Essays in Development Economics

Niklas Heusch

TESI DOCTORAL UPF / ANY 2018

DIRECTORS DE LA TESI
Robin Hogarth i Alessandro Tarozzi
Departament d'Economia i Empresa

Acknowledgements

I am extremely grateful to my advisors Robin and Alessandro for the guidance and supervision with which they have provided me through the years. Your advice and encouragement has meant a lot to me. I am also very grateful and indebted to Karen Macours for her guidance while I was visiting the Paris School of Economics as a PODER fellow, and to the research groups on development and applied economics at UPF and the PSE from whose comments and suggestions I have benefitted. In Ghana, I am also heavily indebted to Madeleen Husselman, Joseph Nyoagbe, Daniel Amaning Danquah, and a great team of enumerators. Across the pond, I would like to thank the Verneys and Jonathan Weis, the greatest teacher I ever had. On this side of it, I thank Dominik, Friedrich, Clara, Sara, and Theresa, as well as Jan, Marianne, and Andreas: you make me feel like I never left.

I sincerely thank Marta Araque and Laura Agusti for their patience and endless help- the debt I have accumulated to you over the years is only surpassed by my gratitude.

No man is an island- thank you, Ana, Ana, Caterina, Christina, Cristina, Francisco, Giulia, Holger, Lorenzo, María, Marta, Miguel, Patrick, Sebastien, and Ricardo- you have been the continent and have made these years as memorable as they were; it's been a pleasure and a privilege. Marcelo, Maddie, Tom, and Flora, the year I spent in Paris will be hard to forget- I am blessed to have met you. Meryem, I am extremely grateful you missed a housemate; Jared, I am extremely grateful you missed your flight; Marco I am forever grateful you missed the sea: your presence has meant everything and I am sorely feeling your absence.

Divya, from Barcelona to Berlin, and Brooklyn to Nairobi- you have been making my days. Thank you for your joy, your impeccable spelling, also in adverse circumstances, and even sharing foie gras.

Last, but foremost, I thank my parents: you have opened the doors to the world for me, instilled in me a sense of wonder, and given me Tobi and all the love I could ask for.

Abstract

This thesis combines three independent articles. In the first chapter, I examine the health care that drug sellers, common medical providers in many developing countries, provide for childhood illness in Ghana, and study its determinants. Overall, I find the quality of treatment to be poor and provide evidence that this is caused by low knowledge of drug sellers, rather than low effort or adverse financial incentives; a simulation exercise suggests that adequate treatment would not reduce drug sellers' profits or increase clients' expenditures.

In the second chapter, I examine rural-urban migration in Tanzania and provide evidence of substantial selection into urban migration among residents of agricultural households: movers to urban areas are substantially better educated and more commonly participate in formal labour markets prior to moving. However, changes to the economic situation of agricultural households have large impacts on this sorting to urban areas, suggesting that households' ability to finance migration might be an important bottleneck.

In the third chapter, I study the (recently debated) performance of proxy means testing (PMT), an econometric approach deducing households' poverty status from easily collectable information on household characteristics. Brown et al. (2016) criticise the performance of PMT; I find these results to be driven by miscalibration: when calibrated to match the poverty rate of the population, PMT performs substantially better and, although far from perfect, might still provide useful information to its users.

Resumen

Esta tesis está compuesta por tres artículos independientes. El primer capítulo examina la atención sanitaria que los farmacéuticos, proveedores médicos muy comunes en algunos países en vías de desarrollo, proveen para enfermedades infantiles en Ghana y estudia los factores determinantes. Encuentro que la calidad de los tratamientos es baja y muestro evidencia de que está causada por el bajo conocimiento de los farmacéuticos, y no por el bajo esfuerzo de éstos o la presencia de incentivos económicos perversos. Un ejercicio de simulación sugiere que el tratamiento adecuado no reduciría los beneficios de los farmacéuticos ni incrementaría los gastos de los clientes.

En el segundo capítulo, examino la migración rural-urbana en Tanzania y proveo evidencia de la existencia de una selección sustancial en la migración urbana dentro de los residentes de los hogares agrícolas: aquellos que deciden mudarse a áreas urbanas son más educados y tienden a participar más en el mercado laboral antes de mudarse. Sin embargo, cambios en la situación económica de los hogares agrícolas tienen grandes impactos sobre la selección, sugiriendo que la habilidad para financiar la migración que tienen los hogares puede ser un obstáculo importante.

En el tercer capitulo estudio el desempeño (debatido recientemente) del "proxy means testing" (PMT), un método econométrico que establece el estatus de pobreza de los hogares según un conjunto de información sobre las características de los hogares que se obtienen fácilmente. Brown et al. (2016) critican el desempeño del PMT; yo encuentro que estos resultados se deben a una calibración errónea: cuando la calibración se realiza para igualar la tasa de pobreza de la población, el PMT funciona mucho mejor y, aunque no es perfecto, puede seguir proveyendo con información útil a sus usuarios.

Preface

This doctoral thesis combines three self-contained essays on development economics, cutting across the fields of health, migration, and measurement. While independent in nature, the articles are united by all having an empirical approach, a research question relevant to policy, and roots in theoretical literature.

Informal health care providers are common across developing countries and drug sellers provide a large share of treatment for common (and often deadly) childhood illnesses in sub-Saharan Africa. The first chapter combines covert observation with formal surveys to study the quality of medical treatment for childhood illness provided by informal drug sellers in such a setting in Northern Ghana. I find the quality of treatment for four common childhood illnesses (malaria, diarrhoea, respiratory infections, and anaemia) to be poor: drug sellers provide appropriate treatment in only one third of interactions. I subsequently examine knowledge, effort, and financial incentives as determinants of provider behaviour and find that inadequate knowledge, rather than low effort or adverse financial incentives, is a main constraint to better treatment. Profit incentives, on the other hand, do not appear to be a bottleneck: a simulation exercise suggests that providing better treatment would not diminish drug sellers' profits or increase clients' expenditure.

The second chapter studies rural-urban migration. In developing countries, urban residents are more productive and consume more than their rural peers; recent research has suggested that these differences might stem from unobserved heterogeneity in skill between urban and rural residents, brought about by sorting. I study domestic migration flows in Tanzania and document three results consistent with this explanation. Firstly, above-median educated individuals are three times as likely to leave their agricultural households and move to urban areas as their below-median educated peers; urban movers are substantially better educated. Secondly, movers to urban areas are more likely to be part of formal labour markets at their prior rural locations, but appear to struggle to find adequate employment there and report unemployment more frequently. Thirdly, the out-migration of more educated individuals from agricultural households to urban areas is highly sensitive to households' economic conditions: being able to finance migration to urban areas appears to be an important pre-condition and bottleneck to the sorting of more educated individuals to urban areas.

The third chapter studies proxy means testing (PMT), which promises to identify poor households using information on household assets and demographic characteristics that can easily be collected and verified through surveys. Recently, however, Brown et al. (2016) have criticised PMT for commonly failing to identify poor households. I hence revisit the authors' results and find that poor calibration is a major driver of the poor performance of PMT they find. When I calibrate the poverty rate predicted by PMT to match the known actual poverty rate of the population, I find that PMT performs substantially better: across the 5 countries I examine, PMT correctly classifies 60-70% of poor households, while chance would only correctly classify 40%. The poorest households are the least likely to be missed by PMT, while the households wrongly predicted to be poor do not tend to be among the richest households. While far from perfect, these results suggest that PMT might still provide useful information to users aware of its limitations.

Contents

List of Figures

List of Tables

Chapter 1

INFORMAL MEDICAL PROVIDERS IN DEVELOPING COUNTRIES: WHAT DO THEY KNOW, DO, AND PROFIT FROM?

1.1. Introduction

Every year 7.6 million children die before reaching their fifth birthday, most of them in sub-Saharan Africa (3.6 million) and South and Southeast Asia (2.1 million). Infectious diseases, most commonly pneumonia, diarrhoea, and malaria, account for 64% of these deaths (Liu et al., 2012). However, a large proportion of these deaths is preventable using treatments (antimalarials, antibiotics, and oral rehydration) widely available even in developing countries.

