-0.046
R^2	0.086	0.093	0.116
Observations	3,116	4,866	2,816
Panel B - Males			
DID	-0.071	0.028	-1.077***
	(0.204)	(0.126)	(0.159)
Mean zbmi at baseline	-0.094	-0.122	-0.082
R^2	0.119	0.114	0.265
Observations	1,146	2,724	1,866
Panel C - Females			
DID	-0.388**	0.209	-0.088
	(0.174)	(0.137)	(0.230)
Mean zbmi at baseline	0.011	-0.077	0.004
R^2	0.081	0.076	0.215
Observations	1,242	2,136	730
Panel D - Underweight children			
DID	0.221	0.579***	-0.783
	(0.477)	(0.206)	(0.719)
Mean zbmi at baseline	-2.968	-2.950	-2.983
R^2	0.681	0.565	0.675
Observations	112	580	88
Panel E - Healthy children			
DID	-0.178	0.104	-0.429***
	(0.124)	(0.085)	(0.114)
Mean zbmi at baseline	-0.385	-0.429	-0.393
R^2	0.061	0.074	0.110
Observations	1,610	3,070	1,594
Panel F - Overweight children			
DID	-1.314***	-0.342**	-0.367
	(0.267)	(0.160)	(0.228)
Mean zbmi at baseline	2.285	2.277	2.280
R^2	0.511	0.389	0.507
Observations	462	1,140	618
Individual variables	Yes	Yes	Yes
Household head variables	Yes	Yes	Yes
Household variables	Yes	Yes	Yes
Rural dummy	Yes	Yes	Yes
Region dummies	Yes	Yes	Yes
Number of entries and exits	Yes	Yes	Yes

Notes: Standard errors are in parentheses. Individual variables include age, gender, and relationship to the head. Household head variables include age, gender, and education level of the head. Household variables include household size, number of working-age members, dependency ratio, and wealth index quintiles. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

3.L Profile of the migrants

Table 3.L.1: Selected characteristics of the migrants, left behind and non-left behind

	Migrants	Left Behind	Non Left Behind	(1)-(2)	(1)-(3)
Less than 14 years old	0.155	0.386	0.361	-0.231***	-0.205***
	(0.020)	(0.015)	(0.005)	(0.029)	(0.027)
15-24 years old	0.646	0.125	0.120	0.521***	0.526***
	(0.027)	(0.010)	(0.003)	(0.024)	(0.019)
25-34 years old	0.158	0.060	0.128	0.099***	0.031
	(0.020)	(0.007)	(0.003)	(0.018)	(0.019)
35-44 years old	0.025	0.113	0.135	-0.088***	-0.110***
	(0.009)	(0.010)	(0.003)	(0.018)	(0.019)
45-64 years old	0.012	0.246	0.180	-0.234***	-0.168***
	(0.006)	(0.014)	(0.004)	(0.024)	(0.021)
Over 65 years old	0.003	0.071	0.076	-0.068***	-0.073***
	(0.003)	(0.008)	(0.003)	(0.014)	(0.015)
Male	0.599	0.473	0.486	0.127***	0.113***
	(0.027)	(0.016)	(0.005)	(0.032)	(0.028)
Head	0.040	0.247	0.348	-0.207***	-0.308***
	(0.011)	(0.014)	(0.005)	(0.025)	(0.027)
Spouse of the head	0.022	0.189	0.159	-0.168***	-0.137***
	(0.008)	(0.012)	(0.004)	(0.022)	(0.020)
Child of the head	0.780	0.462	0.422	0.318***	0.358***
	(0.023)	(0.016)	(0.005)	(0.031)	(0.028)
Grandchild of the head	0.084	0.065	0.040	0.019	0.043***
	(0.015)	(0.008)	(0.002)	(0.016)	(0.011)
Other relationship to the head	0.075	0.037	0.031	0.038***	0.044***
	(0.015)	(0.006)	(0.002)	(0.013)	(0.010)
None or preschool	0.181	0.522	0.435	-0.341***	-0.253***
	(0.022)	(0.016)	(0.005)	(0.030)	(0.028)
Primary education	0.241	0.249	0.247	-0.008	-0.006
	(0.024)	(0.014)	(0.004)	(0.028)	(0.024)
Post-primary education	0.350	0.172	0.220	0.178***	0.130***
	(0.027)	(0.012)	(0.004)	(0.026)	(0.024)
Secondary education	0.178	0.035	0.065	0.143***	0.113***
	(0.021)	(0.006)	(0.002)	(0.016)	(0.014)
Tertiary education	0.050	0.022	0.033	0.028***	0.017
	(0.012)	(0.005)	(0.002)	(0.011)	(0.010)
Married	0.053	0.360	0.341	-0.307***	-0.289***
	(0.012)	(0.015)	(0.005)	(0.028)	(0.027)
Rural	0.776	0.805	0.630	-0.028	0.147***
	(0.023)	(0.013)	(0.005)	(0.026)	(0.027)
Days of work per week[a]	3.480	4.647	4.491	-1.167***	-1.011***
	(0.196)	(0.096)	(0.029)	(0.200)	(0.180)
Paid employed[b]	0.076	0.051	0.100	0.025*	-0.023
	(0.015)	(0.008)	(0.003)	(0.015)	(0.017)
Owner of non-farm enterprise[b]	0.044	0.155	0.190	-0.110***	-0.146***
	(0.012)	(0.012)	(0.004)	(0.022)	(0.022)
Worker in non-farm enterprise[b]	0.083	0.144	0.181	-0.062***	-0.099***
	(0.016)	(0.012)	(0.004)	(0.022)	(0.022)
Owner of farm-plot[b]	0.054	0.281	0.268	-0.227***	-0.214***
	(0.013)	(0.015)	(0.005)	(0.027)	(0.025)
Worker in farm-plot[b]	0.397	0.519	0.396	-0.122***	0.001
	(0.028)	(0.017)	(0.005)	(0.033)	(0.028)
Full-time student[b]	0.514	0.332	0.298	0.182***	0.216***
	(0.028)	(0.016)	(0.005)	(0.032)	(0.026)
Retired or ill[b]	0.000	0.039	0.049	-0.039***	-0.049***
	(0.000)	(0.007)	(0.002)	(0.011)	(0.012)
Full-time homemaker[b]	0.051	0.156	0.150	-0.105***	-0.100***
	(0.012)	(0.012)	(0.004)	(0.022)	(0.020)
Looking for work[b]	0.171	0.078	0.098	0.093***	0.074***
	(0.021)	(0.009)	(0.003)	(0.020)	(0.017)
Moved to Accra region	0.358	0.000	0.000	0.358***	0.358***
	(0.027)	(0.000)	(0.000)	(0.015)	(0.005)
Moved to Ashanti region	0.218	0.000	0.000	0.218***	0.218***
	(0.024)	(0.000)	(0.000)	(0.013)	(0.004)
Observations	322	1,003	10,815	1,325	11,137

Notes: Migrants are those who have migrated for work for more than 6 months.
[a] Information only available for the household head, the first spouse, and one other household member over the age of 12, selected randomly.
[b] The categories related to an individual's activity are not mutually exclusive, meaning an individual can belong to multiple categories. For example, an individual can be a full-time student while also contributing to a household farm. Information only available for individuals from age 7 and above.

3.M Parental migration

In the surveys, I use information about co-residence with parents. I use the following questions: "Is [Name]'s father currently living in this household?" and "Is [Name]'s mother currently living in this household?". The possible answers to these questions are: "Yes", "No, he/she is deceased", and "No, he/she lives in another household".

3.M.1 Parental migration for work

Based on co-residence with the parents, I create a treatment group of children who were co-resident with at least one parent in Wave 2 and are left behind by at least one parent who had migrated away for more than six months to find work between the two survey waves by using the previous treatment. The control group consists of children who were still co-residing with at least one parent in Wave 3 but did not have at least one parent who had migrated away for more than six months to find work between the two survey waves. I drop children who had lost both parents (double orphans) between the two waves. The sample is composed of 2,478 children, of which 19 are left behind.

Table 3.M.1: Results from PSM-DID for children, parental migration for work

	Zbmi			
	(1)	(2)	(3)	(4)
Parental migration*Post	0.111	-0.561***	-0.491***	-0.489***
	(0.197)	(0.199)	(0.188)	(0.187)
Mean zbmi at baseline	-0.061	-0.061	-0.061	-0.061
R^2	0.001	0.228	0.327	0.332
Observations	1,806	1,806	1,806	1,806
Individual variables	No	Yes	Yes	Yes
Household head variables	No	Yes	Yes	Yes
Household variables	No	Yes	Yes	Yes
Rural dummy	No	No	Yes	Yes
Region dummies	No	No	Yes	Yes
Number of entries and exits	No	No	No	Yes

Notes: Standard errors are in parentheses. The variables listed at the bottom of the table are included in the DID, i.e., after matching. Individual variables include age, gender, and relationship to the head. Household head variables include age, gender, and education level of the head. Household variables include household size, number of working-age members, dependency ratio, and wealth index quintiles. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

3.M.2 Parental migration based on non-co-residence

To obtain an alternative measure of parental migration, I also create two other groups, no longer relying on the treatment I previously completed based on migration for job-related reasons for more than six months. This time, it is solely based on co-residence. This treatment is, therefore, less restrictive, as it relies solely on co-residence to define parental migration, which is effectively a parental departure from the household. The treatment group consists of children who were co-residents with at least one of their parents in Wave 2 but are no longer in Wave 3. The control group, on the other hand, comprises children who remain co-residents with at least one of their parents, and no parent has left the household between the two waves, regardless of the reason. Similar to the approach in section 3.M.1, I exclude double orphans. The sample is composed of 2,532 children, of which 145 are left behind.

Table 3.M.2: Results from PSM-DID for children, parental migration related to non-co-residence

	Zbmi			
	(1)	(2)	(3)	(4)
Parental migration*Post	-0.205**	-0.178*	-0.193**	-0.163*
	(0.098)	(0.097)	(0.095)	(0.097)
Mean zbmi at baseline	-0.071	-0.071	-0.071	-0.071
R^2	0.003	0.070	0.112	0.113
Observations	4,888	4,888	4,888	4,888
Individual variables	No	Yes	Yes	Yes
Household head variables	No	Yes	Yes	Yes
Household variables	No	Yes	Yes	Yes
Rural dummy	No	No	Yes	Yes
Region dummies	No	No	Yes	Yes
Number of entries and exits	No	No	No	Yes

Notes: Standard errors are in parentheses. The variables listed at the bottom of the table are included in the DID, i.e., after matching. Individual variables include age, gender, and relationship to the head. Household head variables include age, gender, and education level of the head. Household variables include household size, number of working-age members, dependency ratio, and wealth index quintiles. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

3.N Food consumption patterns

In the questionnaire, I use the questions in Section 11: Consumption module, Part A: Food items consumed. More specifically, I use the value reported by the most knowledgeable household member regarding the value of the quantity of food items (in Ghanaian cedis), whether the food items are sourced from purchases, self-production, or gifts. In this section, I exclude from the sample households that likely have reporting issues. For instance, some households report consuming only a single type of food, such as exclusively cooking oils or only beverages. Considering the unlikely nature of that occurrence, these households are omitted, along with 0.10% of the outliers regarding the share of consumed food items. Also, consistent with the data treatment, I chose a complete case analysis, thus excluding observations with missing outcome values for at least one wave.

First, I investigate the impact of migration on the household's food consumption over the last 30 days, including food obtained through purchases (Column 1 of Table 3.N.1), self-produced food (Column 2 of Table 3.N.1), food received as gifts (Column 3 of Table 3.N.1), as well as the total household consumption (Column 4 of Table 3.N.1). All outcomes are adjusted for household size, yielding per capita outcomes to account for variations in household composition. According to the results, there is a significant and negative effect on food consumed from purchases. In other words, when at least one household member migrates during the survey waves, the household's purchase of food items diminishes. This is consistent with the hypothesized negative impact on household income; indeed, facing a negative income shock from migration, individuals reduce food purchases, which could elucidate the negative effect observed on children's anthropometry. A significant and positive result is noted for the per capita consumption of self-produced food, although not that significant. Post-migration, households appear to increase their food production, possibly as a coping strategy to become more self-reliant, perhaps in response to the income loss from the migrant's absence. However, this increase is not proportional to the observed decrease in purchased food items.

Subsequently, I explore the effect of migration on per capita consumption across various food groups, considering food items that are purchased, received, and produced by the household (Table 3.N.3). Furthermore, I assess how migration affects the composition of the household's total food consumption in terms of the share of different food groups (Table 3.N.4). In summary, migration leads to a reduction in the consumption of fruits and vegetables, whether in terms of total consumption in Cedis (Column 4 of Table 3.N.3) or their share of the household's total food consumption (Column 4 of Table 3.N.4). This may suggest a decrease in the intake of often costlier or less accessible foods, in line with the short-term negative income shock. Such a trend could indicate a decline in dietary quality, particularly if the reduced food groups are those providing essential nutrients such as fruits and vegetables, as well as eggs (Column 5 of Table 3.N.3), potentially explaining the decline in children's z-scores. Moreover, beverages (Column 12 of Table 3.N.3) may be deemed non-essential during an income shock, leading households to curtail their consumption. Finally, according to Column (10) of Table 3.N.4, there is a noticeable increase in the share of sugary foods following migration. If viewed through the lens of an income shock, households may prioritize calorie-dense foods over nutrition to satisfy immediate hunger. Sweetened foods, often energy-dense, provide a quick satiety at a lower cost than more nutritious but expensive options.

Finally, I have constructed two additional measures to assess household food diversity: the Simpson and Shannon indices. The Simpson index (Equation 5 below), bounded between zero and one, indicates

that a higher value corresponds to greater dietary diversity. Conversely, the Shannon index (Equation 6 below) quantifies the concentration of food group consumption, assigning lower weights to subgroups with a larger share of food expenditure and higher weights to those with a smaller share. This index spans from zero to the natural logarithm of the total number of food groups, reaching its apex when expenditure is evenly distributed across all subgroups.
The Simpson index can be expressed as follows:

$$Simpson\ index = 1 - \sum_{i=1}^{n} w_i^2 \tag{1}$$

with w_i the consumption share (purchased, self-produced, and received) of food group i (cereals, starches, pulses and nuts, fruits and vegetables, eggs, fish, meat, dairy products, cooking oils, sugary products, spices, beverages, wild food, and out-of-home food).

The Shannon index is defined as:

$$Shannon\ index = -\sum_{i=1}^{n} w_i log(w_i) \tag{2}$$

where w_i is defined as previously.
Table 3.N.2 reveals a negative and significant impact on the Simpson index, suggesting that migration tends to reduce dietary diversity. This may imply that left-behind households are likely to consume a less varied array of food items, which could affect the nutritional status of those left behind. This finding aligns with the disruptive effect and the assumption of a negative income shock. However, no significant effect is noted on the Shannon index.
Nevertheless, despite the insight gained from analyzing food consumption patterns, some limitations hinder complete confidence in these results. First, the recall period for querying a household member about consumed food items is 30 days. This extended recall period may be prone to errors: individuals might not accurately recall what the household consumed, or, concerning out-of-home food items, it may be challenging for one member to account for what all others ate, especially over the past 30 days, and so forth. Moreover, these measures are at the household, not the individual level. Hence, we lack insight into the distribution of food consumption within the household, particularly between adults and children. Given that our main finding pertains to the negative effect on children's nutritional status, the unknown intra-household distribution remains a critical gap. For these reasons, while the findings are intriguing, they should be interpreted cautiously.