Across many developing countries, private informal medical providers, who are not formally trained or legally recognized and typically operate outside the realm of government regulation, are essential sources of medical advice and treatment and provide a large share of health care (Waters et al., 2003; Bloom et al., 2011; Sudhinaraset et al., 2013). Particularly the poor often seek health care from these providers, as they are more affordable, easier to reach, and positively regarded, even though the quality of medical care they provide is thought to be low (Peters and Bloom, 2012; Makinen et al., 2000). Informally trained village doctors, drug sellers, and midwives have hence become important parts of developing countries' health systems (Mackintosh et al., 2016; Morgan et al., 2016). In sub-Saharan Africa, especially drug sellers are widespread, accounting for 77% of providers in Uganda, 36%-49% of treatments provided in Nigeria, and 33% of treatments provided in Kenya (Sudhinaraset et al., 2013).[1] In Ghana, the focus of this study, informal drug sellers account for 47% of health care providers (Makinen et al., 2011).

Paediatric health care is no exception to this pattern: "private sector providers are [also] the most commonly consulted source of care for child illness in many countries" (Waters et al., 2003). In a comprehensive literature review on care-seeking behaviour for childhood illness in developing countries, Geldsetzer et al. (2014) find that at the median pharmacies and drug vendors account for 32% of care sought for diarrhoea, 32% of care sought for malaria, and 17% of care sought for pneumonia.

Yet, despite the importance of these providers, little is known about the quality of medical care that drug sellers provide for childhood illness and about the determinants of this quality (Smith, 2009; Wafula and Goodman, 2010; Wafula et al., 2012): evidence on the quality of diagnosis and treatment provided by drug sellers for the most common childhood illness conditions is scarce and far between, and the determinants of treatment quality, provider knowledge, effort, and financial incentives have been neglected entirely. However, these are important to understand, as low quality of treatment need not mean that providers lack skill or knowledge: studying health care provision in urban India, for example, Das et al. (2012) find that low effort (rather than low knowledge) is a major bottleneck to better medical care. Economic incentives may similarly affect the provision of treatment: analysing the entry of a high-quality competitor in Uganda, Björkman-Nyqvist et al. (2016) find that incumbent drug sellers strategically increase the quality of treatment provided. When seen as a principal-agent problem where clients can (at best) imperfectly observe the quality of providers' treatment, it be-

[1] Drug sellers are a global phenomenon not restricted to sub-Saharan Africa: In Bangladesh, for example, informal providers are estimated to provide 60%-77% of health care (Sudhinaraset et al., 2013), while a survey of medical providers in 19 Indian states found that informal providers outnumbered trained providers four to one (Das and Hammer, 2014).

comes clear that drug sellers' knowledge might not be the only determinant of treatment quality, but that provider effort and financial incentives may also influence the quality of treatment provided. Thus, while evidence on the quality of health care that commonly utilised drug sellers provide for childhood illness is scarce in itself, research on the determinants of this quality, provider knowledge, effort, and financial incentives is almost completely absent.

This paper aims to provide evidence from Northern Ghana on the quality of diagnosis and treatment that drug sellers provide for the most common childhood illnesses and to study its determinants - drug sellers' knowledge and ability to provide adequate diagnosis and treatment, the effort that drug sellers exert in providing treatment (as measured by the gap between ability and actual practice), and drug sellers' financial incentives (as measured by profits from the sale of drugs). To do so, an original dataset on drug sellers' treatment practices, knowledge, and financial incentives was collected in Northern Ghana. Data on the quality of diagnosis and treatment provided by drug sellers for common childhood illness conditions (malaria, diarrhoea, respiratory infection, and anaemia) was collected covertly through "mystery clients" (surveyors pretending to be clients seeking medical help) in order to minimize observation bias; data on drug sellers' knowledge was collected through structured interviews with drug sellers (employing "vignettes", in which surveyors presented hypothetical illness scenarios to drug sellers in role plays and asked what they would do); and data on wholesale and retail prices of drugs was collected to study drug sellers' profits and financial incentives.

Mystery client visits revealed a dire picture of the quality of treatment provided for childhood illness: on average, drug sellers were almost twice as likely to sell inadequate or harmful drugs (which they did in 64% of interactions) as to provide the treatment recommended by medical guidelines or to refer the client to a skilled provider (36% of interactions). Differences in illness conditions existed: malaria was treated best (where drug sellers sold adequate drugs or referred in 47% of interactions), while respiratory infections were treated particularly poorly (where drug sellers sold adequate drugs or referred in 20% of interactions). A lack of knowledge (rather than low effort) appeared to be the key barrier to better treatment for three of the studied illness conditions: when drug sellers were formally interviewed about their knowledge and asked which treatment they would provide for hypothetical children with the same illnesses, treatment quality was similarly poor for malaria, respiratory infections and anaemia (although drug sellers were more likely to state referring a child). Large differences between drug sellers' knowledge and actual behaviour existed for diarrhoea however, where 85% of drug sellers knew the correct treatment when asked about, but only 34% of drug sellers provided it when secretly observed. Illness-specific low provider effort is a possible but unlikely explanation, as there is no obvious reason why provider effort should be low for one illness condition only. Profit incentives, on the other hand, are more plausible to discourage the provision of appropriate treatment (oral rehydration solution (ORS) and zinc), as they are relatively cheap and bear low profits to drug sellers. Yet, profit incentives need not prevent adequate treatment for the illnesses: a simulation exercise suggests that for all studied illness conditions drug sellers could sell bundles of drugs to clients, which would provide adequate treatment to the child, while still yielding similar profits to drug sellers and not costing more to clients as the largely inadequate treatments currently sold.

This paper mainly speaks to three literatures: research predominantly in public health has investigated the quality of medical care in the informal sector, while a literature in economics has drawn attention to the role of provider effort in the provision of medical treatment (and repeatedly found large "know-do" gaps). Lastly, a related strand of economic literature has highlighted the role of economic incentives in the provision of medical care.

Employing mystery clients, Tawfik et al. (2006) measure the quality of informal providers' diagnosis and treatment in Tanzania[2] and find poor case management for childhood malaria, diarrhoea, and respiratory infections; practitioners provided adequate treatment or referral in less than 16% of mystery client interactions for five of the six studied illness conditions.[3] A lack of knowledge appears to be one major bottleneck among providers, as following a one-day training intervention, the proportion of correctly managed cases increased by an impressive 34-50 percentage-points in four of the six scenarios studied. Nsimba (2007) similarly studies drug sellers' case management for childhood illness in Tanzania,

[2]Traditional birth attendants, drug shops, traditional healers, and private clinics comprise the majority of the authors' sample of informal providers.

[3]Treatment for severe diarrhoea and dehydration seemed to be the exception, where 62% of providers gave ORS and 21% referred the child urgently to another provider.

finding poor dispensing practices and a lack of knowledge among drug sellers. Yet, neither study allows to compare drug sellers' actual behaviour to their knowledge, thereby casting light on the role of effort, or explores the role of profit incentives in dispensing decisions.

Studying the role of effort and knowledge in determining the quality of medical care in developing countries, a second related strand of literature has repeatedly found that poor treatment practices need not imply low provider knowledge: Das and Hammer (2007), Das et al. (2012), and Mohanan et al. (2015) find the quality of medical treatment provided by both qualified and unqualified doctors in urban and rural India to be low, and document substantial differences between the quality of diagnosis and treatment that providers are capable of providing and that providers actually provide: actual practice (as measured through direct clinical observation or mystery patients) is generally substantially worse than the best practice providers are capable of providing (as measured through vignettes), a finding referred to as the "know-do gap". The size of the gap is so large, that using mystery clients in Madhya Pradesh, Das et al. (2012) find no differences between university-trained doctors and untrained informal practitioners in the likelihood of providing a correct diagnosis or correct treatment. In Tanzania, Leonard and Masatu (2005) similarly find that the correlation between physicians' actual behaviour (when observed directly in clinical practice) and physicians' competence (as measured by vignettes) is close to zero: "If the clinician does the right thing on the vignette, there is a 53% chance that he will do the right thing on the DCO [direct clinical observation]." Following the findings of these papers, it is obvious that poor observed treatment practices need not imply a lack of provider knowledge and that "provider effort may be a key determinant of quality in health service provision" (Das et al., 2012). One contribution of this paper is therefore to go beyond only observing the quality of treatment provided by drug sellers and to consider the importance of effort and provider knowledge. Furthermore, while existing research on the know-do gap has focussed on doctors, this paper examines its importance for another class of commonly used medical providers.

A third related strand of literature has investigated the economic incentives of medical providers and drug sellers and has similarly found that factors beyond knowledge - in this case economic incentives - are important determinants of treatment provision. Fitzpatrick (2016) finds that drug sellers in Uganda strategically respond to more informed customers by reducing both the price and the quality of sold antimalarials, as they (correctly) perceive that more informed customers are more likely to shop around and less likely to be repeat customers. Björkman-Nyqvist et al. (2016) find that drug sellers also respond strategically to changes in local market structure in Uganda: upon the entry of a new high-quality competitor, drug sellers become less likely to sell drugs of substandard quality while decreasing pries for antimalarials by approximately 17% on average. Evaluating the entry of a high-quality, low-cost drug retailer chain in Hyderabad, Bennett and Yin (2014) obtain qualitatively similar results: economic forces shape the quality of treatment provided by drug sellers. While this research has advanced our understanding of the impact of market structure and customer information on the price and quality of specific drugs, the wider question of how economic forces, such as financial incentives stemming from differences in profit margins between drugs, impact the quality of treatment provided remains unanswered. By considering profit margins (alongside knowledge and effort) explicitly, and asking whether adequate treatment might simply not be sufficiently profitable to drug sellers, this research aims to remedy this.

This paper is organised as follows: Section 2 describes the local setting, the empirical methods, and the collected data. Section 3 describes and analyses drug sellers' actual behaviour in the treatment and diagnosis of childhood illness, as observed through mystery client visits. Section 4 then focusses on drug sellers' self-reported behaviour in identical situations, thereby exploring drug sellers' knowledge of appropriate practices. Section 5 discusses drug sellers' financial incentives and section 6 concludes.

1.2. Local Setting and Empirical Methods

Drug sellers have a large presence across many developing countries, making for a high number of suitable locations for research. In Ghana, the setting of this study, drug sellers are "the greatest and most accessible source of [health] service[s] in rural and urban-poor districts" according to the World Bank (Makinen et al., 2011, p. 40); at the same time, child health outcomes are relatively poor and under-five mortality is comparatively high. Together, these two facts made Ghana a particularly well-suited study location.

1.2.1. Local Context: Child Health in Ghana

Ghana ranks 140th (out of 188 countries) in the UNDP's 2015 Human Development Index; under-five mortality is comparatively high at 60 deaths per 1,000 live births according to the 2014 Ghana Demographic and Health Survey (DHS). Malaria, diarrhoea, and pneumonia are common in children. In a representative sample of 7550 under-five children, the 2011 MICS reports that in the two weeks prior to the survey 12.7% of children had diarrhoea, 2.9% had suspected pneumonia, and 18.9% had fever (an indication of likely malaria). Microscopic analysis confirmed malaria parasitaemia in 27.5% of children. Yet, treatment for these diseases is highly deficient: 41.5% of children with diarrhoea did not receive oral rehydration therapy (i.e. oral rehydration solution (ORS), recommended home-made fluids, or increased fluids), 44.3% of children with suspected pneumonia did not receive antibiotics, and 47.3% of children with fever did not receive recommended anti-malarials. This lack of treatment is potentially fatal, making pneumonia, malaria, and diarrhoea the three leading causes of under-five death in Africa, accounting for 1.56 million (or 44% of) deaths annually (Liu et al., 2012).

1.2.2. Local Context: Treatment Provision in Ghana

Informally trained drug sellers are widely used as first-line providers for the treatment of illness in Ghana: 11,000-13,000 chemical sellers operate across the country (Dalberg Global Development Advisors and the MIT-Zaragoza International Logistics Programme, 2008, p. 12); the World Bank calls them "the greatest and most accessible source of [health] service[s] in rural and urban-poor districts" (Makinen et al., 2011, p. 40). The 2008-10 Accra Time Use and Health Study found that around 47% of illnesses were treated at a drugstore or pharmacy (Fink et al., 2012), while data from the 2005/06 Ghana Living Standards Survey showed that chemical sellers provided 35% of treatment sought for illnesses episodes

(GSS 2008).[4]

For Ghana, some evidence on drug sellers' practices exists: Buabeng et al. (2010) examine the availability of antimalarials and dispensing practices at drug sellers and find that only 4% of their sample of 58 chemical sellers treated malaria as recommended. Friedman et al. (2015) test whether SMS reminders to drug sellers can improve the treatment of childhood diarrhoea and find that reminders had little effect but that 80% of their 698 surveyed drug sellers adequately treated diarrhoea using ORS in the 3-8 months after a training intervention in any case. Considering sexually transmitted infections (STIs), two earlier studies (Adu-Sarkodie and Steiner, 2000; Mayhew and Nzambi, 2001) find poor case management at drug seller stores. Overall, however, comprehensive research documenting the quality of diagnosis and treatment for the four most relevant childhood illness conditions and investigating drug sellers' knowledge and financial incentives, two fundamental determinants of behaviour, has been missing.

1.2.3. Medical Benchmark for Treatment Practices

In principle, even informal providers, such as drug sellers, could diagnose and treat malaria, diarrhoea, acute respiratory infections, and anaemia in children correctly, as clear and simple medical guidelines for the diagnosis and treatment of common childhood illnesses exist. Using a series of algorithms and flow charts, the "Integrated Management of Childhood Illness" (IMCI) approach developed by the WHO and UNICEF provides minimally trained health providers with a systematic way to diagnose the most common illnesses and their severity (pneumonia, diarrhoea, measles, fever, otitis media, and malnutrition), to ascertain which children require antimalarial or antibiotic treatment, and which are in need of immediate referral or hospitalisation (Black et al., 2016; Tulloch, 1999; Gove, 1997). Upon diagnosis, medical guidelines recommend the treatment of (suspected) malaria with artemisinin combination therapy (ACT), diarrhoea and related dehydration with oral rehydration solution (ORS) and zinc, respiratory infections with amoxicillin, and anaemia with iron and a dewormer (if not already

[4] Buabeng et al. (2007) similarly find that for a sample of 500 patients seeking care for malaria from two hospitals in the Ashanti region, licensed chemical sellers were the most common source of treatment (50%) among those that had sought treatment beforehand (43%).

taken in the prior six months).

To illustrate the approach, figure 1.1 depicts the fever module of the IMCI algorithm. After the initial assessment of the child for danger signs requiring immediate referral (module not depicted), the algorithm requires the health care provider who diagnoses a child with fever to inquire about the illness history, and to look for a stiff neck, runny nose, signs of a bacterial cause of fever (among them swelling, local tenderness, or oral sores), and signs of measles. Based on the results of the assessment, the fever is then classified as a severe febrile disease, malaria, or a fever without malaria, and the appropriate action and medication is indicated. Similar modules exist for diarrhoea, cough, and anaemia, as well as for the initial assessment of the child for general danger signs, and for ear problems, acute malnutrition, and HIV status.

The efficacy of this approach has been established (Bryce et al., 2005; Black et al., 2016), and many low- and middle-income countries, among them Ghana, have adopted the IMCI as their approach for the first-level diagnosis and treatment of childhood illness. The algorithm hence provides a natural (and reasonable) benchmark for evaluating the quality of diagnosis and treatment provided by the drug sellers studied. Furthermore, its existence lends a clear policy implication to this research, as it is suitable for application by minimally skilled providers and drug sellers could hence be trained in its use.