Table 3.N.1: Results from PSM-DID, food consumed per capita from various sources

	Purchased food per capita	Produced food per capita	Received food per capita	Total food consumed per capita
	(1)	(2)	(3)	(4)
Household left-behind*Post	-7.375**	3.868*	-0.992	-4.498
	(3.353)	(2.079)	(1.118)	(4.409)
Mean at baseline	86.512	27.120	9.730	123.389
R^2	0.350	0.090	0.070	0.270
Observations	5,642	5,642	5,642	5,642
Household head variables	Yes	Yes	Yes	Yes
Household variables	Yes	Yes	Yes	Yes
Rural dummy	Yes	Yes	Yes	Yes
Region dummies	Yes	Yes	Yes	Yes
Number of entries and exits	Yes	Yes	Yes	Yes

Notes: Standard errors are in parentheses. The variables listed at the bottom of the table are included in the DID, i.e., after matching. Household head variables include age, gender, and education level of the head. Household variables include number of working-age members, dependency ratio, and wealth index quintiles. Number of entries and exits are two separate variables. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table 3.N.2: Results from PSM-DID, dietary diversity indices

	Simpson index	Shannon index
	(1)	(2)
Household left-behind*Post	-0.008**	-0.017
	(0.004)	(0.013)
Mean at baseline	0.779	1.843
R^2	0.150	0.220
Observations	6,172	6,172
Household head variables	Yes	Yes
Household variables	Yes	Yes
Rural dummy	Yes	Yes
Region dummies	Yes	Yes
Number of entries and exits	Yes	Yes

Notes: Standard errors are in parentheses. The variables listed at the bottom of the table are included in the DID, i.e., after matching. Household head variables include age, gender, and education level of the head. Household variables include household size, number of working-age members, dependency ratio, and wealth index quintiles. Number of entries and exits are two separate variables. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table 3.N.3: Results from PSM-DID, food consumed per capita and per food group

	Total food consumed (purchased, produced & received) per food group													
	Cereals	Starches	Pulses & Nuts	Fruits & Vegetables	Eggs	Fish	Meat	Dairy	Oil	Sugar	Spices	Beverages	Wild	Out-of -home
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Household left-behind*Post	-0.018	0.217	1.651	-2.301***	-0.268**	-0.367	-0.951	-0.108	-0.316	0.123	0.059	-0.600*	-0.216	-1.400
	(1.002)	(0.986)	(1.071)	(0.836)	(0.105)	(0.736)	(0.784)	(0.204)	(0.250)	(0.119)	(0.111)	(0.318)	(0.157)	(1.644)
Mean at baseline	13.395	14.017	18.689	17.676	1.455	13.712	9.068	2.440	4.047	1.584	1.365	3.191	1.197	21.552
R^2	0.100	0.180	0.090	0.280	0.200	0.240	0.090	0.170	0.190	0.080	0.150	0.210	0.050	0.070
Observations	5,642	5,642	5,642	5,642	5,642	5,642	5,642	5,642	5,642	5,642	5,642	5,642	5,642	5,642
Household head variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Household variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Rural dummy	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of entries and exits	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Standard errors are in parentheses. The variables listed at the bottom of the table are included in the DID, i.e., after matching. Household head variables include age, gender, and education level of the head. Household variables include household size, number of working-age members, dependency ratio, and wealth index quintiles. Number of entries and exits are two separate variables. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table 3.N.4: Results from PSM-DID, share of food group's consumption out of total household consumption

	Share of total food consumed (purchased, produced & received) per food group													
	Cereals	Starches	Pulses & Nuts	Vegetables & Fruits	Eggs	Fish	Meat	Dairy	Oil	Sugar	Spices	Beverages	Wild	Out-of -home
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Household left-behind*Post	0.005	0.005	-0.001	-0.010**	-0.001	-0.000	-0.003	-0.000	0.003	0.002**	0.001	-0.002	-0.001	0.002
	(0.006)	(0.005)	(0.005)	(0.002)	(0.001)	(0.004)	(0.005)	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)	(0.006)
Mean at baseline	0.131	0.133	0.160	0.157	0.012	0.114	0.074	0.017	0.038	0.014	0.013	0.021	0.011	0.105
R^2	0.440	0.029	0.100	0.070	0.090	0.170	0.090	0.140	0.050	0.007	0.140	0.180	0.110	0.100
Observations	5,424	5,424	5,424	5,424	5,424	5,424	5,424	5,424	5,424	5,424	5,424	5,424	5,424	5,424
Household head variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Household variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Rural dummy	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of entries and exits	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Standard errors are in parentheses. The variables listed at the bottom of the table are included in the DID, i.e., after matching. Household head variables include age, gender, and education level of the head. Household variables include household size, number of working-age members, dependency ratio, and wealth index quintiles. Number of entries and exits are two separate variables. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Chapter 4

Child Fostering and Nutrition in South Africa

This chapter is a joint work with Christelle Dumas (Professor, University of Fribourg) and Elsa Gautrain (Postdoctoral researcher, University of Fribourg).

4.1 Introduction

Children's living arrangements and family structures are fundamental determinants of their health and developmental outcomes, shaping both immediate well-being and long-term trajectories. While co-residence with parents is often associated with better health outcomes (Bledsoe and Isiugo-Abanihe, 1989; Castle, 1995; Madhavan and Townsend, 2007; Bramlett and Blumberg, 2007; Prall and Scelza, 2017), in low- and middle-income countries, the extended family also plays a crucial role in childcare and social protection (Fafchamps and Quisumbing, 2008). This challenges the assumption that living apart from parents is inherently detrimental. In this paper, we investigate the nutritional outcomes of South African children who are sent away from their biological parents to live with new caregivers, often within the extended family. We develop an empirical framework to assess whether child fostering improves or hinders child nutrition. Although nutrition represents only one dimension of child well-being, focusing on this aspect is essential due to its role as a key determinant of long-term health and human capital development (Hoddinott et al., 2013; McGovern et al., 2017).

South Africa provides a compelling context to explore the relationship between child fostering and health outcomes. Child fostering is particularly widespread, with 26% of mothers reporting at least one child under 15 living outside their household in 2016 (Cotton, 2021). Simultaneously, the country faces persistent malnutrition. Approximately one in four children under the age of five are stunted, a prevalence that has remained essentially unchanged over the past decades. Concurrently, rates of overweight and obesity are rising, making it one of the most affected countries in sub-Saharan Africa (Shekar and Popkin, 2020). This dual challenge reflects the ongoing struggle with the double burden of malnutrition (Azomahou et al., 2022). This combination of widespread child fostering and persistent nutritional challenges underscores the relevance of investigating the effects of this specific living arrangement on children's nutritional status.

We draw on data from the National Income Dynamics Study (NIDS), a nationally representative longitudinal survey in South Africa. The survey includes detailed child tracking, enabling us to observe child out-fostering and examine children's characteristics both before and after placement. Our analysis focuses specifically on children who leave their parents' household to join a new one, where they are cared for by a nonparental caregiver. This precise focus on child-out fostering represents a novel contribution to the existing literature, which often examines fostering more broadly. Pioneering work by Isiugo-Abanihe (1985) in demographic studies conceptualizes child fostering as a form of relocation, where a child moves from their biological home to that of a foster caregiver. This notion of relocation is reflected in the widely used terminology of "sending" and "receiving" households, which describes the dynamics of child-out fostering in related studies. In this respect, we refer to foster children as those who change households and acknowledge that this definition excludes children who are left behind by their parents in their household of origin.

Establishing the causal impact of child fostering on nutrition requires addressing key methodological challenges. Our first contribution lies in implementing an empirical approach that accounts for the pre-fostering characteristics of the child and their sending household. Indeed, child fostering is not a random practice; it is shaped by various factors, including parental marital status (Grant and Yeatman, 2014), social norms (Marazyan, 2015), the number of siblings (Cotton, 2024), as well as the child's age and gender, which could create endogeneity issues when assessing the causal impact of fostering. To the best of our knowledge, few large-scale, representative surveys provide information on household and child characteristics prior to placement (Akresh, 2009; Beck et al., 2015). As a result, most existing studies rely on cross-sectional data, which capture the child's situation only after integration into the host household (see Ariyo et al., 2019, for a review in the African context). Only a limited number of papers tackle this identification challenge.[1] To address these challenges, we rely on two key elements for identification: first, a comprehensive set of pre-fostering characteristics, and second, machine learning techniques, which efficiently handle high-dimensional covariates and capture complex interactions among variables. This approach minimizes the risk of misspecification and strengthens the validity of our estimates.

The second contribution of this paper is to address attrition issues, a common challenge in longitudinal surveys. To achieve this, we take advantage of the NIDS-specific survey design, which tracks a targeted group of individuals who change residence between survey rounds, referred to as Continuing Sample Members (CSMs). Tracking efforts are concentrated on this group, resulting in a particularly low attrition rate. As the proportion of CSMs in a given area increases, tracking becomes more complex, thereby reducing the likelihood of resurveying individuals of any type. These two exogenous factors—CSM status and the proportion of CSMs in an area—explain the probability of being resurveyed and enable us to incorporate an instrumental variable strategy into our empirical framework.

We leverage these features within a Double Machine Learning (DML) estimator developed by Bia et al. (2023) to estimate the causal impact of fostering on children's nutritional status. This represents our third contribution. The DML approach simultaneously addresses both selection into fostering and the likelihood of being observed in the subsequent survey round. More broadly, we introduce the DML methodology to the literature, providing robust results in contexts with double selection issues. Specifically, the estimator calculates two key propensity scores: one for the probability of being observed in the following survey round and another for the probability of being fostered. Outcomes are then estimated for treated and control groups, conditional on covariates and inversely weighted by these propensity scores. Moreover, the estimator is doubly robust, providing unbiased results as long as either the conditional mean outcome model or the propensity score model is correctly specified (Chernozhukov et al., 2018).

Building on this framework, our findings indicate that fostering improves the nutritional status of children under 15. Specifically, we find that it reduces the probability of stunting by 7 percentage points over 2-3 years, corresponding to a 45 percent decline relative to the baseline prevalence. The primary mechanism underlying these improvements is the relocation of foster children to smaller households, often cared for by grandparents who are likely to receive pensions.[2] Fostering reshapes household dynamics for siblings who stay with their parents, as it entails the departure of one child from the household. By leveraging the ability to track both sending and receiving households, we also examine the indirect effects of fostering on children's nutrition in sending households. Our analysis reveals that having a sibling fostered reduces the likelihood of stunting for these children as well, primarily through adjustments in household size. Overall, these findings suggest that child-out fostering can function as a mutually beneficial arrangement, enhancing nutritional outcomes for both foster children and their siblings in the household of origin.

The remainder of the paper is structured as follows. Section 4.2 discusses the potential effects of fostering on nutrition, based on prior literature. Section 4.3 introduces the data, sample characteristics, and key variables, complemented by descriptive statistics. In Section 4.4, the empirical strategy is presented.

[1] For instance, Cichello (2003), Akresh (2004), and Bose-Duker et al. (2021) implement a household fixed effects strategy; Badaoui and Mangiavacchi (2022) utilize an instrumental variable approach; and Beck et al. (2015) use retrospective questions to account for the pre-fostering conditions.

[2] These findings are consistent with previous literature on the impacts of child fostering and the role of grandparents in child development (Bledsoe et al., 1988; Zimmerman, 2003; Duflo, 2003; Serra, 2009; Alber, 2013; Hampshire et al., 2015).

Section 4.5 discusses the results, including robustness checks and sensitivity analyses, and examines the underlying mechanisms. Section 4.6 presents the results for children in the sending households. Finally, Section 4.7 concludes.

4.2 Conceptual framework

This section develops a conceptual framework to understand how fostering influences children's nutrition. The analysis begins with a simplified scenario and progressively incorporates additional complexities.

Child fostering results from several strategic decisions made by both sending and receiving households. Sending households decide whether to foster a child and, if so, which child, when, and where. Receiving households determine their willingness and capacity to host a child and may even request one (Bledsoe and Isiugo-Abanihe, 1989). Thus, the arrangement must be mutually acceptable to both households, impacting not only the foster child but also all households involved.

Foster care involves transferring a child from one household (A) to another (B). To simplify, assume that households A and B initially possess similar resources. After fostering, household A retains fewer children, while household B gains an additional member. This shift reallocates the burden of childcare and resources between the two households. If total household resources remain unchanged, children in household A may experience higher per capita consumption, while those in household B may face reduced resources due to increased competition.

Empirical evidence suggests that foster children are often sent to households with greater resources or opportunities than their households of origin. For instance, Eloundou-Enyegue and Shapiro (2004) have shown that fostering can serve as a mechanism to reduce inequalities among kin by transferring children from poorer to wealthier households. This reflects a strategic decision by sending households to improve the child's living conditions, access to education, or other developmental prospects (Zimmerman, 2003; Akresh, 2004; Beck et al., 2015; Badaoui and Mangiavacchi, 2022). Fostering may also be used to mitigate the effects of adverse events, such as income shocks (Akresh, 2009), climatic shocks (Ronnkvist et al., 2023), excess fertility (Cotton, 2024), or parental divorce (Grant and Yeatman, 2014).

Fostering arrangements may also involve inter-household transfers (Marazyan, 2011; Gautrain, 2023). Sending households may compensate receiving households for fostering, potentially with resources intended for the child's direct benefit. However, the extent to which these transfers improve the child's well-being depends on how they are utilized. If the transfer equals the resources previously allocated to the child in household A, the net impact on both households may be neutral.

However, the relationship between resources and nutritional status may not be monotonic. While an increase in per capita consumption can improve the nutritional status of individuals, the quality of the diet also matters and may depend on the environment. For instance, a child fostered into a household in a region with lower living costs may have better access to food at a given nominal expenditure. Geographic context further shapes nutritional outcomes. Fostering often involves relocating children to urban areas (Isiugo-Abanihe, 1985; Eloundou-Enyegue and Shapiro, 2004; Serra, 2009). Urban settings may provide greater dietary diversity but also increase exposure to processed foods, which can negatively impact nutrition (Popkin, 1999, 2001; Hawkes, 2008). Conversely, rural areas may offer access to fresh, locally produced foods that may provide higher-quality nutrition but remain vulnerable to agricultural shocks that can disrupt food availability (Reardon et al., 2021). Thus, fostering across rural and urban areas may also affect both the quality and quantity of children's diets.

Intra-household allocation adds another layer of complexity. Within households, resources are often distributed unevenly, with foster children potentially receiving less favorable treatment than other children. The role of the caregiver is pivotal in these choices. Caregivers' allocation decisions often reflect a combination of personal motivations, cultural norms, and economic constraints. Biological connections can shape these behaviors, as parents tend to favor their biological children regarding resource allocation (Hamilton, 1964). This can create disparities: foster children may be treated less favorably than biological children. For instance, Prall and Scelza (2017) find that among pastoralists in Namibia, foster children have worse nutritional outcomes compared to children living with their parents. Similarly, in South Africa, children living with maternal female kin have double the likelihood of nutritional deficiencies (Madhavan and Townsend, 2007). Conversely, suppose foster parents perceive fostering as an opportunity to strengthen social ties. In that case, foster parents could be expected not to discriminate against the foster child to preserve their long-term relationship with their network. Furthermore, caregiving practices can vary widely depending on the caregiver's expectations about the child's future contributions. For instance, when the new caregivers are grandparents, particularly grandmothers, which is common in many fostering arrangements (Gautrain, 2023), studies have shown that living with grandmothers can positively impact nutritional outcomes (Schrijner and Smits, 2018). These factors underscore the importance of understanding the caregiver's role in shaping fostering outcomes.

Finally, a last aspect to consider lies in the reasons for fostering, which may not always align with foster children's best interests. In cases where children are fostered for labor rather than developmental purposes, their well-being may be jeopardized. Evidence suggests that children fostered under forced circumstances often face poor outcomes, including lower nutritional status (Castle, 1995). For example, foster children might face greater work demands than biological children in the receiving household (B) or siblings remaining in the sending household (A). However, disentangling the effects of different motives is challenging, as several motives often coexist (Serra, 2009).

In summary, fostering may have varied implications for foster children and both households involved. While it can enhance opportunities by providing access to better resources or education, it may also introduce risks related to resource competition and caregiver intent. This paper aims to provide an average effect of being fostered on nutrition and assess the underlying mechanisms.

4.3 Data

4.3.1 National Income Dynamics Study (NIDS)

We use data from the National Income Dynamics Study (NIDS), collected by the Southern Africa Labour and Development Research Unit (SALDRU) at the University of Cape Town.[3] NIDS explores the living conditions of individuals, addressing aspects such as economic activity, poverty, labor market participation, education, and health. This longitudinal survey has been conducted every two to three years since 2008, with the most recent wave covering 2017. It tracks 28,000 individuals residing in 7,305 households as of the first round. The sample was representative of the national population in 2008, and a two-stage cluster sampling design was employed.

The survey encompasses two types of interviewees: Continuing Sample Members (CSMs) and Temporary Sample Members (TSMs). For the CSMs, the survey design induces individual tracking regardless of

[3] The NIDS portal is available at: `https://www.datafirst.uct.ac.za/dataportal/index.php/catalog/NIDS`. Last accessed: 03 June 2024.

whether they change households between two survey waves. The CSMs are either from the initial surveyed sample in 2008 or children born to female CSMs in subsequent waves. When such a person moves outside her household within the boundaries of South Africa, she can be followed to her new household, which will then be surveyed. TSMs are co-residents of a CSM during a survey wave. The individual tracking does not apply to these individuals, and they are only surveyed if they still reside in a household with a CSM in subsequent waves.[4]

This tracking system is a notable feature of the NIDS, which is particularly valuable in the context of child fostering and rare in longitudinal household surveys.[5] Although not all CSMs are successfully followed, the distinction between CSM and TSM in tracking is clear, and we leverage this feature of the survey in our empirical framework.[6]

4.3.2 Sample

We focus on the last two survey rounds of the NIDS: Wave 4 (2014/2015) and Wave 5 (2017).[7] This decision is based on several considerations. Firstly, these rounds provide the most recent perspective on child fostering in South Africa. The second reason concerns data constraints related to tracking and missing data. The pair of waves 4 and 5 has a high re-interview rate, along with the lowest rate of missing anthropometric data.[8] Finally, the prevalence of nutritional statuses, particularly stunting rates among children under five, confirms that the sample from waves 4 and 5 aligns with external data sources.[9]

In our analysis, we focus on children who were aged 0 to 12 years during Wave 4. This age range is selected to ensure that these individuals are still considered children in Wave 5 (under age 15), which is essential for a study on fostering where children need to remain young enough to be considered dependent.[10]

4.3.3 Children's nutritional status

To assess the impact of fostering on children's nutritional outcomes, we rely on anthropometric indicators, specifically z-scores derived from NIDS data. Using BMI-for-age and height-for-age z-scores, we define three nutritional statuses (WHO, 2006):

- healthy, defined as BMI-for-age between -2 standard deviation (SD) and +1 SD;
- overweight/obese, defined as BMI-for-age $\geq +1$ SD;
- and stunted, defined as height-for-age ≤ -2 SD.