1.2.4. Study Location, Sample, and Surveys

Substantial variation in child health outcomes exists within the ten regions of Ghana; the Northern Region fares worst among several indicators (2014 DHS) and was hence an obvious choice of location; the study took place in Tamale, the regional capital. Government registration records list 228 drug sellers operating in Tamale, accounting for 41% of the 558 of drug sellers registered in the Northern Region. A census of drug sellers in Tamale undertaken for a recent study (Raifman et al., 2014) served as the sampling frame for this study. 80 drug sellers were randomly drawn from the 153 drug sellers found during the census survey and formed the sample of drug sellers used in this study. Drug sellers that could not be found during surveys were replaced with drug sellers randomly drawn from the

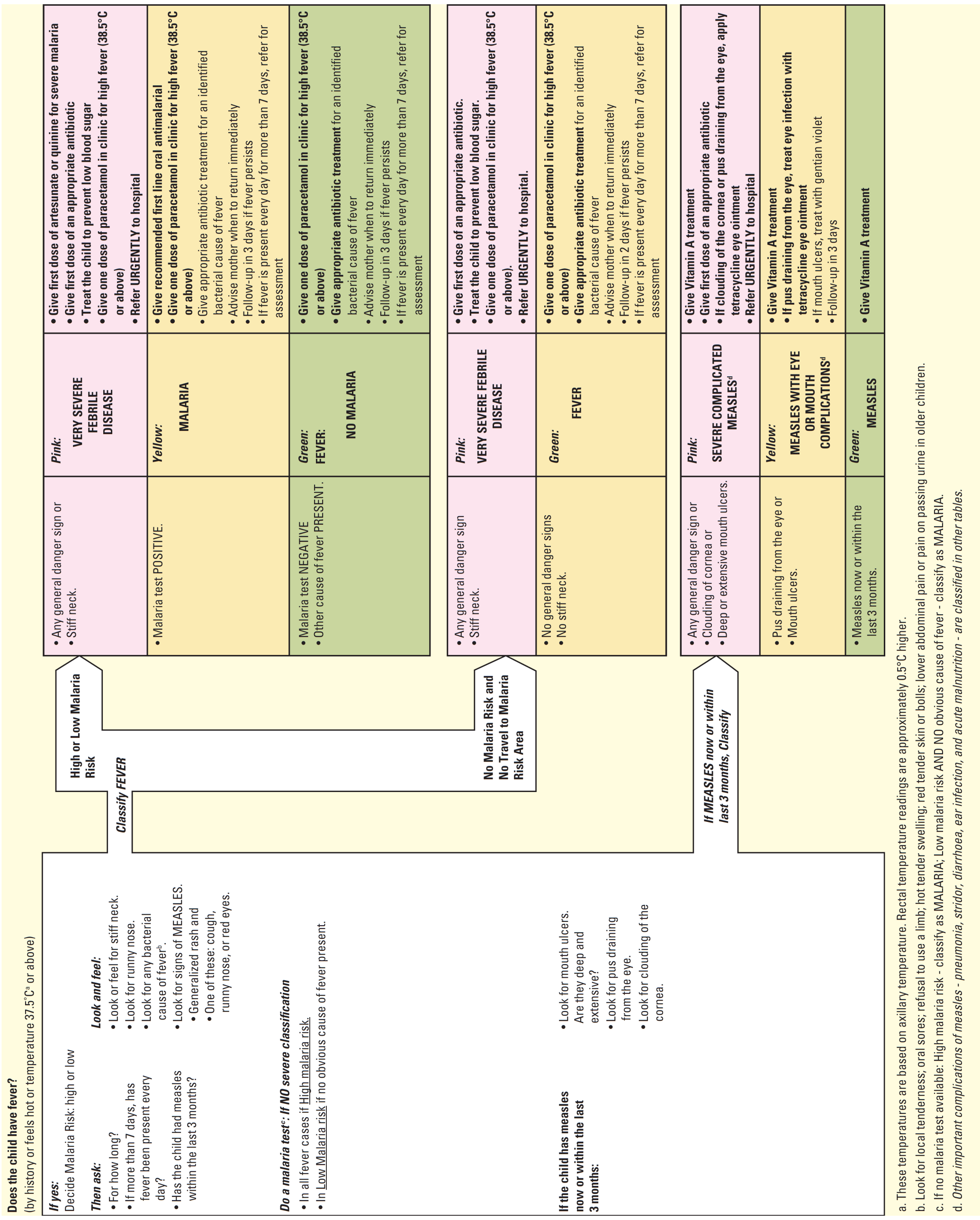


Figure 1.1: Fever Algorithm from 2014 IMCI (source: Black et al. (2016))

sampling frame located in the same neighbourhood.

"Mystery client" visits were employed to covertly collect detailed information on the diagnosis and treatment that drug sellers actually provided; facility surveys, which included "vignettes" for the four illnesses studied were employed to collect data on drug sellers' knowledge and competence in diagnosing and treating the same four illnesses, as well as to obtain data on store operations. Furthermore, in order to estimate drug sellers' profit margins on drugs, data on wholesale prices for drugs was obtained from local wholesalers, which - combined with the price data from the drugs actually purchased during mystery client visits - permitted to estimate drug sellers' profit margins and financial incentives. The three surveys are discussed in turn below.

Mystery Client Visits

Mystery clients are an established method for studying the behaviour of health care providers while minimizing observation bias (Madden et al., 1997; Watson et al., 2006); economists have similarly used them in various contexts (Das and Hammer, 2014; Banerjee et al., 2014). Unlike conventional surveys, mystery client surveys employ "undercover" surveyors, who engage with the survey respondent pretending to be an ordinary customer or patient. During the interaction, mystery clients follow an exact script and have been trained to reply to questions; upon completing the interaction, mystery clients leave and subsequently record details of the interaction in a structured questionnaire.

Data on drug sellers' actual practices were collected by mystery clients who posed as mothers and sought treatment for their supposedly sick child exhibiting symptoms of malaria, diarrhoea, acute respiratory infection, or anaemia. Each drug seller was visited once for each condition, by a different mystery client each time, for a total of four mystery client visits. Scripts for mystery clients were elaborated using the WHO's IMCI manual (WHO, 2014), and subsequently reviewed by Dr. Dorothy Addo-Yofo (paediatrician at 37 Military Hospital, Accra). Each script required the mystery client to describe their child's condition (such as "my child is ill and has a fever" in the case of malaria) and to ask for diagnosis and treatment. Mystery clients were prepared to answer questions on the child and

Table 1.1: Mystery client visits: survey outcomes

	all	Malaria	Diarr.	Resp. Inf.	Anaem.
interaction: sold any drugs	292	76	77	70	69
interaction: referred w/o drugs	7	0	0	1	6
unavailable: no drugs or staff	16	3	3	7	3
refused: need to see child	5	1	0	2	2
Observations	320	80	80	80	80

his/her illness, and provided further details when asked by the drug seller. Mystery clients purchased the drugs advised by the drug seller and upon leaving the store completed a survey on the outcomes and details of the interaction. Apart from the drugs sold, their cost, and the diagnosis given, the surveyors recorded which questions the drug seller had asked diagnosing the illness of the child, whether the drug seller asked or insisted to see the child, and whether the drug seller referred the mother to a doctor or hospital. In line with other studies employing mystery client visits, mystery clients did not bring their supposedly ill 2-year old child with them to the drug seller, but (if asked) explained to the drug seller they had left their child at home.

Table 1.1 summarises outcomes of the mystery client survey: 320 mystery client visits were conducted in total (four visits, one per illness condition, for a total of 80 drug sellers). Drug sellers sold drugs without referral or asking to see the child in 292 interactions. In 16 interactions, the drug seller was either not encountered or stated not to have appropriate drugs; in 7 interactions, drug sellers referred the client without making a sale. Lastly, in 5 interactions, drug sellers refused to engage with the mother without the sick child present.

Facility Surveys

Information on drug sellers' reported treatment practices, knowledge, and store operations was collected through structured interviews with drug sellers at their stores. Vignettes were used to elicit drug sellers' knowledge of diagnosis

Table 1.2: Facility survey: survey outcomes

	all	Malaria	Diarr.	Resp. Inf.	Anaem.
interaction: sold any drugs	257	72	74	60	51
interaction: referred w/o drugs	31	0	0	13	18
unavailable: no drugs or staff	0	0	0	0	0
refused: need to see child	8	2	0	1	5
attrited: no informed consent	24	6	6	6	6
Observations	320	80	80	80	80

and treatment for childhood illness: surveyors asked sellers to engage in a role-play and to imagine that the surveyor was a mother coming into their store, describing her child's principal symptom, and asking what to do and which drug(s) to buy. Surveyors explained that they would like to know what questions the drug seller would ask the mother (to which they would provide answers in turn), which medication the drug seller would recommend, and what advice the drug seller would give to the mother in this situation. Reported treatment practices for the four illness conditions (malaria, diarrhoea, acute respiratory infection, and anaemia) were thus elicited. In line with previous studies, elicited data on drug sellers' self-reported treatment practices are taken to reflect the knowledge of drug sellers on adequate treatment practices. Facility surveys also collected information on the availability and prices of drugs for the treatment of childhood illness (antimalarials, antibiotics, ORS, zinc, and dewormers), and on details of the store operation and the individuals working there. To ensure that the visit of a surveyor did not affect drug sellers' actual treatment practices as measured through mystery client visits (i.e. if drug sellers suspected they were under supervision following the facility survey), facility surveys took place only after mystery client visits had been completed.

Table 1.2 summarises the results of data collection during facility surveys: of the total sample of 80 drug sellers (accounting for 320 vignettes), six refused to participate in the survey, leading to 24 missing vignettes. Table 1.3 presents summary statistics on drug sellers and their operations. Importantly, drug sellers

Table 1.3: Facility survey: summary statistics

Summary statistics	
mean number of clients (per day)	45.47
mean client expenditure (GhC)	8.27
mean duration of client interaction (minutes)	6.07
proportion of clients who the drug seller knows	0.32
Advice on appropriate treatments	
proportion of clients who ask for advice	0.49
proportion of clients who ask for drug directly	0.51
Store operations	
mean years of operation	13.66
proportion where owner lives within 500m	0.39
Relations with wholesalers	
proportion procuring drugs from local wholesalers	0.89
Observations	74

overwhelmingly purchased drugs from local wholesalers (suggesting that using the wholesale prices of local wholesalers to estimate unit profits is a valid approach), commonly provided advice to customers (rather than just selling a specific drug they were asked for), and did not know a substantial share of their customers (making mystery clients a feasible strategy).

Wholesaler Drug Price Survey

Local drug wholesalers in Tamale (from whom drug sellers commonly purchase their drugs) were surveyed to obtain the wholesale prices of all drugs encountered during mystery client visits and facility survey interviews.[5] Using these

[5]Facility surveys asked drug sellers about the names of the wholesalers from which they purchased their drugs. In total, drug sellers mentioned 19 wholesalers. The four most commonly

wholesale prices, unit profits from the sale of each drug were estimated for each interaction by calculating the difference between the price at which a drug seller sold a drug and the average price at which wholesalers sold the drug to drug sellers. Mystery clients purchased a total of 505 drugs during their visits; wholesale prices could be obtained for 419 of these drugs (83%). In cases where the wholesale price of a drug could not be obtained, the average wholesale price of similar drugs (containing the same active ingredient) was used as a proxy.

1.3. Treatment and Knowledge

1.3.1. What do drug sellers do?

Mystery client visits resulted in a gloomy picture of the treatment provided by drug sellers: overwhelmingly, the drugs purchased by mystery clients were inadequate for the treatment of the presented illness condition and violated guidelines, while referrals were rare. When selling drugs, drug sellers were twice as likely to sell inadequate or harmful drugs as to sell the drugs recommended by treatment guidelines: averaged across all illness conditions, drug sellers provided adequate treatment[6] in 31% of interactions (and referred the mystery client to another provider in a further 5% of interactions). Between illness conditions, the quality of treatment varied substantially: while 47% of mystery clients for malaria and 43% of mystery clients for anaemia were sold an adequate drug or referred, this was the case for only 34% of mystery clients presenting cases of diarrhoea and 20% of mystery clients presenting symptoms of acute respiratory infection.

Table 1.4 summarises these findings on the quality of treatment that drug sellers provide; table 1.A3 in the appendix presents them in greater detail. Two findings stand out: antibiotics were omnipresent at drug sellers' stores and were commonly sold for the treatment of diarrhoea: 86% of mystery client visits pre-

mentioned wholesalers (accounting for 44 of 81 mentions) and two smaller outlets were surveyed. Furthermore, these wholesalers also commonly sold the drugs of four drugmakers that had been separately mentioned as wholesalers by drug sellers. Including these, the surveyed wholesalers hence accounted for 69 of 81 mentions.

[6] I define "adequate" treatment generously: any interaction in which a drug with the appropriate active ingredient is sold is counted as adequate, even when other potentially harmful drugs were also sold in the same interaction.

Table 1.4: Treatments provided (mystery client visits)

	all	Malaria	Diarr.	Resp. Inf.	Anaem.
adequate treatment or referral	0.36	0.47	0.34	0.20	0.43
adequate treatment	0.31	0.46	0.34	0.17	0.25
- but also sold harmful drugs	0.07	0.01	0.22	0.03	0.00
referred	0.05	0.01	0.00	0.03	0.17
- but also sold harmful drugs	0.00	0.00	0.00	0.00	0.01
inadeq. treatm. & no referral	0.64	0.53	0.66	0.80	0.57
- and also sold harmful drugs	0.27	0.03	0.66	0.27	0.13
illness-specific					
- sold monotherapies (malaria)	.	0.39	.	.	.
- sold antibiotics (diarrhoea)	.	.	0.86	.	.
Observations	299	76	77	71	75

Proportion of interactions. Sample: interactions in which drugs were sold (with or without referral) or the child was referred (without drugs sold). This covers 93% of actual interactions.

senting cases of diarrhoea were told to purchase an antibiotic (and did so). Secondly, monotherapies were commonly sold to mystery clients presenting symptoms of malaria instead of recommended combination therapies (monotherapies were sold in 39% of interactions, while combination therapies were sold in 46%), although monotherapies are no longer considered fully effective due to growing parasite resistance.

Generally, drug sellers spent little time and effort to ask about (and understand) the child's symptoms and quickly guessed a diagnosis based on the principal symptom the mystery client mother mentioned. Across illness conditions drug sellers asked relatively few questions (on average 2.2, most commonly about the age of the child and the duration of symptoms) that could help to understand the probable cause of the child's symptoms, and then recommended a drug (or several, 1.7 on average) to the client.[7] To select the drug(s), drug sellers appeared to use a simple heuristic and seemed to associate each principal symptom presented by mothers with an illness, for which then they sold the drugs they believed to be appropriate. Drug sellers commonly associated fever with malaria, for example (pronouncing malaria as the diagnosis in 68% of interactions when mothers presented fever as the child's principal symptom), and paleness with anaemia, a "lack of blood," or a "blood problem" (in 65% of interactions). This worked well for three of the four illness conditions (malaria, diarrhoea, and anaemia) where drug sellers generally happened to guess correctly, but failed when a different illness than the one that drug sellers generally associated with the symptom was its true cause, as was the case for respiratory infections.[8]

The consequences of drug sellers' failure to sufficiently inquire about the child's symptoms and illness history extends beyond respiratory infections, however: fever in children might be caused by illnesses other than malaria (such as measles or other infections) and these illnesses hence need to be ruled out (the IMCI medical guidelines therefore only classify fever as malaria if the child does

[7] Table 1.A1 and figure 1.A1 in the appendix present information on the diagnostic questions drug sellers asked.

[8] In the case of respiratory infections, drug sellers overwhelmingly failed to understand that the child presented to them as having a cough and difficulty breathing was also breathing fast and was hence, following the IMCI algorithm, to be diagnosed as having a respiratory infection and to be treated with amoxicillin (after excluding severe pneumonia as a diagnosis by verifying that the child did not also exhibit a stridor).

not have a stiff neck, runny nose, rush, cough, or red eyes).[9] Yet, drug sellers rarely asked the questions required to rule out such alternative causes for fever (measles, infections) or detect complications (such as severe dehydration for diarrhoea) and took a leap of faith when pronouncing their diagnosis, since based on the information drug sellers had obtained, other illnesses were equally plausible (even if potentially less likely) causes. By not taking steps to confirm or rule out these illness as potential causes of the symptoms, drug sellers hence revealed that these illnesses would be, in fact, misdiagnosed and subsequently inadequately treated. The quality of treatment and diagnosis by drug sellers would therefore have been even lower if the study had considered other illnesses with similar principal symptoms, such as measles instead of malaria.

Unsurprisingly, the common provision of inappropriate treatments, as well as sales of medically unnecessary products, such as cough syrups and vitamin drinks, also led to unnecessarily high expenditures on drugs by mystery clients: on average, the majority of money spent by mystery clients on the treatments that drug sellers recommended was wasted and only a small fraction of the total expenditure was spent on drugs appropriate for the treatment of the illnesses they presented. Across illness conditions, mystery clients spent on average 9.2 Ghana Cedis (approximately USD 2.3) on the drugs recommended by drug sellers; less than a quarter (24%, or GhC 2.2) of this expense fell on drugs deemed appropriate and beneficial by medical guidelines, while a similar share (23%) fell on drugs deemed harmful, and 46% (4.2 Ghana Cedis) fell on drugs deemed unnecessary or not strictly necessary. Tables 1.5 and figure 1.2 (in the following section) provide a detailed summary of the expenditures on different types of drugs across illness conditions.