[4] The survey process is detailed in Section 4.1.2 of the Panel User Manual (Brophy et al., 2018).

[5] This feature also exists in the Indonesian Family Life Survey and the Mexican Family Life Survey.

[6] Among individuals surveyed in Wave 4 but not interviewed in Wave 5, 70.45% are TSMs and 29.55% are CSMs. Conversely, 8.28% of CSMs interviewed in Wave 4 are not interviewed in Wave 5, whereas 41.01% of TSMs interviewed in Wave 4 are not interviewed in Wave 5.

[7] Wave 3 will also be used for placebo tests.

[8] Re-interview rates are the following: 78.86% of individuals between waves 1 and 2, 82.77% between waves 2 and 3, 81.50% between waves 3 and 4, and 79.56% between waves 4 and 5. Anthropometric data (height-for-age for children) are missing at the following rates: 25.03% in Wave 1, 45.93% in Wave 2, 20.02% in Wave 3, 12.64% in Wave 4, and 15.81% in Wave 5.

[9] In Wave 4 (2014-2015), 21.71% of children were stunted, and in Wave 5 (2017), 20.46%. These figures align with external data; UNICEF, WHO, and WB Joint Child Malnutrition Estimates reported 22.3% in 2015 and 22.4% in 2017. Data: https://www.who.int/data/gho/data/indicators/indicator-details/GHO/gho-jme-country-children-aged-5-years-stunted-(-height-for-age--2-sd). Last accessed: 24 June 2024.

[10] In many studies, children aged 2 years or younger are often excluded due to the higher likelihood of measurement errors in length data for this age group (WHO and UNICEF, 2019). Therefore, as a robustness check, we restrict the sample to the 2-12 years in Wave 4 (results can be found in Appendix Table 4.C.4.

For each status, we define binary outcomes equal to 1 if the child falls into the status and 0 otherwise. Stunting reflects chronic malnutrition, capturing medium- to long-term effects of undernutrition. Its consequences may appear rapidly, including impaired cognitive, motor, and language development, alongside increased morbidity and mortality (Stewart et al., 2013). The other categories, healthy and overweight/obesity, derived from BMI-for-age z-scores, provide insights into dietary or lifestyle changes, which may, for instance, occur when a child moves to a household with different eating habits. Note that the category healthy is not the exact counterpart to overweight/obese; the gap includes underweight children, who are excluded from the categories. While studying the underweight status would have been interesting, we have too few observations concerned to achieve convergence with our estimation method. Moreover, although undernutrition is not negligible in South Africa, its prevalence is lower relative to stunting or overweight/obesity among children (Kruger et al., 2023). Overall, these three measures are complementary in assessing the impacts of fostering on children's nutrition.

4.3.4 Defining fostering and groups of children

We classify children using information on their co-residence with parents. Our sample comprises children who, in Wave 4, are not fostered. Since it is rare for children to live solely with their father if the mother is alive, we focus on children living with their mother in Wave 4 to determine whether the child is included in the sample.[11] We exclude children who become orphans during the period. As a result, our sample is composed of children who live with their mother in Wave 4 and have at least one parent alive in Wave 5. We then define the treatment group (foster children) and the control group (non-foster children).

While there is no explicit, universally agreed-upon definition of fostering, we define child fostering as moving out from the household of the parents. This notion is reflected in the frequent use of terms in the literature that emphasize a child's movement, such as "relocation," "movement," "sent to live," "sending households," "receiving households," or "fostering-out," (see, for instance, Isiugo-Abanihe, 1985; Bledsoe and Isiugo-Abanihe, 1989; Castle, 1995; Ainsworth, 1996; Zimmerman, 2003; Akresh, 2004, 2005, 2009; Serra, 2009; Alber, 2013; Hampshire et al., 2015; Beck et al., 2015). These terms emphasize the transfer of a child between household units, framing fostering as a change in the child's living arrangement and environment. Thus, all children left behind by migrant parents are excluded.[12] However, we also produce estimations using an alternative definition of fostering that includes left-behind children, with results discussed in Section 4.5.

Since we use longitudinal data, some individuals interviewed in Wave 4 may not have been followed up in Wave 5. Consequently, each group is further divided into two subgroups based on whether the children were surveyed in Wave 5 or not.

[11] In Wave 4, 3.13% of children aged 0-12 co-reside only with their father while their mother is alive, consistent with other national estimates (e.g., `http://childrencount.uct.ac.za/indicator.php?domain=1&indicator=2`, last accessed 4 October 2024).

[12] Fostering can be interpreted more broadly. For instance, when a parent migrates but leaves their child in the care of another caregiver within the same household unit, the child does not physically move and is often categorized as left behind in the literature (Démurger, 2015). However, this scenario blurs the distinction between fostering and being left behind. Bose-Duker et al. (2021), for example, focus on fostering, resulting specifically from the relocation of parents rather than children.

- **Treatment group.** For those surveyed in Wave 5, we identify foster children as those who change household units and who no longer live with a parent in Wave 5 (and at least one parent is still alive).

  The other subgroup of foster children includes those not surveyed in Wave 5, while their mothers stayed in the household of origin. Given that in South Africa, it is uncommon for children to live solely with their father, we consider that these children are fostered.[13]

- **Control group.** The counterfactual group consists of non-foster children. For those surveyed in Wave 5, non-foster children are those who continue to reside with a parent in Wave 5.

  For those not surveyed in Wave 5, we classify them as non-foster children if their mother is also not surveyed in Wave 5. We assume that it is very likely that the mother and her children stay together as it is the most common living arrangement. Thus, we classify the child as co-residing with the mother and, therefore, non-foster.[14]

Table 4.1 below summarizes the definitions of the treatment and control groups depending on whether children are surveyed or not in Wave 5.

Table 4.1: Description of treatment and control groups

	Surveyed in W5	**Definition**	**N**
Treatment groups: foster children	Yes	Co-resident with their mother in Wave 4 but no longer in Wave 5 and at least one parent still alive in Wave 5. The child moved to a different household unit.	201
	No	Co-resident with their mother in Wave 4 but not surveyed in Wave 5, while the mother remained in the original household unit.	83
Control groups: non-foster children	Yes	Co-resident with their mother in Wave 4 and still co-resident with at least one parent in Wave 5.	5,159
	No	Co-resident with their mother in Wave 4, but neither the child nor the mother is surveyed in Wave 5.	893

Observability/Non-observability For children included in the sample in Wave 4, there are two reasons why their outcomes may not be observed in Wave 5. First, this could be due to individuals not being re-surveyed in Wave 5, either by design or due to unsuccessful follow-up contact. Since the treatment definition involves changing housing units, child fostering is very likely to be related to the probability of being re-surveyed. Indeed, 29.23% of foster children were not surveyed in Wave 5, compared to 14.76% of non-foster children. Second, even if an individual is re-surveyed in Wave 5, anthropometric outcomes in Wave 5 may be missing for various reasons, such as refusal to provide information, unavailability of the child at the time of the survey (for measurement and weighing), or

[13] According to the latest data, in 2022, 3.7% of South African children lived with their father only. This percentage was around 3.6% in 2014 and 3.3% in 2017. The data are available at: http://childrencount.uct.ac.za/indicator.php?domain=1&indicator=2. Last accessed: 4 October 2024.

[14] The control group also includes children who may be indirectly affected by fostering, i.e., children from both the sending and receiving households. These groups constitute 5.60% and 1.45% of the non-foster children, respectively. We also present the results excluding these children from our analysis for robustness. In summary, this exclusion does not alter our results.

postponement of the interview, leading to incomplete data collection. However, nearly all cases of non-observability of the outcome are due to individuals not being surveyed in Wave 5 (89.05%), with only 10.95% attributable to missing data despite being surveyed. Therefore, non-observability related to missing data is a secondary concern. From this point onward, we refer to non-observability in Wave 5 without specifying the different conditions leading to being unobserved.

4.3.5 Descriptive statistics

We now describe our baseline sample from Wave 4. First, we examine children's characteristics by fostering status to identify the main determinants of fostering. Next, we analyze the selection into observability, i.e., the differences between children observed in Wave 5 and those not.

Selection into treatment Tables 4.2 and 4.3 present the baseline characteristics of the child, mothers, and household by fostering status of children. According to Table 4.2, 4.48% of children aged 0-12 at Wave 4 will be fostered within the next two to three years. When comparing foster children to all non-foster children before fostering, we observe several differences at the child level: foster children are more likely to be born out of wedlock, to be African, to have a fair subjective health status (indicating potentially worse health conditions), and are less likely to be covered by medical aid or to reside with their maternal grandparents. At the mother and household level (Table 4.3), foster children tend to have younger mothers, reside in smaller households, with a younger household head, and live in poorer households, among other characteristics.

These differences reveal that foster and non-foster children differ in baseline characteristics, suggesting selection into fostering occurs at various levels. However, there is no significant difference between foster and non-foster children based on their residence; specifically, children from rural areas are neither more nor less likely to be fostered. Finally, Table 4.A.1 presents the intra-household selection of foster children. Foster children are more often the household head's children, have a lower birth order, a higher height-for-age z-score, but a lower subjective health status.

Selection into observability In Tables 4.A.2 and 4.A.3 in the Appendix, we display the differences between the samples of observed and non-observed individuals, regardless of their treatment status. Non-observability correlates with several baseline characteristics, particularly with baseline nutritional status and treatment. This suggests a selection bias in observability, which we address in our empirical strategy.

Table 4.2: Baseline differences between foster and non-foster children, child's characteristics

	(1) Foster	(2) Non-foster	(1)-(2)
Female	0.553	0.508	0.045
	(0.030)	(0.006)	(0.030)
Age	5.546	5.834	-0.288
	(0.217)	(0.046)	(0.217)
Birth order	2.201	2.370	-0.170*
	(0.100)	(0.021)	(0.098)
Born out of wedlock	0.683	0.579	0.104***
	(0.028)	(0.006)	(0.030)
Ethnicity: African	0.880	0.829	0.051**
	(0.019)	(0.005)	(0.023)
Zbmi	0.328	0.298	0.030
	(0.085)	(0.017)	(0.079)
Overweight/Obese	0.271	0.268	0.003
	(0.026)	(0.006)	(0.027)
Healthy	0.687	0.700	-0.013
	(0.028)	(0.006)	(0.028)
Zhfa	-0.873	-0.845	-0.029
	(0.085)	(0.016)	(0.076)
Stunted	0.187	0.155	0.032
	(0.023)	(0.005)	(0.022)
Ill/disable	0.056	0.038	0.019
	(0.014)	(0.002)	(0.012)
Covered by medical aid	0.053	0.089	-0.036**
	(0.013)	(0.004)	(0.017)
Poor subjective health status	0.007	0.003	0.004
	(0.005)	(0.001)	(0.003)
Fair subjective health status	0.035	0.018	0.018**
	(0.011)	(0.002)	(0.008)
Good subjective health status	0.148	0.157	-0.009
	(0.021)	(0.005)	(0.022)
Very good subjective health status	0.278	0.326	-0.047*
	(0.027)	(0.006)	(0.028)
Father alive	0.905	0.922	-0.017
	(0.017)	(0.003)	(0.016)
Father and paternal grandparent coresident	0.032	0.038	-0.006
	(0.010)	(0.002)	(0.012)
Paternal grandparents alive	0.099	0.125	-0.027
	(0.018)	(0.004)	(0.020)
Mother and maternal grandparent coresident	0.254	0.323	-0.069**
	(0.026)	(0.006)	(0.028)
Maternal grandparents alive	0.391	0.400	-0.009
	(0.029)	(0.006)	(0.030)
Observations	284	6,052	6,336

Notes: Sample of children aged 0-12 years at baseline (Wave 4). Foster are children who will move out from the parental household in Wave 5. Baseline corresponds to Wave 4 of the NIDS. Standard errors are in parentheses. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table 4.3: Baseline differences between foster and non-foster children, mother's and household's characteristics

	(1) Foster	(2) Non-foster	(1)-(2)
Mother: age	30.356	32.603	-2.248***
	(0.452)	(0.102)	(0.482)
Mother: nb years in union	1.106	1.093	0.012
	(0.234)	(0.047)	(0.221)
Mother: separated/divorced	0.000	0.016	-0.016**
	(0.000)	(0.002)	(0.008)
Mother: widow	0.028	0.034	-0.006
	(0.010)	(0.002)	(0.011)
Mother: primary educated	0.077	0.083	-0.005
	(0.016)	(0.004)	(0.017)
Mother: secondary educated	0.722	0.700	0.022
	(0.027)	(0.006)	(0.028)
Mother: tertiary educated	0.169	0.186	-0.017
	(0.022)	(0.005)	(0.024)
Mother: diabetic	0.004	0.005	-0.001
	(0.004)	(0.001)	(0.004)
Mother: HIV	0.053	0.040	0.013
	(0.013)	(0.003)	(0.012)
Mother: decides about expenses	0.511	0.503	0.007
	(0.030)	(0.006)	(0.030)
Mother: poor subjective health status	0.007	0.012	-0.005
	(0.005)	(0.001)	(0.007)
Mother: fair subjective health status	0.049	0.051	-0.002
	(0.013)	(0.003)	(0.013)
Mother: good subjective health status	0.306	0.283	0.023
	(0.027)	(0.006)	(0.027)
Mother: very good subjective health status	0.306	0.336	-0.029
	(0.027)	(0.006)	(0.029)
Rural	0.486	0.488	-0.002
	(0.030)	(0.006)	(0.030)
Household size	6.151	6.746	-0.595***
	(0.181)	(0.044)	(0.207)
Head: female	0.711	0.656	0.055*
	(0.027)	(0.006)	(0.029)
Head: age	43.042	45.831	-2.789***
	(0.900)	(0.191)	(0.904)
Wealth index [0-1]	0.454	0.549	-0.095***
	(0.011)	(0.002)	(0.011)
Expenditures (per adult equivalent)	1,444.534	1,898.354	-453.820**
	(117.766)	(39.668)	(184.896)
Head: primary educated	0.162	0.175	-0.013
	(0.022)	(0.005)	(0.023)
Head: secondary educated	0.553	0.532	0.021
	(0.030)	(0.006)	(0.030)
Head: tertiary educated	0.183	0.150	0.034
	(0.023)	(0.005)	(0.022)
Household with retired pension	0.229	0.252	-0.023
	(0.025)	(0.006)	(0.026)
Agricultural household	0.236	0.231	0.005
	(0.025)	(0.005)	(0.026)
Household received remittances	0.236	0.279	-0.043
	(0.025)	(0.006)	(0.027)
Household sent remittances	0.162	0.126	0.036*
	(0.022)	(0.004)	(0.020)
Distance to police station	2.574	2.630	-0.056
	(0.064)	(0.014)	(0.066)
Negative household income shock last 2 years	0.123	0.108	0.016
	(0.020)	(0.004)	(0.019)
Observations	284	6,052	6,336

Notes: Sample of children aged 0-12 years at baseline (Wave 4). Foster are children who will move out from the parental household in Wave 5. Baseline corresponds to Wave 4 of the NIDS. Standard errors are in parentheses. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

4.4 Empirical strategy

In this section, we outline our approach to estimate the causal effect of fostering on children's nutritional status, using a model that simultaneously addresses selection into treatment and selection into observability, building on the framework of Bia et al. (2023).

4.4.1 Identification strategy

To describe our empirical strategy, we refer to the potential outcomes framework (Rubin, 1974). At first, we set aside the observability issue, which we incorporate later. Let D_i be the treatment variable for individual i, taking the values 0 (non-foster child) and 1 (foster child). Under the stable unit treatment value assumption,[15] Y_i represents the nutritional outcome in Wave 5 for child i, with potential values $Y_i(1)$ if the child is fostered and $Y_i(0)$ if the child is not fostered. The observed outcome is $Y_i = Y_i(0) + [Y_i(1) - Y_i(0)] \cdot D_i$. The causal effect of fostering for individual i is $\Delta_i = Y_i(1) - Y_i(0)$. We aim to estimate the average treatment effect (ATE) $E(\Delta)$.