[9] As a diagnostic algorithm, the IMCI requires 7-17 questions to be asked (depending on the principal symptom) in order to rule out alternative diagnoses, such measles and infections for fever, dysentery and cholera for diarrhoea, detect complications, such as severe dehydration for diarrhoea, and detect children with danger signs in need of immediate referral for treatment. 5 initial questions should always be asked, plus additional symptom-specific questions: an additional 12 questions should be asked for fever (to detect and rule out causes of fever other than malaria), 6 additional questions for diarrhoea (to detect dehydration, chronic diarrhoea, and cholera), 4 additional questions for cough and breathing problems, and 2 additional questions for paleness. Drug sellers fell markedly short of this standard, as also shown in table 1.A1

Table 1.5: Expenditure on drugs (mystery client visits)

	all	Malaria	Diarr.	Resp. Inf.	Anaem.
Total expenditure	9.21	12.34	6.88	9.01	8.63
- on appropriate drugs	2.19	3.89	0.86	0.86	3.09
- on unnecessary drugs	4.20	5.71	0.60	6.07	4.59
- on harmful drugs	2.18	0.21	5.41	2.08	0.95
- on outdated antimalarials	0.64	2.53	0.00	0.00	0.00
Observations	299	76	77	71	75

Average expenditure in Ghana Cedis. Sample: interactions in which drugs were sold (with or without referral) or the child was referred (without drugs sold). This covers 93% of actual interactions.

Among all illness conditions, the share of expenditure on adequate drugs was highest for malaria and anaemia (at 32% and 36% on average, respectively), since appropriate treatments for both conditions (artemisinin combination therapies for malaria, and haematinic tonics for anaemia) were relatively expensive; unnecessary drugs (painkillers or monotherapies for malaria and vitamin tonics without sufficient iron for anaemia) accounted for the remainder of the expenditure. For diarrhoea and respiratory infections, on the other hand, adequate treatments accounted for only a minimal share of expenditure (13% and 10%, respectively). Total treatment expenditure was lowest for diarrhoea, as both antibiotics and oral rehydration solution (ORS), the treatments most commonly sold, were both relatively cheap; unsurprisingly, given their common sale (in 86% of interactions), harmful antibiotics also accounted for the lion's share (79%) of expenditure for diarrhoea. For respiratory infections, on the other hand, unnecessary drugs (mostly cough syrups) accounted for the lion's share (at 67% of expenditure).

1.3.2. Do drug sellers know better?

The previous section established through mystery client visits that drug sellers overwhelmingly provided inadequate treatments for the four common childhood illnesses studied. Asking what causes this low quality of treatment is hence a natural question: a lack of knowledge among drug sellers, adverse financial incentives (in the form of lower profits from providing appropriate treatments), or simply a lack of effort (a low quality of treatment when drug sellers have sufficient knowledge and no adverse financial incentives) are all potentially plausible explanations that any investigation needs to consider. In principle, demand side effects (whereby clients demand inappropriate treatments) could also lead to the provision of inadequate drugs by drug sellers.[10] However, since mystery clients never demanded any specific kind of drug, but simply described their child's symptoms, answered potential questions, and then purchased any drug(s) that the drug seller suggested, (first-order) demand side effects are a less plausible explanation in this setting.

Among the potential explanations, examining drug sellers' knowledge is a natural starting point to understand the low quality of their treatment, since without knowledge of the appropriate treatment and diagnostic procedures (which unskilled providers could also plausibly know and adhere to, as discussed in section 1.2.3), it is hard to see how drug sellers could provide correct treatments. Data on drug sellers' knowledge and ability to diagnose and treat the common childhood illnesses that mystery clients presented was collected through vignettes, a methodology well established in public health research and discussed in section 1.2.4: enumerators visited drug sellers to survey them and engaged them in a role play, asking what drug sellers would do if clients presenting a given symptom sought help. In correspondence with the mystery client visits, four different vignettes (one for each condition) were presented to drug sellers, and drug sellers were asked what they would do in such a situation (and asked specifically which questions, if any, they would ask and which drugs, if any, they would sell to such a client). As is common practice (Amin et al., 2008; Mohanan et al., 2015), the behaviour of drug sellers in response to vignettes is interpreted as a measure of drug sellers' knowledge.

Findings from vignettes suggest that a lack of knowledge is the principal bottleneck to better treatment for three of the four illnesses: figure 1.2 and table 1.A3

[10]Previous research has studied such effects in other settings: Bennett et al. (2014) for example find that competition among health care providers leads increases the prescription of antibiotics demanded by patients, even when providers know that they are a not an effective treatment.

describe the knowledge of drug sellers as measured by vignettes and compare the treatments that drug sellers say they would provide to the treatments that drug sellers actually provide as observed by mystery clients. Responses in vignettes and in actual interactions were very similar for cases of malaria, respiratory infections, and anaemia: in 46% of vignettes for malaria, drug sellers said they would sell an appropriate antimalarial (and actually did so in 46% of mystery client visits), while in 18% of vignettes for respiratory infections drug sellers said they would sell an antibiotic (and did so in 17% of mystery client visits). When presenting symptoms of anaemia, drug sellers stated they would sell a haematinic liquid in 10% of interactions, while they actually did so in 25% of mystery client interactions. However, this seemingly worse treatment in vignettes is explained by a larger proportion of referrals: drug sellers stated they would refer the client in 49% of vignettes, while they only did so in 17% of mystery client interactions. The large correspondence between actual behaviour and knowledge as displayed in vignettes hence suggests that for three of the four examined illnesses - malaria, respiratory infections, and anaemia - a lack of knowledge is the principal impediment to better treatment by drug sellers. The fact that drug sellers asked as few questions that would help to understand their clients' symptoms (and investigate potential medical causes) when put on the spot in vignettes as they did in mystery client interactions[11] supports the argument that low knowledge rather than a lack of effort in actual interactions is a principal bottleneck.

However, beyond the correspondence between drug sellers' knowledge and practices for the three illnesses, two significant differences between actual and self-reported behaviours exist: firstly, drug sellers are significantly more likely to *state* that they would refer a child than to actually refer it (23% vs. 5%); this is the case for all illness conditions, although differences are particularly large for anaemia, as discussed above. Secondly however, drug sellers appeared to know well

[11] Figure 1.A1 and the corresponding table 1.A2 (both in the appendix) provide a detailed overview of the questions that drug sellers asked for the diagnosis of each condition. Across questions, there are no significant differences in the proportion of drug sellers that actually ask a given question when surveyed by mystery clients and that ask a given question in vignettes. Drug sellers' history-taking and diagnosis were hence similarly poor in actual and hypothetical interactions; and drug sellers appear to do exactly what they say they would do in this regard, suggesting that they do not deem it useful to ask asking further questions in order to arrive at a diagnosis valuable, or that they do not know which further questions to ask.

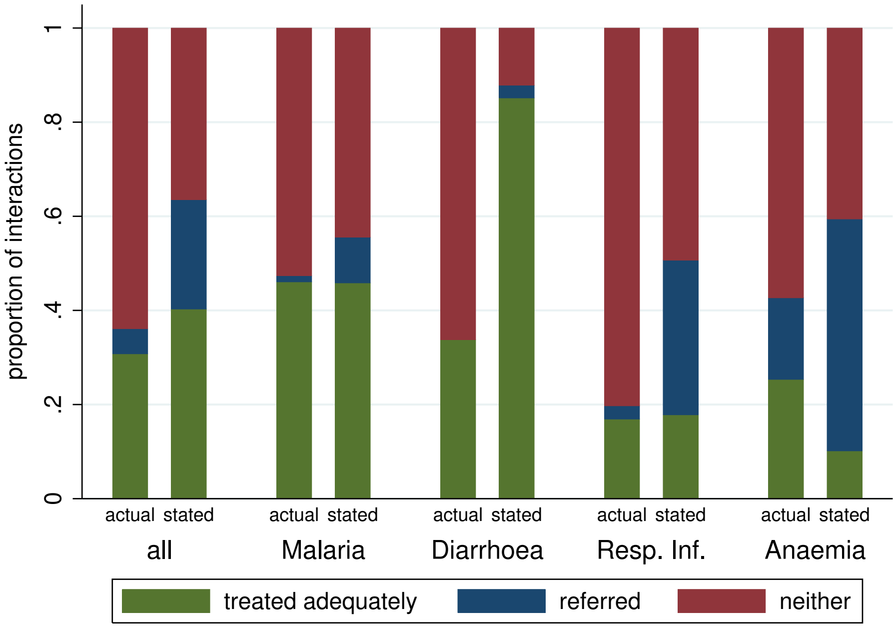


Figure 1.2: Treatments provided (actual vs. stated). Figure 1.A2 includes numerical labels.

what the appropriate treatment for diarrhoea was and overwhelmingly provided such correct treatments in hypothetical vignettes: 85% of all vignette interactions concluded with the sale of oral rehydration salts (ORS), while drug sellers only sold these correct treatments to 34% of mystery clients. Drug sellers were also significantly less likely to sell inappropriate antibiotics in vignettes, but commonly did so in mystery client interactions (30% vs. 86% of interactions). Given their knowledge of the adequate treatment for diarrhoea, a lack of knowledge can hardly explain the poor treatment practices observed by mystery clients.