Under the assumption of conditional independence of the treatment, various methods have been proposed to estimate the ATE. This assumption posits that the covariates in the data are comprehensive enough to control for the influence of any confounder.[16] One method involves conditioning on the covariates to derive the conditional average treatment effect (CATE): $\Delta_x = E[Y(1)|X = x] - E[Y(0)|X = x] = \mu_1(x) - \mu_0(x)$. Subsequently, averaging across the values x appearing in the population yields the ATE: $\Delta = E[\mu_1(x) - \mu_0(x)]$.

Alternatively, under the same assumption of conditional independence of the treatment, one may also predict the likelihood of being treated based on the covariates, obtain a propensity score for treatment $p(X)$, and then weigh observations according to this propensity score to ensure that the treatment and control groups have comparable characteristics. The Inverse Probability Weighting (IPW) approach is one of those methods and computes the sample analog to $\Delta = E\left[\frac{Y \cdot D}{p(X)} - \frac{Y \cdot (1-D)}{1-p(X)}\right]$ to estimate the ATE.

Combining both approaches is more efficient (Chernozhukov et al., 2018). The ATE can be expressed as $\Delta = E[\phi(X)]$ where $\phi(X) = \mu_1(X) - \mu_0(X) + \frac{[Y-\mu_1(X)] \cdot D}{p(X)} - \frac{[Y-\mu_0(X) \cdot (1-D)]}{1-p(X)}$. The estimator is the sample analog to the previous expression and is doubly robust, i.e., the resulting estimator maintains consistency under the correct specification of either the conditional mean outcomes or the propensity score model. A key advantage of this approach lies in its robustness to model misspecification: estimation errors in either component are attenuated through their interaction. In practice, combining a regression approach (to estimate conditional means outcomes) with a propensity score estimation (via logistic regression, for instance) yields more efficient estimates than either method alone, although both approaches rely on the same conditional independence assumption (but not the same specification). It has also been shown to have better properties in small samples (Rothe and Firpo, 2013). Part of the identification strategy consists of using ML non-parametric estimation to obtain the components of ϕ. This data-driven approach allows us to incorporate a rich set of covariates, thereby strengthening the plausibility of the conditional independence assumption while reducing model specification concerns.

[15] Child i's nutrition is only influenced by its own treatment value.
[16] We discuss this assumption in more detail below.

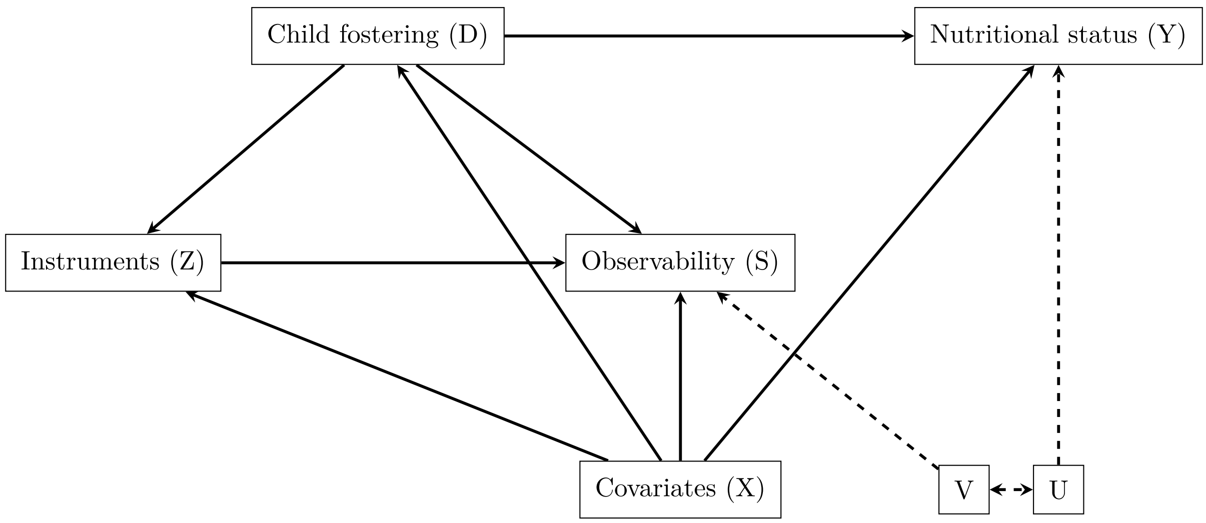


Figure 4.1: Hypotheses of the model, adapted from Bia et al. (2023)

We now add the additional block of the non-random selection into outcome observability. As depicted in Figure 4.1, the causal model accounts for this selection process, adapted from Bia et al. (2023). The arrows indicate a potential effect between variables, while the absence of an arrow indicates that such an effect is assumed to be negligible or non-existent. For instance, while unobserved variables (U) affect nutritional status (Y), they are assumed not to influence child fostering (D), as stated earlier. However, these unobserved characteristics are correlated with the ones (V) affecting Observability S, which determines whether Y is observed. The identification strategy relies on instrumental variables (Z) that correlate with S, may correlate with the treatment D, and covariates X, but should not be affected by omitted factors (U, V).

To address endogenous observability, we introduce propensity scores for the probability of being observed, noted as $\pi = \pi_1(X, Z)$ for treated individuals and $\pi_0(X, Z)$ for untreated individuals. The above notations need to be slightly adjusted to account for the joint determination of the likelihood of being treated and observed.

$$
\begin{aligned}
&\Delta = E[\phi_1 - \phi_0], \quad \text{with} \\
&\phi_1 = \frac{D \cdot S \cdot [Y - \mu_1(S = 1, X, \Pi)]}{p(X, \Pi) \cdot \pi_1(X, Z)} + \mu_1(S = 1, X, \Pi) \\
&\phi_0 = \frac{(1 - D) \cdot S \cdot [Y - \mu_0(S = 1, X, \Pi)]}{(1 - p(X, \Pi)) \cdot \pi_0(X, Z)} + \mu_0(S = 1, X, \Pi).
\end{aligned}
\tag{4.1}
$$

where $p(X, \Pi)$ denotes the propensity score for treatment conditional on covariates and the propensity scores of observability, and $\mu_1(S = 1, X, \Pi)$ represents the average outcome among treated, observed individuals with covariates X and probability of being observed Π.

The identification of the ATE relies on the following set of assumptions:

- Assumption 1: Conditional independence of the treatment.

$$Y(0), Y(1) \perp D|X = x \quad \text{for all } x \text{ in the support of } X$$

- Assumption 2: Instrument for selection into observability. The instrument Z, which may depend on D, is correlated with S but orthogonal to the outcome Y, conditional on the treatment and covariates.[17]

- Assumption 3: Common support for the treatment.

$$0 < p(x, \pi) < 1, \quad \text{for all } x, z \text{ in the support of } X, Z.$$

- Assumption 4: Common support for selection.

$$\pi_0(x, z) > 0 \quad \text{and} \quad \pi_1(x, z) > 0 \quad \text{for all } x, z \text{ in the support of } X, Z.$$

- Assumption 5: Conditional effect homogeneity.

$$E[Y(1) - Y(0)|S = 1, X = x, V = v] = E[Y(1) - Y(0)|X = x, V = v]$$
$$\text{for all } x, v \text{ in the support of } X, V.$$

A sufficient condition for effect homogeneity is the separability of observed and unobserved components in the outcome equation, that is $Y = \eta(D, X) + \nu(U)$ with η and ν general functions (Huber, 2014).

Under this set of assumptions, the expression in (4.1) is doubly robust and can be estimated by plugging-in the sample analogs.

4.4.2 Specification choices

Among our identifying assumptions, the conditional independence of the treatment requires particular attention. Its validity rests on two components: first, our ability to control for all relevant covariates that jointly determine fostering decisions and nutritional outcomes, and second, on our ability to use a well-specified model.

Covariates We employ a rich set of pre-treatment covariates, including child-level variables (gender, age, relationship to the head, ethnic group, etc.), parental characteristics (information on the father's living conditions, co-residence with maternal/paternal grandparents, mother's marital status, etc.), and

[17] The formal assumption 2 is the following:

(a) The instrument Z (which may be a function of D, $Z = Z(D)$) is conditionally correlated with S, that is: $E[Z \cdot S|D, X] \neq 0$, and satisfies (i) $Y(d, z) = Y(d)$ and (ii) $Y \perp Z|D = d, X = x$ for all $d \in \{0, 1\}$ and x in the support of X.

(b) $S = I\{V \leq \chi(D, X, Z)\}$, where V is a scalar (index of) unobservables and χ is a general function with a strictly monotonic cumulative distribution function conditional on X.

(c) $V \perp (D, Z)|X$.

household characteristics (area of residence, household size, characteristics of the household head, district dummies, etc.). In line with Angrist and Pischke (2009), we do not use post-treatment variables to avoid introducing inappropriate controls that may have been affected by the treatment itself.

A crucial feature of our specification is the inclusion of baseline nutritional indicators (BMI-for-age and height-for-age z-scores in Wave 4). This allows our model to measure changes in nutritional status for foster children compared to non-foster ones. In essence, disregarding the aspect related to observability, this approach is similar to a value-added model, where fostering may either improve or worsen a child's nutrition relative to their pre-treatment level. The exhaustive list of baseline covariates employed in our estimation procedure is detailed in Table 4.A.4 of the Appendix.

Model specification Our setting potentially presents a methodological challenge: while we have access to a large and rich set of potential covariates, the number of treated children remains limited. ML methods are particularly well-suited to deal with such situations, as they offer greater flexibility and choose the relevant covariates in a data-driven way rather than with ad hoc selection. We adopt a fully non-parametric method, which is detailed below. All the plug-ins that are necessary to estimate ϕ_0 and ϕ_1 $(\mu_0(.), \mu_1(.), p(.), \pi_0(.), \pi_1(.))$ will be obtained using this ML method.

Instruments for selection We exploit two instruments from the NIDS tracking design to address selection into observability. As previously mentioned, the survey design distinguishes between Continuing Sample Members (CSM) and Temporary Sample Members (TSM): biological children of CSM mothers are tracked across survey waves even if they move and change households, while children of TSM mothers are not. We leverage this specific survey feature to build two exogenous instruments.

The first instrument is a binary variable indicating whether a child has a CSM mother, which, by survey design, should increase the likelihood of being observed in Wave 5. The second instrument is a continuous variable representing the district-level share of children with CSM mothers in Wave 5. The intuition behind this continuous instrument is that enumerators may face an increased burden in tracking individuals in districts with higher proportions of children with CSM mothers. In other words, it may pose more significant logistical challenges for enumerators conducting follow-up surveys. This increased difficulty likely results in higher attrition rates in the area, as the effort required to maintain consistent follow-up with these families becomes more demanding and resource-intensive. Interestingly, this second instrument likely affects the capacity to re-survey any individual, not only the CSMs. The survey design should thus generate exogenous variation in the probability to be observed in Wave 5.[18] We assess the instrument's predictive power in Section 4.4.3.1.

Threats to identification The estimates could be biased if (a) we fail to account for variables that affect both the treatment and the potential outcomes or (b) the instrument correlates with unobservables affecting the observability of the child or the fostering decision.

One potential threat falling in the (a) category is if parents have specific preferences favoring human capital that lead them to foster a specific child. For this reason, we control for birth order, which is

[18]The continuous instrument has missing values due to the absence of district of residence information in Wave 5. This affects 11.21% of the sample (710 out of 6,336 individuals) and only concerns the non-observed group, i.e., those for whom we cannot observe an outcome in Wave 5. They represent 64.78% of the non-observed children. Our strategy involves replacing these missing values with the mean value of this variable at the regional level, tacking the region of residence at Wave 4. In other words, we substitute each missing value with the regional mean of the share of children (by district) with CSM mothers. Regarding the distribution of the replaced missing values by fostering status, we proxy the continuous instrument for 10.58% of non-foster children and 24.65% of foster children.

known to predict different parental investments. We also control for several variables related to the mother's health and education, as well as the child's health (prior nutrition, serious illnesses, perceived health status, etc.). The ML algorithm we chose is non-parametric and, therefore, potentially selects any "interaction" between these variables. Thus, we assume that if the household has specific preferences, these would be captured by the available variables in Wave 4.

A second potential threat is the case where a shock impacts both the decision to foster and the child's nutrition. While we control for adverse income shocks happening over the two years before Wave 4, a shock that happens after Wave 4 may trigger the fostering decision. To rule out this possibility, we assess whether fostering is associated with shocks happening between Wave 4 and Wave 5, and we show below that this is not the case.

Related to (b), the instrument's validity is questionable if Z, despite being determined by the survey design, correlates with characteristics that also explain the observability of the child or the fostering decision. TSM women, having joined the panel later, might exhibit less residential stability and, thus, a greater propensity for child fostering. We account for this by controlling for maternal marital status, a strong correlate of movement between households (Gautrain, 2023), as well as the number of rounds where the mother is present in the surveyed household unit of Wave 4. These controls capture systematic differences in women's propensity to change households.

Implementation Among various ML techniques, we opt for the Random Forest (RF) procedure, a classifier that allows multiple decision trees to be drawn to classify observations based on covariates. Each tree looks at different random subsets of the data and variables. Therefore, RF can automatically explore and capture interactions between variables. For example, suppose two variables interact in a way that affects the outcome. In that case, the trees in the forest can capture this by making different splits based on various combinations of these variables. Combining all the trees, the RF can account for many potential interactions, even if they are complex or involve multiple variables. The RF procedure is also fully non-parametric and does not restrict the functional form between the outcomes and the explanatory variables nor between the observability, the treatment, and the covariates. This, therefore, limits the risk of misspecification, which would lead to a biased estimate of the causal effect.

We use the *treatselDML* command available in the *causalweight* package in *R* (Bodory and Huber, 2018), using 3-fold cross-fitting and the RF with default options of the *SuperLearner* package as machine learner.[19] We run estimations with 6,336 individuals and 120 covariates measured at baseline. We exclude observations where the product of the estimated treatment and observability propensity scores is below a threshold of 0.01, in order to guarantee that the common support assumptions are satisfied.[20]

[19] We conducted a preliminary screening to remove highly correlated variables. In line with Bia et al. (2023), we have standardized all continuous variables to a mean of 0 and a standard deviation of 0.5 to enhance the machine learning-based estimation of nuisance parameters.

[20] We also provide the results with a different trim level.

4.4.3 Assessing the validity of the method

Before presenting the results, we conduct key tests to assess the validity of our model. Specifically, we justify the relevance of our instruments and demonstrate the validity of the propensity score approach used in the double machine learning.

4.4.3.1 Instrumentation

We first present the results analogous to a "first-stage" analysis, predicting observability based on the instrumental variables. Figure 4.2 illustrates the importance of the explanatory variables in predicting selection into observability. In this context, variable importance reflects the contribution of each variable to reducing prediction errors across the decision trees in the forest. Variables ranked higher have a more significant impact on the model's predictive accuracy. The algorithm identifies the two instrumental variables as the most critical variables, ranking them first and second among all the covariates. This indicates that our instruments Z are highly effective in predicting observability.

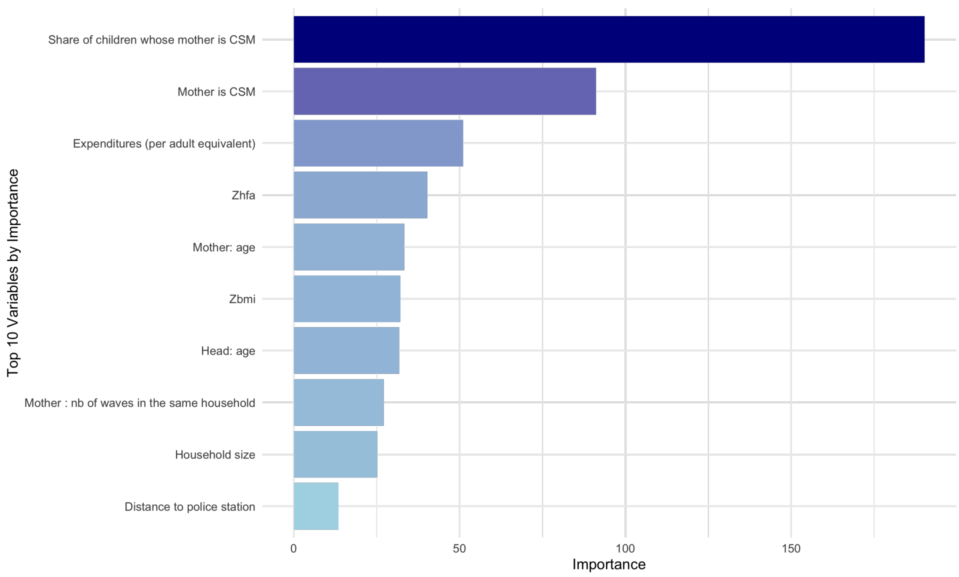


Figure 4.2: Importance of the instruments for explaining selection into observability

Notes: Sample of children aged 0-12 at baseline (Wave 4). The share of children whose mother is CSM is measured at the district level. The graph shows the importance of covariates on an outcome variable, following a random forest with 3-fold cross-fitting. The outcome variable is a dummy equal to one if health outcomes are observable in Wave 5. Covariates are measured pre-treatment; the list is provided in Appendix 4.A.4.