An examination of clients' expenditures on the drugs suggested by drug sellers (actually incurred by mystery clients and hypothetical in the case of vignettes) provides a first indication that financial incentives could explain the difference between practice and knowledge in the treatment of diarrhoea. Figure 1.3 and table 1.A4 in the appendix compare the cost of treatment[12] for each disease in mystery

[12] expressed in Ghana Cedis (GhC); at the time of field work, GhC 4 approximately equalled USD 1

client interactions and vignettes.[13] Across all illness conditions, treatments in vignettes cost approximately 20% less than the treatments sold to mystery clients, as drug sellers were less likely to sell unnecessary and harmful drugs in vignettes, resulting in reduced expenditures on these items. For malaria, respiratory infections, and anaemia, differences in the actual and stated cost of treatment were small (at 5%, 8%, and a more notable 19%, respectively). For diarrhoea, however, expenditures differed substantially between (covertly observed) actual interactions and vignettes: while mystery clients spent on average GhC 6.88, expenditure in vignettes was only GhC 3.34. The lower use of (and subsequently lower expenditure on) antibiotics in vignettes was the main driver of this reduction: mystery clients spent on average GhC 5.41 on antibiotics when seeking treatment for diarrhoea, while hypothetical clients in vignettes spent 73% less on antibiotics (GhC 1.47), as drug sellers were significantly less likely to state selling them in vignettes. Conversely, expenditure on ORS and zinc were significantly higher in vignettes (increasing by GhC 0.92 from GhC 0.86 to GhC 1.79). However, as ORS and zinc are relatively cheap, this increase was too small to compensate for the substantial loss of revenue from selling fewer antibiotics in vignettes. Providing the correct diarrhoea treatment in vignettes (and not selling antibiotics) thereby resulted in a significant loss of revenue to drug sellers.

Obviously, this difference between actual and stated behaviours leaves room for several (non-rival) interpretations: firstly, although stating that ORS and zinc are the adequate treatment for diarrhoea, drug sellers might believe antibiotics to be the more effective treatment, but not admit this as they are legally not allowed to dispense antibiotics. Secondly, clients might so commonly demand antibiotics for the treatment of diarrhoea, that drug sellers sell them in anticipation of their clients' demands, even to clients - such as the mystery clients - who do not demand any particular treatment and although drug sellers know that antibiotics are not the adequate treatment. Thirdly, ORS and zinc simply might not be sufficiently lucrative for drug sellers (as the lower total cost of such treatment leaves less

[13] The cost of treatment is the expenditure on drugs that mystery clients incurred, and the price of the drugs that drug sellers say they would sell in vignettes. Since drug sellers were significantly more likely to refer children in hypothetical interactions, average expenditures were calculated only for interactions in which drugs were sold (this was the case for 90% of mystery client visits and 80% of vignettes).

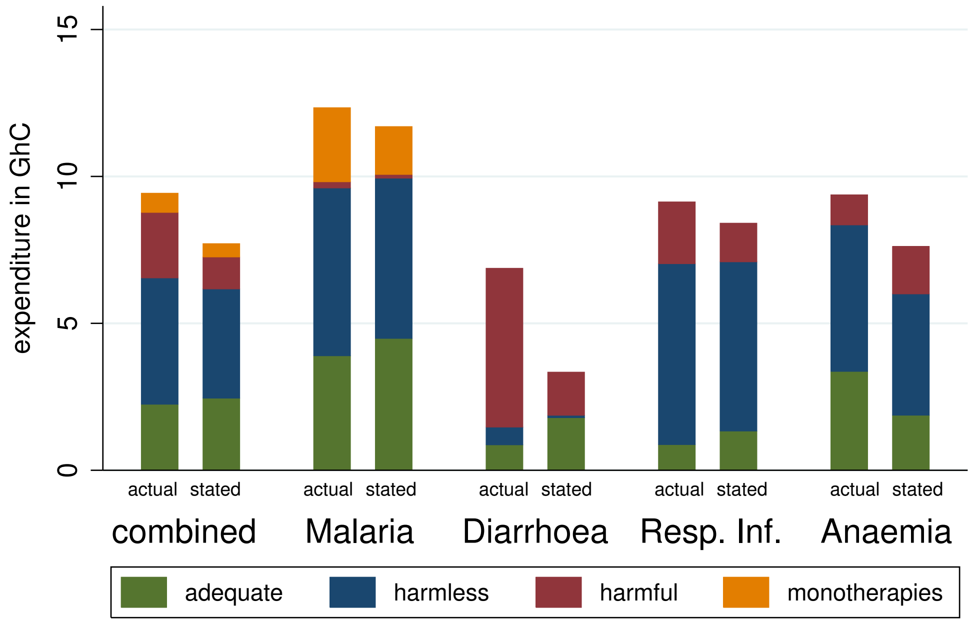


Figure 1.3: Expenditure on treatment by condition (actual vs. stated). Figure 1.A3 includes numerical labels.

room for profit margins), such that drug sellers resort to selling more expensive and profitable antibiotics instead, even though they know the optimal treatment to be ORS and zinc. Economic incentives are then the driving force of drug sellers' behaviour, and the provision of knowledge to drug sellers would not translate into a higher quality of treatment as drug sellers would find their profits reduced. The following section hence investigates further whether drug sellers' economic incentives might impede the provision of better treatments- do drug sellers have economic incentives to provide inferior treatments?

1.4. Results: Economic incentives

Drug sellers operate private, for-profit enterprises. Unlike other medical providers, they neither charge a consultation fee nor are paid by the government (or health insurance schemes); instead their income stems from the sale of drugs. Buying drugs from wholesalers and retailing them to customers at higher prices, the business model of drug sellers is hence relatively simple: drug sellers earn profits when selling drugs to clients.[14] Different drugs (among the wide sample of drugs purchased by mystery clients) have different wholesale (and retail) prices, and thus, profits vary across drugs. Profit-maximizing behaviour may hence plausibly influence what drugs (and thereby what quality of treatment) drug sellers provide, and induce them to sell more expensive, unnecessary, or even wrong treatments to earn higher profits. To understand whether economic incentives discourage drug sellers from providing medically appropriate treatments, I return to examine the data on drug sellers' treatment practices collected through mystery clients. The analysis in section 3 showed that inappropriate and harmful drugs account for a large share of (mystery) clients' expenditures, which suggests that they might also generate a substantial share of the profits that drug sellers earn. This in turn might lead one to worry that - even when known to drug sellers - adequate treatments could simply be too unprofitable for drug sellers to provide.

In this section, I hence examine these questions empirically, building on data from mystery client visits and a market survey of local drug wholesalers, from which drug sellers commonly purchase their drugs (as discussed in section 1.2.4). Combining these two sources of data, I estimate for each drug and drug seller the profit that a drug's sale yielded to the drug seller (the difference between the retail price paid by the mystery client and the average price at which wholesalers sell the drug to drug sellers). Analogously to the analysis of expenditure on drugs (in section 1.3.1), I first examine the size and sources of drug sellers' profits in mystery client visits. I then investigate through two simulation exercises how providing adequate treatments would affect the profits that drug sellers earn (and the expenditure that clients face). The motivation behind this exercise is simple: if increases in the quality of treatment that drug sellers provide lead to reductions in drug sellers' profits or increases in clients' expenditures, then providing better treatment is not necessarily in the interest of drug sellers (or affordable to clients). In this case, interventions aimed at improving the knowledge of drug sellers (such as training) might be able to increase *knowledge*, but would fail to improve the

[14]Throughout the following discussion, the term "profit" will refer to the difference between a drug's wholesale and retail prices, which drug sellers retain when selling a drug to customers. It thereby does not account for costs to the drug seller (i.e. capital and labour) which are fixed in the moment the client enters the drug store.

actual practices of drug sellers as they reduce profits. If, on the other hand, the provision of adequate treatment does not reduce drug sellers' profits, then financial incentives do not pose an obvious barrier to the provision of better treatment.