To further evaluate the instrument's validity, we run an OLS regression on "being observed in Wave 5", including Z, X, and D as regressors. Table 4.4 shows that having a CSM mother increases the likelihood of being observed in Wave 5 (column 1), while the share of CSM mothers in the district decreases this likelihood (column 2). Both effects remain significant when combined (column 3), justifying the use of both variables in our preferred specification. Although the method requires a continuous instrument for identification, relying on only one might weaken identification, so we include both to strengthen the model. In column 3, the F-statistic is around 268, indicating strong predictive power when the instruments are used jointly.[21]

[21] In column 3, the F-statistic for the continuous IV alone is 10.32.

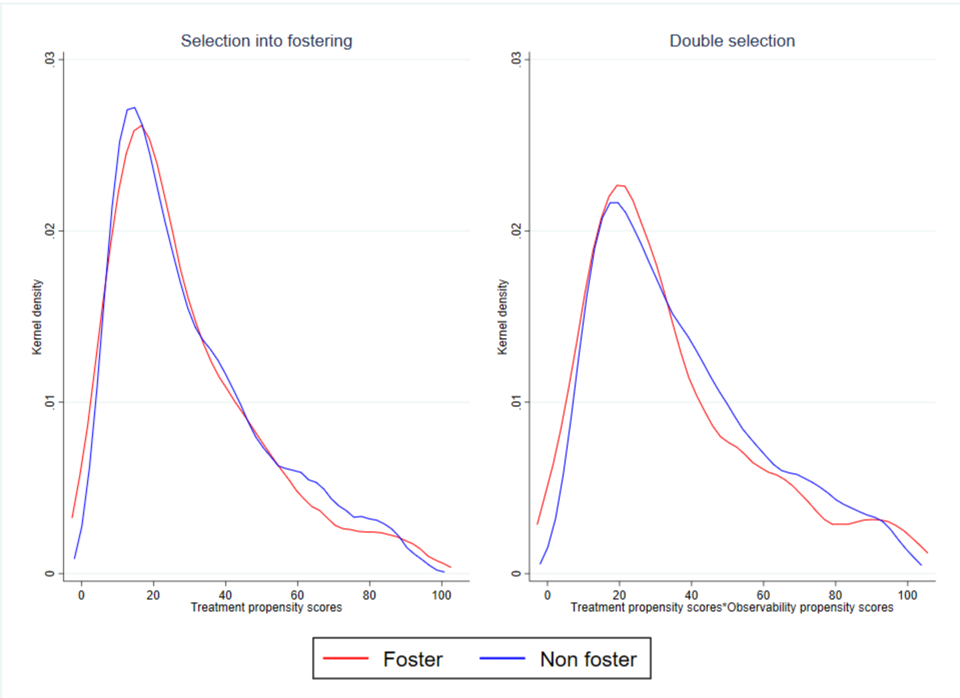
Table 4.4: OLS estimates, selection into observability and the instruments

	(1)	(2)	(3)
Outcome: Being observed			
Mother is CSM	0.262***		0.263***
	(0.011)		(0.011)
Share of children whose mother is CSM		-0.432**	-0.588***
		(0.190)	(0.183)
Adjusted R^2	0.158	0.088	0.159
Observations	6,336	6,336	6,336
Average outcome	0.827	0.827	0.827
F	524.598	5.161	267.854

Notes: Sample of children aged 0-12 years at baseline (Wave 4). Share of children whose mother is CSM is measured at the district level. The outcome variable is a dummy equal to one if anthropometric measures are observed in Wave 5. Covariates are measured pre-treatment, they include the treatment (fostering) and the list provided in Appendix 4.A.4. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$

4.4.3.2 Common support assumptions

Figure 4.3 displays two kernel density plots showing the distribution of propensity scores for foster and non-foster children after the trimming. The left-hand plot shows a clear overlap in the treatment propensity scores. The right-hand plot incorporates treatment and observability propensity scores (reflecting the double selection) and demonstrates the overlap. These figures confirm that the common support assumption is validated after trimming.



Notes: Sample of children aged 0-12 at baseline (Wave 4). Propensity scores of treatment and observability are extracted from the DML: random forest with 3-fold cross-fitting and a trimming threshold of 0.01. Trimmed observations are excluded. The graph on double selection is drawn only for observed individuals. The outcome variable is being stunted and is measured in Wave 5. Covariates are measured pre-treatment; the list is provided in Appendix 4.A.4.

Figure 4.3: Distribution of the propensity scores

4.5 Results

4.5.1 Impacts on foster children

Table 4.5 presents the Double Machine Learning (DML) estimates. We find a negative and statistically significant effect of fostering on the stunted outcome, indicating that fostering significantly reduces the probability of stunting. The estimated effect of -0.070 corresponds to approximately a 45% reduction in the likelihood of being stunted, given the baseline prevalence of stunting at 15.64%.[22] In contrast, no significant effect is found on the healthy, or overweight/obese outcomes, suggesting that fostering has no substantial impact on these dimensions of children's nutrition.

Table 4.5: ATE estimates based on the DML procedure

Outcome	ATE	Standard error	p-value	Baseline mean	N
Healthy	0.054	0.035	0.119	0.699	6,336
Overweight/Obese	-0.029	0.034	0.394	0.268	6,336
Stunted	-0.070***	0.018	0.000	0.156	6,336

Notes: Sample of children aged 0-12 years. The treatment is fostering. The outcome variables are nutritional statuses based on BMI-for-age z-scores (healthy and overweight/obese) and height-for-age z-scores (stunted). Covariates are measured pre-treatment, and the list is provided in Appendix 4.A.4. DML: random forest with 3-fold cross-fitting and a trimming threshold of 0.01. Standard errors are based on asymptotic approximations using the estimated variance of the efficient score functions. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$

Overall, our findings reveal that child relocation substantially lowers the incidence of stunting, highlighting that foster children are likely to experience significantly better long-term nutritional outcomes. However, fostering appears to have no noticeable effect on the short- to medium-term nutritional status of children (healthy and overweight/obese). These findings diverge from previous studies in sub-Saharan Africa, which often report negative effects of fostering on children's nutrition (Ariyo et al., 2019). These discrepancies are likely due to variations in treatment definitions, such as excluding orphans or those who do not relocate, as well as the use of different methods. Regarding the method, we also apply more standard empirical approaches to estimate average treatment effects in Section 4.5.2. As for the definition, excluding cases of parental migration—which may negatively impact children's nutrition (Gosselin-Pali, 2025)—could also explain some of the differences observed compared to previous findings. To explore this, we expand the definition of foster children to include those left behind by parents in their household of origin between waves 4 and 5. This group could be considered fostered, as they stop co-residing with either parent between the two survey rounds, although both parents are still alive. However, as they do not relocate, they were excluded from our previous analyses.[23] The results for this alternative definition of fostering are shown in Table 4.C.8. They indicate that living away from one's parents, regardless of whether it involves relocation, still has a negative and significant effect on the probability of stunting. However, the estimated effect of -0.028 is smaller in magnitude than the results in Table 4.5. This difference is discussed in Section 4.5.4.

[22] The percentage reduction in stunting is calculated as: $\left(\frac{0.070}{0.1564}\right) \times 100 \approx 44.76\%$.

[23] In total, there are 284 foster children according to our definition. However, when we also include children who stop co-residing with any parent in Wave 5 and did not move out (left behind), the number increases to 644 foster children. It is important to note that for left-behind children, it is not possible to include those who were not observable in Wave 5.

4.5.2 Results from standard methods

In this section, we provide the findings from standard impact evaluation methods. Although these methods are not directly comparable with our empirical framework due to differing identification assumptions, static OLS is widely used in the child fostering literature, while Difference-in-Differences (DID) is typically applied to estimate ATEs in longitudinal data.

Cross-sectional results In the literature, most studies assess the impact of child fostering using cross-sectional data (Ariyo et al., 2019). Therefore, we apply OLS without using the longitudinal nature of the data nor addressing any selection issue to provide a comparison. This entails a difference with our approach in the identification of the treated group. Foster children are identified as those living without parents in Wave 5 and who have at least one parent alive; the control group comprises children residing with at least one parent in Wave 5. This change in the definition of foster children leads us to include children who, unlike the definition used in the DML procedure, do not change household units and for whom we cannot confirm that they were previously living with their parents in Wave 4. Incidentally, we use the stock of children living away from their parents rather than the flow of children leaving their parents. As we only consider Wave 5 of the survey, the covariates included in the estimates are only measured post-fostering. These constraints underscore the limitations of cross-sectional analyses in estimating the causal impact of fostering. As summarized in Table 4.D.1 in the Appendix, none of the OLS specifications yield statistically significant results, even when varying the set of controls.

Panel results without addressing selection issues Another approach exploits the longitudinal nature of the data but neglects the attrition issue. We focus on the subsample of children surveyed in Wave 4 who are also re-surveyed in Wave 5.[24] Then, pooling the sample for the two waves, we control for (future) treatment in Wave 5 and estimate the effect of being fostered on the interaction between Foster and Wave 5. This identification is akin to a DID strategy but requires panel data to identify the treated group, not just several cross-sections. We rely on the parallel trend assumption between foster and non-foster children conditional on pre-treatment characteristics. This method is applied in Table 4.D.2 in the Appendix. The results indicate a statistically significant negative effect of fostering on the probability of being stunted in most specifications, except for column (3), suggesting that fostering reduces the likelihood of stunting within this framework. It is reassuring to see that changing the method and not relying on machine learning provides comparable results. Moreover, the fact that we do not find a clear difference with our main results suggests that taking into account the endogenous attrition may not be crucial. However, this claim should be approached with caution, as we use a panel with individual tracking, and the attrition rate for foster children remains low compared to more standard panel surveys.

4.5.3 Robustness checks

4.5.3.1 Sample-based checks

Excluding non-foster children in sending and receiving households from the control group In the primary analysis, the control group, in addition to including children with no connection to child fostering, also comprises children residing in sending and receiving households. Indeed, we consider

[24] Re-surveyed children represent 70.77% of the treated group and 85.24% of the control group, accounting for approximately 84.60% of the overall sample.

the counterfactual of a foster child to be a non-foster child, regardless of their specific characteristics. However, children in sending and receiving households could be indirectly affected by fostering. They represent, respectively, 5.60% and 1.45% of the non-foster children. As the treatment, child fostering, could have spillover effects on these children, their inclusion may lead to a violation of the SUTVA assumption. As a robustness check, we re-estimate the results after excluding children from sending and receiving households from the control group to ensure that potential spillover effects do not drive the findings.

The results, shown in Appendix Table 4.C.1, indicate that the estimated effect on the likelihood of being stunted is slightly smaller in magnitude but remains negative and significant at the 1% level. While the significance of the results changes for the other two outcomes, healthy and overweight/obese, the direction of the effects remains consistent: fostering increases the likelihood of being healthy and decreases the likelihood of being overweight or obese. These findings, therefore, support a positive normative effect of fostering on children's nutritional status.[25]

Excluding children aged 0-2 years In many studies, children younger than 2 are excluded due to the higher likelihood of measurement errors in length data for this age group (WHO and UNICEF, 2019). Therefore, we tested an alternative sample excluding children under 2 years old at baseline. According to Table 4.C.4 in the Appendix, the estimated ATE for stunted remains negative, highly significant, and of a similar magnitude (-6.2 percentage points) to that in Table 4.5.

4.5.3.2 Model-based check

In the default version of the DML procedure, 1% of observations are discarded due to extreme propensity scores. This trimming rule is applied to avoid excessively small denominators in weighting by the inverse of the propensity scores. In a robustness analysis, we adjust this threshold to 2%. The results, available in Appendix Table 4.C.5, confirm that the main findings on the probability of being stunted remain robust under this alternative trimming rule (-5.5 percentage points).

4.5.3.3 Placebo tests

As a first placebo test, we estimate the ATEs using the DML method, where the outcomes are measured in Wave 4, while the treatment occurs between Waves 4 and 5, and the covariates are taken from Wave 3. Thus, the analysis focuses only on children with complete information in waves 3 and 4, aged 2-12 in Wave 4. In this sample, all children are observed in both waves, and the DML procedure addresses only the selection into treatment.[26] As expected, our treatment does not affect baseline nutritional status when controlling for covariates from Wave 3. Indeed, as shown in Appendix Table 4.C.6, the ATEs are insignificant.

In a second placebo test, we simulate a random treatment assignment. Specifically, individuals are randomly assigned to either a treatment or a control group, with the proportion of treated individuals matching the percentage of foster children (4.48%). Appendix Table 4.C.7 confirms that no significant ATEs are observed under random treatment assignment.

[25] In Tables 4.C.2 and 4.C.3, we also present the results after separately excluding children from sending households and host households, respectively. The results remain consistent.

[26] This is done with the *treatDML* command from the *causalweight* package in *R* (Bodory and Huber, 2018).

4.5.4 Mechanisms

Our main findings show that fostering positively impacts children's nutritional status, mainly by reducing the likelihood of stunting. To explore potential mechanisms, we apply the DML framework to post-treatment mediators, such as household and caregiver characteristics. The rationale is to identify pathways through which the positive impact on nutrition might occur.

Change in household characteristics First, the improvement in children's nutritional status may be attributed to a change in the living conditions of the child. The analysis of mediators (Table 4.6) suggests that a key mechanism appears to be changes in household composition. Children who are fostered typically join smaller households with fewer working-age adults compared to households of non-foster children. Additionally, foster children tend to integrate into households with a significantly higher proportion of adults aged 65 and over. This finding is consistent with the higher likelihood of foster children residing in households with a retired pension beneficiary. The difference in household income sources may be advantageous for foster children, as pension income is a more stable and predictable resource, less vulnerable to idiosyncratic economic shocks. This observation aligns with Hamoudi and Thomas (2014), who document the relationship between pension income and co-residence between grandparents and grandchildren in South Africa. The importance of income from old-age benefits for child outcomes is also supported by

Table 4.6 also provides evidence that foster children are more likely to reside in rural areas than non-foster children. This has nutritional implications, as rural households are likely to engage in agricultural activities, which can enhance access to diverse and nutritious foods (Govender et al., 2017). The rural location could, therefore, contribute to improved dietary diversity and nutrition, potentially explaining the reduced likelihood of stunting.

In addition, child fostering seems to increase the probability of residing in households that received remittances over the past year. At this stage, however, we cannot determine whether these remittances are sent by the origin household of foster children, nor whether they are a result of hosting the child, or if they precede the hosting decision. Therefore, caution should be exercised in interpreting this result, and we return to this issue below (see Section 4.6).

Finally, the alternative analysis we made, which estimates the effects of fostering using a broader definition that includes left-behind children as fostered, also confirms that relocating a child to a different environment is a key factor influencing nutrition. Specifically, comparing the main results based on the definition of fostering as child relocation (Table 4.5) with those when fostering includes children that should be considered left behind (Table 4.C.8) reveals a smaller impact of fostering on the probability of being stunted when left-behind children are included. This difference suggests that child out-fostering may represent a more deliberate, child-centered decision, while leaving children behind may reflect broader household strategies, such as income diversification, which likely have distinct impacts on nutrition. Overall, this explanation, along with the roles of household composition and location as critical mediators, aligns with the observed decrease in the impact when left-behind children are classified as fostered.

Changing caregivers Fostering, by definition, involves having a new caregiver, which may also explain the positive impacts on children's well-being. The mediators' analysis in Table 4.6 points out that foster children are significantly more likely to be under the direct care of the household head or the head's spouse in their host households.[27] This change in caregiving arrangements is noteworthy, as it suggests that foster children are placed under the direct care of the primary decision-makers in host households. Such a position could enhance these children's access to resources through more favorable intra-household allocation decisions. This would reconcile the observed improvement in nutritional outcomes with the absence of change in expenditures for the household as a whole, aligning with previous evidence on the role of proximity to the household head (Brown et al., 2019; De Vreyer and Lambert, 2021).

Changes in household composition can also result in variations in the availability of potential caregivers. As receiving households have more often elderly people, these individuals may have more time to dedicate to children. Although it is unclear how this directly translates into improved nutrition, it is notable that the shift in household composition significantly increases a child's likelihood of being cared for by a grandparent. This suggests that, in our context, the caregiver, even if he or she is not the parent, is likely to be quite altruistic towards the child, and this may explain the positive impact.[28]

Table 4.6: ATE estimates based on the DML procedure, alternative outcomes

Outcome	ATE	Standard error	*p*-value	Baseline mean	N
Household size	-0.847***	0.272	0.002	6.720	6,336
Nb of household members 0-14	-0.043	0.158	0.784	3.118	6,336
Nb of household members 15-64	-0.850***	0.140	0.000	3.313	6,336
Nb of household members 65+	0.102**	0.048	0.032	0.192	6,336
Wealth index	-0.031**	0.015	0.041	0.529	6,336
Food expenditures	-115.624	70.309	0.100	1327.142	6,336
Expenditures	-2078.826***	345.629	0.000	5018.046	6,336
Food expenditures (per capita)	25.268	28.217	0.371	232.547	6,336
Expenditures (per capita)	-265.334**	117.032	0.023	937.297	6,336
Food expenditures (per adult equivalent)	47.401	47.866	0.322	473.573	6,336
Expenditures (per adult equivalent)	-565.947***	190.377	0.003	1878.013	6,336
Food expenditures (per calories adult equivalent)	42.071	33.938	0.215	289.440	6,336
Household with retired pension	0.212***	0.049	0.000	0.251	6,336
Rural	0.207***	0.039	0.000	0.488	6,336
Negative household income shock last 2 years	0.009	0.036	0.795	0.108	6,336
Cared by head or head's spouse	0.259***	0.025	0.000	0.575	6,303
Grandchild of the head	0.274***	0.046	0.000	0.331	6,336
Household received remittances	0.111**	0.046	0.016	0.277	6,336
Household sent remittances	-0.010	0.035	0.773	0.128	6,336

Notes: Sample of children aged 0-12 years at baseline (Wave 4). The treatment is fostering. Covariates are measured pre-treatment, and the list is provided in Appendix 4.A.4. DML: random forest with 3-fold cross-fitting and a trimming threshold of 0.01. Standard errors are based on asymptotic approximations using the estimated variance of the efficient score functions. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$

4.6 Spill-overs

4.6.1 Indirect impacts on non-foster children in sending households

With our data, we can also explore the effects of fostering on children who remain in the sending households. Using the same DML model, we estimate the effect of fostering, with the treatment group consisting of children from sending households. The control group now includes all children not involved in fostering and excludes all children in host households.

[27] This is further supported by Table 4.E.3 in the Appendix, which shows that 86.6% of the foster children are under the care of the household head or their spouse.

[28] Evidence from sub-Saharan Africa indicates that children living with grandmothers have significantly lower odds of being stunted (Schrijner and Smits, 2018).

As shown in Table 4.7, the results indicate that fostering has a negative and significant effect on the probability of stunting for children who remain in the household of origin. In other words, this suggests that fostering also reduces the likelihood of stunting among siblings staying with their parents, compared to non-foster children. The effect is, however, smaller, with a reduction in stunting by 4.3 percentage points, compared to a reduction of 7 percentage points for foster children (Table 4.5). For the other outcomes, the coefficients are not statistically significant.

The key takeaway from these results, along with those for foster children, is that although the indirect effect of fostering on children remaining in the household of origin is smaller, fostering appears to create a potentially mutually beneficial arrangement for both groups of children involved in the practice of fostering, particularly in terms of long-term nutritional outcomes.[29]

Table 4.7: ATE estimates based on the DML procedure, children from sending households as the treated group

Outcome	ATE	Standard error	p-value	Baseline mean	N
Healthy	0.024	0.034	0.483	0.700	5,964
Overweight/Obese	-0.012	0.032	0.697	0.267	5,964
Stunted	-0.043**	0.022	0.048	0.154	5,964

Notes: Sample of children aged 0-12 years at baseline (Wave 4). The treated group consists of children from the sending households, and the control group consists of non-foster children (excluding host siblings of foster children). The outcome variables are nutritional statuses based on BMI-for-age z-scores (healthy and overweight/obese) and height-for-age z-scores (stunted), measured in Wave 5. Covariates are measured pre-treatment, and the list is provided in Appendix 4.A.4. DML: random forest with 3-fold cross-fitting and a trimming threshold of 0.01. Standard errors are based on asymptotic approximations using the estimated variance of the efficient score functions. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$

4.6.2 Mechanisms for non-foster children in sending households

Parallel to our analysis of receiving households, we examine the mechanisms at play in sending households through the study of mediators (Table 4.E.4). The results reveal that a primary channel operates through changes in household composition: fostering leads to a reduction in both household size and the number of children aged 0-14. By lowering the number of dependents, fostering reduces the number of mouths to feed and alleviates resource constraints, potentially freeing up resources. However, the reduction in household size and number of children is not equivalent to the departure of one child. Despite this, it seems that out-fostering a child serves as a mechanism to regulate household composition. Following Cotton (2024), fostering may be seen as a strategy for reducing the burdens of childrearing, thus enabling non-foster children to receive greater attention in terms of health and nutrition and face less competition for resources. Additionally, while we observe a decline in the wealth index, food expenditures remain stable, indicating that resources allocated to food may be preserved.

Interestingly, we observe no change in remittances sent by the origin household following the fostering, and we cannot confirm whether these remittances are directed to the host household. Our findings in Table 4.6 show that hosting households benefit from additional remittances, leaving open the question of whether this increase is a result of fostering. Given that the origin household does not send remittances

[29] Unfortunately, we cannot assess the impact on non-foster children in receiving households. Descriptive statistics show that foster children typically enter households with few, if any, host children, resulting in a limited sample size (88 individuals), which is insufficient for robust estimates. This also suggests likely no spillover effects for this group, as foster children generally arrive alone in host households.

for fostering and it is unlikely that another household would contribute financially for this fostering, this is an indication that the host household may have been selected based on its ability to access cash resources (such as pensions and remittances), which, in turn, supports improved nutrition for the child.

4.7 Conclusion

This paper investigates the impact of child fostering on children's nutritional outcomes in South Africa. We both leverage the unique characteristics of the NIDS longitudinal dataset and employ advanced machine learning techniques. The analysis takes advantage of two unique features of the NIDS: the ability to track individuals over time and a survey design that intensively follows specific predefined individuals, mitigating non-observability issues post-treatment. Additionally, by employing a double machine learning approach, which accounts for selection into fostering and outcome observability, our research addresses critical limitations in the existing literature.

Our findings provide evidence that child fostering through relocation significantly improves children's nutritional outcomes in South Africa, particularly by reducing stunting. This improvement in long-term health outcomes is achieved through two main mechanisms: changes in living environments and caregiving arrangements. Foster children are often placed in smaller, rural households with older caregivers, frequently grandparents, who are more likely to receive stable old-age pension income. Furthermore, foster children are often directly cared for by household heads, which likely improves their intra-household resource allocation. Importantly, fostering appears to create a mutually beneficial dynamic: children remaining in the origin household also experience improved nutritional outcomes, primarily due to reduced household size and decreased competition for resources.

These findings highlight fostering as a strategy with the potential to enhance child health in both sending and receiving households. However, it also raises an important question: what is the impact on caregivers who host foster children? Could the positive outcomes for children come at the expense of the caregivers' own health and nutrition? In other words, might caregivers be sacrificing their own well-being to prioritize the nutrition of foster children? These questions present intriguing avenues for future research.

References of Chapter 4

Ainsworth, M. (1996): "Economic aspects of child fostering in Cote d'Ivoire," *Research in Population Economics*, 8, 25–62.

Akresh, R. (2004): "Adjusting Household Structure: School Enrollment Impacts of Child Fostering in Burkina Faso," *SSRN Electronic Journal*.

——— (2005): "Risk, Network Quality, and Family Structure: Child Fostering Decisions in Burkina Faso," *SSRN Electronic Journal*.

——— (2009): "Flexibility of Household Structure: Child Fostering Decisions in Burkina Faso," *Journal of Human Resources*, 44, 976–997.

Alber, E. (2013): "The transfer of belonging: theories on child fostering in West Africa reviewed," in *Child Fostering in West Africa*, ed. by E. Alber, J. Martin, and C. Notermans, Leiden, Boston: Brill, 79–107.

Angrist, J. and J.-S. Pischke (2009): *Mostly Harmless Econometrics: An Empiricist's Companion*.

Ariyo, E., D. Mortelmans, and E. Wouters (2019): "The African child in kinship care: A systematic review," *Children and Youth Services Review*, 98, 178–187.

Azomahou, T. T., B. Diene, and A. Gosselin-Pali (2022): "Transition and persistence in the double burden of malnutrition and overweight or obesity: Evidence from South Africa," *Food Policy*, 113, 102303.

Badaoui, E. and L. Mangiavacchi (2022): "Assessing the impact of fostering on children's outcomes in Niger," *Economics & Human Biology*, 46, 101121.

Beck, S., P. D. Vreyer, S. Lambert, K. Marazyan, and A. Safir (2015): "Child Fostering in Senegal," *Journal of Comparative Family Studies*, 46, 57–73.

Bia, M., M. Huber, and L. Lafférs (2023): "Double Machine Learning for Sample Selection Models," *Journal of Business and Economic Statistics*.

Bledsoe, C. and U. Isiugo-Abanihe (1989): "Strategies of child fosterage among Mende grannies in Sierra Leone," in *Reproduction and Social Organization in Sub-Saharan Africa*, ed. by R. Lesthaeghe, Berkeley: University of California Press, 442–474.

Bledsoe, C. H., D. C. Ewbank, and U. C. Isiugo-Abanihe (1988): "The effect of child fostering on feeding practices and access to health services in rural Sierra Leone," *Social Science & Medicine*, 27, 627–636.

Bodory, H. and M. Huber (2018): "The causalweight package for causal inference in R," FSES Working Papers 493, Faculty of Economics and Social Sciences, University of Freiburg/Fribourg Switzerland.

Bose-Duker, T., M. Henry, and E. Strobl (2021): "Child fostering and the educational outcomes of Jamaican children," *International journal of educational development*, 87, 102483.

Bramlett, M. and S. Blumberg (2007): "Family structure and children's physical and mental health," *Health Affairs (Millwood)*, 26, 549–558.

Brophy, T., N. Branson, R. C. Daniels, M. Leibbrandt, C. Mlatsheni, and I. Woolard (2018): "National Income Dynamics Study panel user manual," .

Brown, C., M. Ravallion, and D. van de Walle (2019): "Most of Africa's Nutritionally Deprived Women and Children Are Not Found in Poor Households," *The Review of Economics and Statistics*, 101, 631–644.

Castle, S. E. (1995): "Child fostering and children's nutritional outcomes in rural Mali: The role of female status in directing child transfers," *Social Science & Medicine*, 40, 679–693.

Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins (2018): "Double/debiased machine learning for treatment and structural parameters," *The Econometrics Journal*, 21, C1–C68.

Cichello, P. L. (2003): "Child Fostering and Human Capital Formation in KwaZulu-Natal: An Economist's Perspective," *Social Dynamics*, 29, 177–212.

Cotton, C. (2021): "An Enduring Institution? Child Fostering in Sub-Saharan Africa," *Population and Development Review*, 47, 1179–1206.

——— (2024): "Child Fostering and Family Size Preferences in Uganda," *Studies in Family Planning.*

De Vreyer, P. and S. Lambert (2021): "Inequality, Poverty, and the Intra-Household Allocation of Consumption in Senegal," *The World Bank Economic Review*, 35, 414–435.

Duflo, E. (2003): "Grandmothers and Granddaughters: Old–Age Pensions and Intrahousehold Allocation in South Africa," *The World Bank Economic Review*, 17, 1–25.

Démurger, S. (2015): "Migration and families left behind," *IZA World of Labor.*

Eloundou-Enyegue, P. M. and D. Shapiro (2004): "Buffering inequalities: The safety net of extended families in Cameroon," *Cornell Food and Nutrition Policy Program Working Paper*, 177.

Fafchamps, M. and A. R. Quisumbing (2008): "Household formation and marriage markets in rural areas," *Handbook of development economics*, 4, 3187–3247.

Gautrain, E. A. (2023): "Three Essays on Family and Migration Economics," phdthesis, Université Clermont Auvergne.

Gosselin-Pali, A. (2025): "Migration and Nutrition of the Left Behind: Evidence from Ghana," *World Development*, forthcoming.

Govender, L., K. Pillay, M. Siwela, A. Modi, and T. Mabhaudhi (2017): "Food and Nutrition Insecurity in Selected Rural Communities of KwaZulu-Natal, South Africa — Linking Human Nutrition and Agriculture," *International Journal of Environmental Research and Public Health*, 14, 17.

Grant, M. J. and S. Yeatman (2014): "The Impact of Family Transitions on Child Fostering in Rural Malawi," *Demography*, 51, 205–228.

Hamilton, W. D. (1964): "The genetical evolution of social behaviour. I," *Journal of Theoretical Biology*, 7, 1–16.

Hamoudi, A. and D. Thomas (2014): "Endogenous coresidence and program incidence: South Africa's old age pension," *Journal of Development Economics*, 109, 30–37.

Hampshire, K., G. Porter, S. Agblorti, E. Robson, A. Munthali, and A. Abane (2015): "Context matters: fostering, orphanhood and schooling in sub-Saharan Africa," *Journal of Biosocial Science*, 47, 141–164.

Hawkes, C. (2008): "Dietary Implications of Supermarket Development: A Global Perspective," *Development Policy Review*, 26, 657–692.

Hoddinott, J., J. R. Behrman, J. A. Maluccio, P. Melgar, A. R. Quisumbing, M. Ramirez-Zea, A. D. Stein, K. M. Yount, and R. Martorell (2013): "Adult consequences of growth failure in early childhood," *The American Journal of Clinical Nutrition*, 98, 1170–1178.

Huber, M. (2014): "Treatment Evaluation in the Presence of Sample Selection," *Econometric Reviews*, 33, 869–905.

Isiugo-Abanihe, U. C. (1985): "Child fosterage in west Africa," *Population and development review*, 53–73.

Kruger, H. S., M. Visser, L. Malan, L. Zandberg, M. Wicks, C. Ricci, and M. Faber (2023): "Anthropometric nutritional status of children (0-18 years) in South Africa 1997-2022: a systematic review and meta-analysis," *Public Health Nutrition*, 26, 2226–2242.

Madhavan, S. and N. Townsend (2007): "The social context of children's nutritional status in rural South Africa," *Scandinavian Journal of Public Health. Supplement*, 69, 107–117.

Marazyan, K. (2011): "Effects of a Sibship Extension to Foster Children on Children's School Enrolment: A Sibling Rivalry Analysis for Indonesia," *The Journal of Development Studies*, 47, 497–518.

——— (2015): "Foster Children's Educational Outcomes when Child Fostering is Motivated by Norms: The Case of Cameroon," *Journal of Comparative Family Studies*, 46, 85–103.

McGovern, M. E., A. Krishna, V. M. Aguayo, and S. Subramanian (2017): "A review of the evidence linking child stunting to economic outcomes," *International Journal of Epidemiology*, 46, 1171–1191.

Popkin, B. M. (1999): "Urbanization, Lifestyle Changes and the Nutrition Transition," *World Development*, 27, 1905–1916.

——— (2001): "The Nutrition Transition and Obesity in the Developing World," *The Journal of Nutrition*, 131, 871S–873S.

Prall, S. P. and B. A. Scelza (2017): "Child fosterage and sex-biased nutritional outcomes among Namibian pastoralists," *American Journal of Human Biology*, 29, e23058.

Reardon, T., D. Tschirley, L. S. O. Liverpool-Tasie, T. Awokuse, J. Fanzo, B. Minten, R. Vos, M. Dolislager, C. Sauer, R. Dhar, C. Vargas, A. Lartey, A. Raza, and B. M. Popkin (2021): "The processed food revolution in African food systems and the double burden of malnutrition," *Global Food Security*, 28, 100466.

Ronnkvist, S. R., B. C. Thiede, and E. Barber (2023): "Child fostering in a changing climate: evidence from sub-Saharan Africa," *Population and Environment*, 45, 29.

Rothe, C. and S. Firpo (2013): "Semiparametric Estimation and Inference Using Doubly Robust Moment Conditions," IZA Discussion Paper No. 7564.

Rubin, D. B. (1974): "Estimating causal effects of treatments in randomized and nonrandomized studies," *Journal of Educational Psychology*, 66, 688–701.

Schrijner, S. and J. Smits (2018): "Grandparents and Children's stunting in sub-Saharan Africa," *Social Science & Medicine*, 205, 90–98.

Serra, R. (2009): "Child fostering in Africa: When labor and schooling motives may coexist," *Journal of Development Economics*, 88, 157–170.

Shekar, M. and B. Popkin (2020): *Obesity: Health and economic consequences of an impending global challenge*, The World Bank.

Stewart, C. P., L. Iannotti, K. G. Dewey, K. F. Michaelsen, and A. W. Onyango (2013): "Contextualising complementary feeding in a broader framework for stunting prevention," *Maternal & Child Nutrition*, 9 Suppl 2, 27–45.

WHO (2006): *WHO child growth standards: length/height-for-age, weight-for-age, weight-for-length, weight-for-height and body mass index-for-age ; methods and development*, Geneva: WHO Press.

WHO and UNICEF (2019): *Recommendations for data collection and analysis and reporting on anthropometric indicators in children under 5 years old*, Geneva.

Zimmerman, F. J. (2003): "Cinderella Goes to School: The Effects of Child Fostering on School Enrollment in South Africa," *The Journal of Human Resources*, 38, 557–590.

Appendix to Chapter 4

4.A Additional descriptive statistics

Table 4.A.1: Baseline differences between foster and non-foster children within households that will foster, child's characteristics

	(1) Foster	(2) Non-foster	(1)-(2)
Female	0.553	0.506	0.047
	(0.030)	(0.028)	(0.041)
Age	5.546	5.500	0.046
	(0.217)	(0.203)	(0.297)
Coresident mother	1.000	0.758	0.242***
	(0.000)	(0.024)	(0.025)
Birth order	2.201	2.500	-0.299**
	(0.100)	(0.089)	(0.135)
Born out of wedlock	0.683	0.661	0.022
	(0.028)	(0.030)	(0.041)
Child of the head	0.553	0.365	0.188***
	(0.030)	(0.027)	(0.040)
Zbmi	0.328	0.159	0.170
	(0.085)	(0.075)	(0.114)
Overweight/Obese	0.271	0.223	0.048
	(0.026)	(0.025)	(0.037)
Healthy	0.687	0.745	-0.058
	(0.028)	(0.026)	(0.038)
Zhfa	-0.873	-1.099	0.226**
	(0.085)	(0.075)	(0.113)
Stunted	0.187	0.201	-0.015
	(0.023)	(0.024)	(0.033)
Ill/disable	0.056	0.054	0.002
	(0.014)	(0.013)	(0.019)
Covered by medical aid	0.053	0.035	0.018
	(0.013)	(0.010)	(0.017)
Poor subjective health status	0.007	0.003	0.004
	(0.005)	(0.003)	(0.006)
Fair subjective health status	0.035	0.016	0.019
	(0.011)	(0.007)	(0.013)
Good subjective health status	0.148	0.115	0.033
	(0.021)	(0.018)	(0.028)
Very good subjective health status	0.278	0.447	-0.169***
	(0.027)	(0.028)	(0.039)
Father alive	0.905	0.871	0.034
	(0.017)	(0.019)	(0.026)
Observations	284	318	602

Notes: Sample of children aged 0-12 years at baseline (Wave 4). Foster are children who will move out from the parental household at Wave 5. The 318 non-foster children are other children observed in sending households at Wave 4. For those children there are between 3 and 40 missing values in each variable except age and female. Birth order and born out of wedlock can be defined only if the mother is coresident. Baseline corresponds to Wave 4 of the NIDS. Standard errors are in parentheses. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table 4.A.2: Non random selection, child's characteristics

	(1) Observed	(2) Non observed	(1)-(2)
Foster	0.037	0.083	-0.046***
	(0.003)	(0.008)	(0.007)
Female	0.506	0.529	-0.023
	(0.007)	(0.015)	(0.017)
Age	5.931	5.297	0.634***
	(0.049)	(0.111)	(0.119)
Birth order	2.417	2.102	0.315***
	(0.023)	(0.042)	(0.054)
Born out of wedlock	0.584	0.582	0.002
	(0.007)	(0.015)	(0.016)
Ethnicity: African	0.833	0.823	0.010
	(0.005)	(0.012)	(0.012)
Zbmi	0.288	0.357	-0.070
	(0.018)	(0.039)	(0.043)
Overweight/Obese	0.263	0.294	-0.031**
	(0.006)	(0.014)	(0.015)
Healthy	0.705	0.671	0.034**
	(0.006)	(0.014)	(0.015)
Zhfa	-0.834	-0.903	0.069*
	(0.017)	(0.043)	(0.042)
Stunted	0.150	0.189	-0.039***
	(0.005)	(0.012)	(0.012)
Ill/disable	0.038	0.043	-0.005
	(0.003)	(0.006)	(0.006)
Covered by medical aid	0.082	0.112	-0.030***
	(0.004)	(0.010)	(0.009)
Poor subjective health status	0.003	0.004	-0.001
	(0.001)	(0.002)	(0.002)
Fair subjective health status	0.018	0.022	-0.004
	(0.002)	(0.004)	(0.004)
Good subjective health status	0.157	0.153	0.004
	(0.005)	(0.011)	(0.012)
Very good subjective health status	0.326	0.313	0.013
	(0.006)	(0.014)	(0.016)
Father alive	0.917	0.939	-0.022**
	(0.004)	(0.007)	(0.009)
Father and paternal grandparent coresident	0.036	0.047	-0.011*
	(0.003)	(0.006)	(0.006)
Paternal grandparents alive	0.127	0.109	0.018
	(0.005)	(0.009)	(0.011)
Mother and maternal grandparent coresident	0.327	0.286	0.040***
	(0.006)	(0.014)	(0.015)
Maternal grandparents alive	0.401	0.395	0.005
	(0.007)	(0.015)	(0.016)
Observations	5,240	1,096	6,336

Notes: We recall that selection is defined for all reasons of outcome non-observability, including when there are missing values for the outcome variables even though the individual is tracked in wave 5. In this regard, the selection variable is here constructed using information on the missing values for stunting. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table 4.A.3: Non random selection, mother's and household's characteristics

	(1) Observed	(2) Non observed	(1)-(2)
Mother: age	32.802	31.069	1.733***
	(0.110)	(0.233)	(0.263)
Mother: nb years in union	1.093	1.099	-0.007
	(0.051)	(0.104)	(0.121)
Mother: separated/divorced	0.015	0.017	-0.002
	(0.002)	(0.004)	(0.004)
Mother: widow	0.037	0.019	0.018***
	(0.003)	(0.004)	(0.006)
Mother: primary educated	0.091	0.045	0.046***
	(0.004)	(0.006)	(0.009)
Mother: secondary educated	0.700	0.709	-0.009
	(0.006)	(0.014)	(0.015)
Mother: tertiary educated	0.178	0.221	-0.043***
	(0.005)	(0.013)	(0.013)
Mother: diabetic	0.006	0.002	0.004
	(0.001)	(0.001)	(0.002)
Mother: HIV	0.041	0.037	0.004
	(0.003)	(0.006)	(0.007)
Mother: decides about expenses	0.511	0.466	0.045***
	(0.007)	(0.015)	(0.017)
Mother: poor subjective health status	0.013	0.010	0.003
	(0.002)	(0.003)	(0.004)
Mother: fair subjective health status	0.053	0.042	0.011
	(0.003)	(0.006)	(0.007)
Mother: good subjective health status	0.289	0.259	0.030**
	(0.006)	(0.013)	(0.015)
Mother: very good subjective health status	0.334	0.339	-0.005
	(0.007)	(0.014)	(0.016)
Rural	0.495	0.455	0.039**
	(0.007)	(0.015)	(0.017)
Household size	6.681	6.902	-0.221*
	(0.047)	(0.103)	(0.113)
Head: female	0.665	0.630	0.034**
	(0.007)	(0.015)	(0.016)
Head: age	45.877	44.889	0.988**
	(0.204)	(0.465)	(0.495)
Head: primary educated	0.177	0.162	0.014
	(0.005)	(0.011)	(0.013)
Head: secondary educated	0.536	0.520	0.016
	(0.007)	(0.015)	(0.017)
Head: tertiary educated	0.146	0.176	-0.030**
	(0.005)	(0.012)	(0.012)
Wealth index [0-1]	0.542	0.559	-0.017***
	(0.003)	(0.006)	(0.006)
Total expenses per adult equivalent	1,784.219	2,326.442	-542.223***
	(35.915)	(138.798)	(100.969)
Household with retired pension	0.252	0.246	0.006
	(0.006)	(0.013)	(0.014)
Agricultural household	0.232	0.228	0.004
	(0.006)	(0.013)	(0.014)
Household received remittances	0.278	0.275	0.003
	(0.006)	(0.013)	(0.015)
Household sent remittances	0.128	0.127	0.001
	(0.005)	(0.010)	(0.011)
Distance to police station	2.642	2.558	0.084**
	(0.015)	(0.033)	(0.036)
Negative household income shock last 2 years	0.108	0.110	-0.002
	(0.004)	(0.009)	(0.010)
Observations	5,240	1,096	6,336

Notes: We recall that selection is defined for all reasons of outcome non-observability, including when there are missing values for the outcome variables even though the individual is tracked in wave 5. In this regard, the selection variable is here constructed using information on the missing values for stunting. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table 4.A.4: List of pre-treatment covariates used in the double machine learning

Child's characteristics	Gender (female)
	Age (dummy variables for each age from 0 to 12 years)
	Child of the head
	Born out of wedlock
	Ethnic group (African)
	Birth order
	Perceived health status (dummy variables from 1 to 5)
	Child already had serious illnesses or disabilities
	Child is covered by medical aid
	Body Mass Index z-score
	Height-for-age z-score
Parents' characteristics	Alive father
	Coresidence with father
	Coresidence with maternal grandparents
	Coresidence with paternal grandparents
	Coresidence with father and alive paternal grandparent
	Coresidence with mother and alive maternal grandparent
	Mother's age
	Mother's marital status (dummy variables for being a widow, or divorced/separated)
	Mother's number of years of union
	Mother's education level (dummy variables: uneducated, primary, secondary, tertiary educated)
	Mother's perceived health status (dummy variables from 1 to 5)
	Mother diagnosed with diabetes
	Mother diagnosed with HIV
	The mother is in charge of decisions about household expenditures.
Household characteristics	Household size
	Household: Rural
	Household: Female head
	Household: Head's age
	Household: Wealth index (dummy variables for each quintile of a standardized wealth index)
	Household: Total expenditures (food and non-food) per adult equivalent
	Household: District of residence (dummy variables for each district from 1 to 52)
	Household: Region
	Household: Head's education level (dummy variables: uneducated, primary, secondary, tertiary educated)
	Household with a pensioner receiving the old age grant
	Agricultural household
	Household: Distance to the police station
	Household has experienced a negative shock over 2 years
	Remittances recipient household (internal/intra-province/abroad, family/non-family, head/other members)
	Remittances sending household (internal/intra-province/abroad, family/non-family, head/other members)

4.B Additional materials for the empirical strategy

Table 4.B.1: Procedure for estimating ATE with DML for sample selection models

Step 1	**Estimate propensity scores using machine learning**
	a. Split the dataset into two parts: a training set (2/3) and an estimation set (1/3) to implement cross-fitting and prevent overfitting.
	b. Estimate the propensity scores for observability, where observability is modeled as a function $\pi(d, X, Z)$, predicting the probability of being observed.
	c. Estimate the propensity scores for treatment, conditional on the covariates X and the estimated propensity scores for observability (which indirectly reflect Z).
	d. Using observations where the outcome is available, train a model to predict potential outcomes in the test set, based on both the covariates and treatment status.
Step 2	**Trim observations below threshold**
	Discard observations where the product of the propensity scores for observability and treatment falls below a specified threshold.
Step 3	**Estimate outcomes using machine learning**
	For the remaining observations, estimate the outcomes for both treated and control groups, weighting them by their respective propensity scores. Apply machine learning techniques to adjust for covariates in this estimation.
Step 4	**Calculate the average treatment effect (ATE)**
	Compute the ATE as the difference between the weighted outcomes of the treated and control groups.

4.C Additional results

Table 4.C.1: ATE estimates based on the DML procedure, excluding sending and host children

Outcome	ATE	Standard error	p-value	Baseline mean	N
Healthy	0.088***	0.029	0.002	0.698	5,940
Overweight/Obese	-0.063**	0.028	0.023	0.269	5,940
Stunted	-0.060***	0.022	0.005	0.156	5,940

Notes: Sample of children aged 0-12 years at baseline (Wave 4). The treatment is fostering. The outcome variables are nutritional statuses based on BMI-for-age z-scores (healthy and overweight/obese) and height-for-age z-scores (stunted), and are measured in Wave 5. Covariates are measured pre-treatment, and the list is provided in Appendix 4.A.4. In this sample, non-fostered children from both sending and receiving households are excluded from the control group. DML: random forest with 3-fold cross-fitting and a trimming threshold of 0.01. Standard errors are based on asymptotic approximations using the estimated variance of the efficient score functions. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$

Table 4.C.2: ATE estimates based on the DML procedure, excluding host children only

Outcome	ATE	Standard error	p-value	Baseline mean	N
Healthy	0.044	0.038	0.257	0.699	6,248
Overweight/Obese	-0.048	0.032	0.132	0.268	6,248
Stunted	-0.057***	0.019	0.002	0.156	6,248

Notes: Sample of children aged 0-12 years at baseline (Wave 4). The treatment is fostering. The outcome variables are nutritional statuses based on BMI-for-age z-scores (healthy and overweight/obese) and height-for-age z-scores (stunted), and are measured in Wave 5. Covariates are measured pre-treatment, and the list is provided in Appendix 4.A.4. In this sample, non-fostered children from receiving households are excluded from the control group. DML: random forest with 3-fold cross-fitting and a trimming threshold of 0.01. Standard errors are based on asymptotic approximations using the estimated variance of the efficient score functions. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$

Table 4.C.3: ATE estimates based on the DML procedure, excluding sending children only

Outcome	ATE	Standard error	p-value	Baseline mean	N
Healthy	0.044	0.037	0.230	0.697	5,997
Overweight/Obese	-0.030	0.035	0.390	0.270	5,997
Stunted	-0.052***	0.018	0.003	0.157	5,997

Notes: Sample of children aged 0-12 years at baseline (Wave 4). The treatment is fostering. The outcome variables are nutritional statuses based on BMI-for-age z-scores (healthy and overweight/obese) and height-for-age z-scores (stunted), and are measured in Wave 5. Covariates are measured pre-treatment, and the list is provided in Appendix 4.A.4. In this sample, non-fostered children from sending households are excluded from the control group. DML: random forest with 3-fold cross-fitting and a trimming threshold of 0.01. Standard errors are based on asymptotic approximations using the estimated variance of the efficient score functions. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$

Table 4.C.4: ATE estimates based on the DML procedure, children aged 2-12 at baseline

Outcome	ATE	Standard error	p-value	Baseline mean	N
Healthy	0.022	0.047	0.632	0.739	5,431
Overweight/Obese	-0.019	0.042	0.644	0.231	5,431
Stunted	-0.062***	0.019	0.001	0.133	5,431

Notes: Sample of children aged 2-12 years at baseline (Wave 4). The treatment is fostering. The outcome variables are nutritional statuses based on BMI-for-age z-scores (healthy and overweight/obese) and height-for-age z-scores (stunted), and are measured in Wave 5. Covariates are measured pre-treatment, and the list is provided in Appendix 4.A.4. DML: random forest with 3-fold cross-fitting and a trimming threshold of 0.01. Standard errors are based on asymptotic approximations using the estimated variance of the efficient score functions. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$

Table 4.C.5: ATE estimates based on the DML procedure, trimming threshold of 0.02

Outcome	ATE	Standard error	p-value	Baseline mean	N
Healthy	0.027	0.036	0.447	0.699	6,336
Overweight/Obese	-0.006	0.034	0.869	0.268	6,336
Stunted	-0.055**	0.024	0.022	0.156	6,336

Notes: Sample of children aged 0-12 years at baseline (Wave 4). The treatment is fostering. The outcome variables are nutritional statuses based on BMI-for-age z-scores (healthy and overweight/obese) and height-for-age z-scores (stunted), and are measured in Wave 5. Covariates are measured pre-treatment, and the list is provided in Appendix 4.A.4. DML: random forest with 3-fold cross-fitting and a trimming threshold of 0.02. Standard errors are based on asymptotic approximations using the estimated variance of the estimated efficient score functions. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$

Table 4.C.6: ATE estimates based on the DML procedure, first placebo test between Wave 3 and Wave 4, treatment in Wave 5

Outcome	ATE	Standard error	p-value	Baseline mean	N
Healthy	-0.109	0.084	0.192	0.609	2,933
Overweight/Obese	0.096	0.075	0.199	0.342	2,933
Stunted	0.024	0.059	0.682	0.212	2,933

Notes: Sample of children aged 0-12 years at baseline (Wave 4). The treatment is fostering. The outcome variables are nutritional statuses based on BMI-for-age z-scores (healthy and overweight/obese) and height-for-age z-scores (stunted), measured in Wave 4. Covariates are measured pre-treatment, and the list is provided in Appendix 4.A.4, except lagged nutritional outcomes. DML: random forest with 3-fold cross-fitting and a trimming threshold of 0.01. Standard errors are based on asymptotic approximations using the estimated variance of the efficient score functions. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$

Table 4.C.7: ATE estimates based on the DML procedure, second placebo test: randomly assigned treatment

Outcome	ATE	Standard error	p-value	Baseline mean	N
Healthy	-0.006	0.034	0.850	0.699	6,336
Overweight/Obese	0.022	0.033	0.499	0.268	6,336
Stunted	0.040	0.030	0.189	0.156	6,336

Notes: Sample of children aged 0-12 years at baseline (Wave 4). The treatment is assigned at random. The outcome variables are nutritional statuses based on BMI-for-age z-scores (healthy and overweight/obese) and height-for-age z-scores (stunted), measured in Wave 4. Covariates are measured pre-treatment, and the list is provided in Appendix 4.A.4. DML: random forest with 3-fold cross-fitting and a trimming threshold of 0.01. Standard errors are based on asymptotic approximations using the estimated variance of the efficient score functions. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$

Table 4.C.8: ATE estimates based on the DML procedure, with fostering defined to also include left-behind children

Outcome	ATE	Standard error	p-value	Baseline mean	N
Healthy	0.038*	0.021	0.078	0.701	6703
Overweight/Obese	-0.032	0.020	0.103	0.267	6703
Stunted	-0.028**	0.013	0.038	0.158	6703

Notes: Sample of children aged 0-12 years at baseline (Wave 4). The treatment is fostering and left-behind children, i.e. these are children that are not coresiding with any parent at Wave 5 and have at least a parent alive, while they were co-resident with at least one parent in Wave 4, disregarding of the change of household unit. The outcome variables are nutritional statuses based on BMI-for-age z-scores (healthy and overweight/obese) and height-for-age z-scores (stunted), and are measured in Wave 5. Covariates are measured pre-treatment, and the list is provided in Appendix 4.A.4. In this sample, non-fostered children from sending households are excluded from the control group. DML: random forest with 3-fold cross-fitting and a trimming threshold of 0.01. Standard errors are based on asymptotic approximations using the estimated variance of the efficient score functions. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$

4.D Alternative methods

Table 4.D.1: OLS estimates of the impact of child fostering

	(1)	(2)	(3)	(4)
Outcome: Healthy				
Foster	-0.001	-0.001	0.008	0.015
	(0.037)	(0.040)	(0.041)	(0.041)
Average outcome	0.752	0.752	0.752	0.752
Observations	8,096	8,096	8,096	8,096
Outcome: Overweight/Obese				
Foster	0.014	0.009	-0.006	-0.013
	(0.035)	(0.038)	(0.038)	(0.038)
Average outcome	0.215	0.215	0.215	0.215
Observations	8,096	8,096	8,096	8,096
Outcome: Stunted				
Foster	-0.007	-0.025	-0.009	-0.011
	(0.029)	(0.031)	(0.031)	(0.031)
Average outcome	0.127	0.127	0.127	0.127
Observations	8,136	8,136	8,136	8,136
Exogenous child's characteristics	Yes	Yes	Yes	Yes
Endogenous child's characteristics	No	Yes	Yes	Yes
Endogenous household's characteristics	No	No	Yes	Yes
District FE	No	No	No	Yes

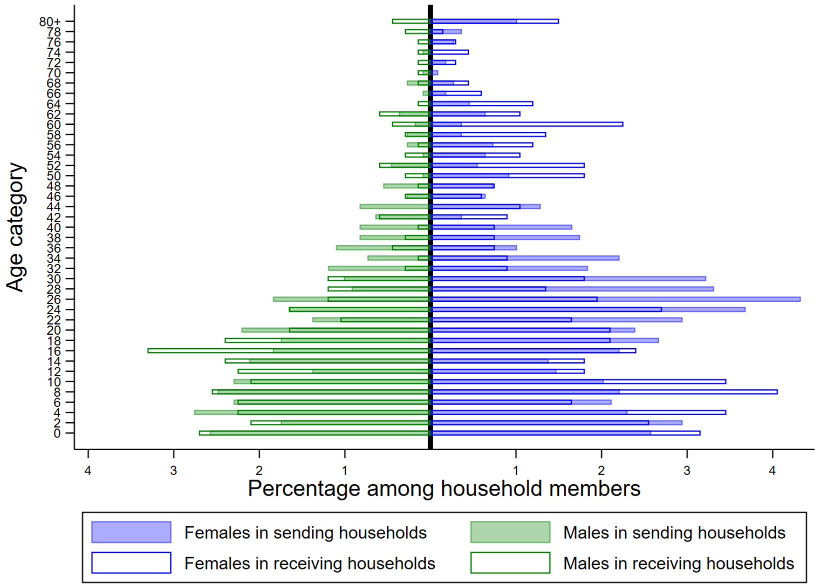
Notes: The sample includes children aged 2-14 in Wave 5, for whom the outcome and the covariates are not missing. Foster is a dummy equal to one if children are not coresiding with any parent at Wave 5 and have at least a parent alive. The outcome variables are nutritional statuses based on BMI-for-age z-scores (healthy and overweight/obese) and height-for-age z-scores (stunted), and are measured in Wave 5. Covariates are measured in Wave 5. Exogenous child's characteristics include age dummies, gender, birth order, born out of wedlock dummy, African dummy. Endogenous child's characteristics include dummies about relationship to the head, illness/disability, covered by medical aid, and father alive. Endogenous household's characteristics include a rural dummy, household size, female head, head's age, head's education, dummies of wealth quintile, total expenditure per adult equivalent, distance to the nearest police station, dummies indicating if the household received a retired pension, remittances, sent remittances, is practising agriculture, and has faced a negative income shocks over the two past years. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$

Table 4.D.2: Difference-in-differences estimates of the impact of child fostering

	(1)	(2)	(3)	(4)
Outcome: Healthy				
Foster	0.003	0.002	0.000	-0.004
	(0.032)	(0.031)	(0.032)	(0.032)
Post	0.013*	0.013*	0.014*	0.012*
	(0.007)	(0.007)	(0.008)	(0.008)
Foster*Post	0.003	-0.012	-0.008	-0.005
	(0.039)	(0.039)	(0.039)	(0.040)
Average outcome	0.736	0.736	0.736	0.736
Observations	10,083	10,083	10,083	10,083
Outcome: Overweight/Obese				
Foster	-0.013	-0.012	-0.011	-0.006
	(0.030)	(0.030)	(0.030)	(0.030)
Post	-0.015**	-0.014**	-0.015**	-0.014**
	(0.007)	(0.007)	(0.007)	(0.007)
Foster*Post	0.013	0.024	0.016	0.014
	(0.035)	(0.036)	(0.036)	(0.036)
Average outcome	0.230	0.230	0.230	0.230
Observations	10,083	10,083	10,083	10,083
Outcome: Stunted				
Foster	0.033	0.031	0.023	0.034
	(0.024)	(0.024)	(0.024)	(0.025)
Post	-0.001	-0.000	-0.001	-0.002
	(0.006)	(0.006)	(0.006)	(0.006)
Foster*Post	-0.063**	-0.067**	-0.045	-0.054*
	(0.029)	(0.029)	(0.030)	(0.030)
Average outcome	0.133	0.133	0.133	0.133
Observations	10,083	10,083	10,083	10,083
Exogenous child's characteristics	Yes	Yes	Yes	Yes
Endogenous child's characteristics	No	Yes	Yes	Yes
Endogenous household's characteristics	No	No	Yes	Yes
District FE	No	No	No	Yes

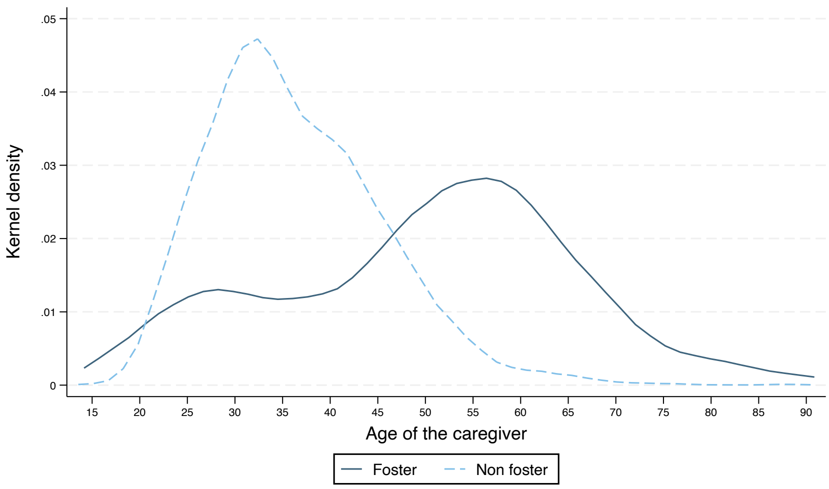
Notes: The sample includes children who were interviewed in Wave 4 and successfully re-interviewed in Wave 5. It includes children aged 2-14 in Wave 5, for whom the outcome and the covariates are not missing. Foster is a dummy equal to one if children are not coresiding with any parent at Wave 5 and have at least a parent alive. The outcome variables are nutritional statuses based on BMI-for-age z-scores (healthy and overweight/obese) and height-for-age z-scores (stunted), and are measured in Waves 4 & 5. Covariates are also measured in Waves 4 & 5. Exogenous child's characteristics include age dummies, gender, birth order, born out of wedlock dummy, African dummy. Endogenous child's characteristics include dummies about relationship to the head, illness/disability, covered by medical aid, and father alive. Endogenous household's characteristics include a rural dummy, household size, female head, head's age, head's education, dummies of wealth quintile, total expenditure per adult equivalent, distance to the nearest police station, dummies indicating if the household received a retired pension, remittances, sent remittances, is practising agriculture, and has faced a negative income shocks over the two past years. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$

4.E Discussion materials



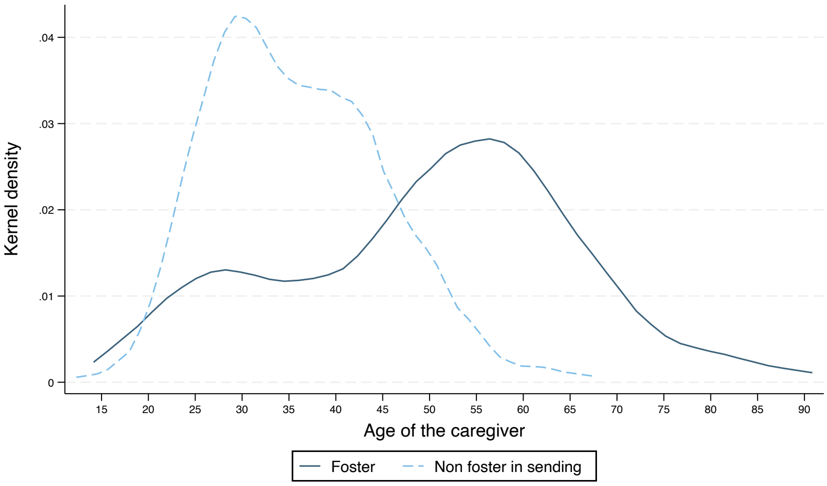
Notes: Sample of households members living with a foster children in Wave 4 (sending households) and living with a foster children in Wave 5 (receiving households). The graph bars show the proportion of individual per age category and gender in each type of households (sending and receiving). For sending households the household composition considered is the one of Wave 4 and for receiving households the one that prevails in Wave 5.

Figure 4.E.1: Distribution of the household members by age and gender in sending and receiving households



Notes: Sample of children aged 0-12 at baseline (Wave 4). The age of the caregiver is measured in Wave 5 for children who are observed.

Figure 4.E.2: Age distribution of caregivers for foster and non-foster children in Wave 5



Notes: Sample of children aged 0-12 at baseline (Wave 4), who are either foster or live in a sending households in Wave 5. The age of the caregiver is measured in Wave 5 for children who are observed.

Figure 4.E.3: Age distribution of caregivers for foster and non-foster (in sending households) children in Wave 5

Table 4.E.3: Relationship to the head of the caregivers in Wave 5

	(1)	(2)	(1)-(2)
	Foster	Non foster	
Head or spouse/partner of the head	0.866	0.621	-0.244***
	(0.024)	(0.007)	(0.035)
Child(-in-law) of the head	0.030	0.255	0.225***
	(0.012)	(0.006)	(0.031)
Parent(-in-law) of the head	0.010	0.007	-0.003
	(0.007)	(0.001)	(0.006)
Other family member of the head	0.055	0.099	0.044**
	(0.016)	(0.004)	(0.021)
Other non-family member	0.000	0.001	0.001
	(0.000)	(0.001)	(0.003)
Missing or absent of the household	0.090	0.019	-0.070***
	(0.020)	(0.002)	(0.010)
Observations	201	5,159	5,360

Notes: Sample of caregivers for children aged 0-12 years at baseline (Wave 4), observed in Wave 5. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table 4.E.4: ATE estimates based on the DML procedure, sending children as the treated group and alternative outcomes

Outcome	ATE	Standard error	p-value	Baseline mean	N
Household size	-0.415**	0.201	0.039	6.718	5,964
Nb of household members 0-14	-0.350***	0.121	0.004	3.110	5,964
Nb of household members 15-64	-0.167	0.105	0.111	3.319	5,964
Nb of household members 65+	0.017	0.027	0.528	0.195	5,964
Wealth index	-0.032**	0.013	0.012	0.534	5,964
Food expenditures	-10.570	171.124	0.951	1340.490	5,964
Expenditures	-684.629	810.764	0.398	5103.980	5,964
Food expenditures (per capita)	15.475	31.736	0.626	234.466	5,964
Expenditures (per capita)	-131.085	153.233	0.392	953.529	5,964
Food expenditures (per adult equivalent)	17.740	66.084	0.788	477.776	5,964
Expenditures (per adult equivalent)	-258.935	320.149	0.419	1910.787	5,964
Food expenditures (per calories adult equivalent)	23.314	40.630	0.566	291.145	5,964
Household with retired pension	-0.006	0.032	0.862	0.252	5,964
Rural	0.001	0.024	0.976	0.488	5,964
Negative household income shock last 2 years	0.055*	0.032	0.084	0.108	5,964
Cared by head or head's spouse	-0.005	0.031	0.860	0.577	5,931
Grandchild of the head	-0.021	0.037	0.580	0.330	5,964
Household received remittances	-0.021	0.040	0.598	0.278	5,964
Household sent remittances	0.058	0.038	0.131	0.126	5,964

Notes: Sample of children aged 0-12 years at baseline (Wave 4). The treated group consists of children from the sending households and the control group consists of non-foster children (excluding host siblings of foster children). Covariates are measured pre-treatment, and the list is provided in Appendix 4.A.4. DML: random forest with 3-fold cross-fitting and a trimming threshold of 0.01. Standard errors are based on asymptotic approximations using the estimated variance of the efficient score functions. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$

Chapter 5

General Conclusion

Malnutrition in all its forms remains a widespread and persistent issue globally, with particularly strong impacts on individuals and households in developing countries. This dissertation, structured around three empirical essays utilizing longitudinal household survey data, investigates the impacts of various economic and social phenomena on nutritional outcomes in sub-Saharan Africa. Overall, it contributes to several strands of development economics, health and nutrition economics, and family and migration economics literature. In particular, this thesis contributes to the literature on the double burden of malnutrition, the effects of migration on left-behind individuals, and the impacts of informal child fostering practices.

The dissertation addresses three main research questions: (i) What are the dynamics of the double burden of malnutrition within South African households, and how is it reflected at the individual level? (ii) How does internal migration in Ghana affect the nutritional status of left-behind individuals? (iii) What are the implications of child fostering in South Africa for the nutritional outcomes of both foster children and those remaining in the household of origin?

The main findings of the second chapter reveal that the double burden of malnutrition is transient at the household level in South Africa. This transitory nature reflects individual-level dynamics, where underweight individuals are unlikely to remain underweight over time. In contrast, overweight/obese individuals tend to stay in that condition for extended periods, contributing to persistence when examined at the household level. Subsequently, the third chapter demonstrates that internal migration in Ghana negatively impacts the nutritional status of left-behind individuals, with children being particularly affected. This negative effect is primarily driven by the short-term disruptive impact of migration, likely resulting from an income shock caused by the migrant's departure, although remittances may offer a positive offset in the longer term. Finally, the last chapter shows that, in South Africa, child fostering significantly reduces the probability of stunting among foster children, likely due to improved living conditions in host households and enhanced caregiving arrangements. Moreover, fostering also improves the nutrition of children remaining in the household of origin, highlighting that it can serve as a mutually beneficial practice for both groups of children.

This work also opens up avenues for future research. In the context of the second chapter, which analyzes nutritional trajectories over a few years, future research could explore these dynamics across lifetimes. In fact, utilizing longer-term data could provide deeper insights into the implications over time. In the third chapter, a limitation is the inability to track individuals who change households due to migration, which often involves further mobility or household dissolution. Future research could investigate the impacts of migration on these relocated individuals and the consequences of their new living arrangements. Finally, the fourth chapter highlights that child fostering can significantly enhance child nutrition in both sending and receiving households. However, it raises critical questions regarding trade-offs faced by caregivers. Understanding whether the benefits for foster children come at the cost of caregivers' health or well-being would shed light on the broader implications of fostering practices and open avenues for further exploration.

All things considered, as long as malnutrition continues to persist in its various forms, studying it remains a pressing priority. Nutrition is a fundamental determinant of human health, and understanding its patterns, drivers, and impacts is crucial for designing effective interventions.