1.4.1. Profits from treatment provision

Mystery client visits collected detailed information on the treatment that each drug sellers provided and recorded the price of every drug that was sold to them. This allows me, combined with data on the (wholesale) prices at which drug sellers buy these drugs themselves, to estimate a drug sellers' profit margin on each drug sold to a mystery client. As drug sellers overwhelmingly purchased their drugs from local wholesalers (Table 1.3), data on relevant wholesale prices of drugs was obtained by surveying them (as discussed in section 1.2.4); drug sellers' profit margins were subsequently calculated as the difference between the price a drug seller charged and the average wholesale price.

Drug sellers' margins on drugs are substantial, as table 1.6 (in combination with table 1.7) shows: across the four illness conditions, the average treatment sold by drug sellers cost 9.49 Ghana Cedis, of which 4.31 Ghana Cedis (45%) were drug sellers' profits (as the average treatment cost 5.18 Ghana Cedis when procured from wholesalers). Differences in profit margins across illness conditions were relatively small: with the exception of diarrhoea (56%), margins were between 41% and 44%; absolute profits were also relatively similar across illness conditions and ranged from GhC 3.85 (diarrhoea) to GhC 5.02 (malaria). Tables 1.6 and 1.7 provide detailed information on average drug sellers' profits and average client expenditures when seeking treatment.[15]

[15] To only consider interactions in which drug sellers provided the full treatment themselves, this analysis excludes mystery client visits in which drug sellers referred mystery clients to a hospital or to another provider. However, results are similar when also including observations in which drug sellers referred the mother and are reported in the tables 1.A5 and 1.5.

Table 1.6: Profits from drugs: by illness condition

	all	Malaria	Diarr.	Resp. Inf.	Anaem.
Total profits	4.31	5.02	3.85	4.06	4.31
- from appropriate drugs	1.00	1.40	0.63	0.51	1.54
- from unnecessary drugs	1.88	2.54	0.31	2.57	2.27
- from harmful drugs	1.17	0.12	2.92	0.98	0.50
- from outdated antimalarials	0.26	0.97	0.00	0.00	0.00
Observations	283	75	77	69	62

Average profit in Ghana Cedis. Sample: interactions in which the drug sellers sold drugs for treatment and did not refer the client to another provider.

Table 1.7: Expenditure on drugs: by illness condition

	all	Malaria	Diarr.	Resp. Inf.	Anaem.
Total expenditure	9.49	12.31	6.88	9.18	9.69
- on appropriate drugs	2.24	3.81	0.86	0.84	3.60
- on unnecessary drugs	4.31	5.72	0.60	6.20	5.08
- on harmful drugs	2.27	0.21	5.41	2.14	1.02
- on outdated antimalarials	0.68	2.56	0.00	0.00	0.00
Observations	283	75	77	69	62

Average expenditure in Ghana Cedis. Sample: interactions in which the drug sellers sold drugs for treatment and did not refer the client to another provider.

Drugs not indicated by treatment guidelines accounted not only for a substantial part of expenditure, but also for a substantial share of drug sellers' profits: across illness conditions, unnecessary drugs (mostly cough syrups, painkillers, and vitamin tonics) accounted for 45% of client expenditures and 44% of drug seller profits. Harmful drugs (mostly antibiotics sold for cases of diarrhoea) accounted for another 24% of client expenditures and 27% of drug seller profits. Overall, the sale of unnecessary and harmful drugs therefore yielded substantial financial benefits to drug sellers. 77% of drug sellers' profits stem from the sale of unnecessary or even harmful drugs, while profits from the sale of medically appropriate drugs only account for 23% of total profits (table 1.6).

Given the substantial profits that the sale of inadequate (unnecessary and even harmful) drugs generates to drug sellers, it is not obvious that providing adequate treatments would be equally profitable. To investigate this, I hence conduct a simulation exercise: for each drug seller and illness condition, I analyse whether providing the adequate treatment could be equally profitable to the drug seller (and whether it would be more expensive to the client). The next section describes this simulation and its findings in detail.

1.4.2. Scenario 1: Strict adherence to treatment guidelines

The IMCI guidelines (discussed in section 1.2.3) specify the appropriate drugs for the treatment of each illness condition: these are artemisinin combination therapy (ACT) for the treatment of malaria, oral rehydration solution (ORS) and zinc for the treatment of diarrhoea, amoxicillin for the treatment of respiratory infections, and haematinic solution combined with dewormer for the treatment of anaemia. Although other drugs, such as cough syrups, pain killers, and vitamin syrups are commonly also sold by drug sellers, the drugs outlined above are sufficient from the medical perspective of the IMCI. Hence, in a first simulation exercise, I consider how strict adherence to the IMCI guidelines, a scenario in which drug sellers sell only the drugs indicated by the IMCI, but nothing else, would affect the profits of drug sellers. To do so, I construct for each drug seller and illness condition a portfolio of (one or several) drugs that adheres to the IMCI guidelines. This is a simple task in this set-up: as the IMCI specifies the appropriate drug for the treatment of each illness, the only choice lies in selecting a particular brand for each drug.[16] For each brand of drug, I set the retail price equal to the median retail price encountered by mystery clients;[17] the profit a portfolio would yield to

[16]I restrict my choice set to a subset of brands encountered during mystery client visits and only consider drugs that were in stock among the wholesalers surveyed and thereby had verifiable wholesale prices. This results in a choice between 8 brands of ACT, 3 brands of ORS, 1 brands of zinc, 3 brands of amoxicillin, 9 brands of haematinic liquid, and 7 brands of dewormer.

[17]Constructing such alternative portfolios implies by definition that I am considering sales of drugs by drug sellers which I have not observed through mystery clients. As I have no information about the retail price at which a drug seller *would sell* a specific brand of drug, I set the hypothetical

Table 1.8: Expenditures and profits: IMCI-recommended treatments only

	all	Malaria	Diarr.	Resp. Inf.	Anaem.
Benchmark expenditure	9.49	12.31	6.88	9.18	9.69
Altern. portfolio: expenditure	7.21	10.00	3.88	5.80	9.56
- of which: beneficial drugs	7.21	10.00	3.88	5.80	9.56
- of which: unnecessary drugs	0.00	0.00	0.00	0.00	0.00
Benchmark profit	4.31	5.02	3.85	4.06	4.31
Altern. portfolio: profit	3.69	4.82	2.59	3.19	4.22
- of which: beneficial drugs	3.69	4.82	2.59	3.19	4.22
- of which: unnecessary drugs	0.00	0.00	0.00	0.00	0.00
Observations	283	75	77	69	62
share of visits: lower profits	0.57	0.51	0.71	0.51	0.52

Each column in the table summarises a portfolio of drugs constructed for the treatment of one illness condition. Vertically, the table is split into two halves: the top half considers client expenditure and disaggregates this (for each portfolio) into spending on drugs prescribed by treatment guidelines and on unnecessary drugs (on which spending is zero by definition in this simulation exercise); the bottom half of the table similarly disaggregates drug sellers' profits into profits from drugs prescribed by treatment guidelines and from unnecessary drugs (from which profits are also zero by construction). The rows 'benchmark expenditure' and 'benchmark profits' display the clients' average expenditure on drugs and the drug sellers' average profits from the drugs sold for each illness condition as observed through mystery client visits. Drug portfolios for each illness conditions are constructed so that they strictly follow the IMCI guidelines, using the average prices (retail and wholesale) of each drug.

the drug seller is hence the difference between its retail price and its wholesale price. Among all the portfolios appropriate to treat an illness, I then choose for each drug seller and illness condition the most profitable portfolio that is not more expensive than the portfolio sold to mystery clients.[18]

retail price for each brand to be equal to the median retail price for the brand observed by mystery clients.

[18] I do this to ensure that every client would still be willing (and able) to purchase this portfolio. However, for 5 mystery client visits for respiratory infection and 8 for anaemia, no sufficiently cheap adequate portfolio exists. For these cases, I hence manually set the retail price of the drug to be equal to the total price mystery clients paid in the corresponding interaction. This retail price reduction lowers profits accordingly; however, they remain positive for 12 of these 13 cases.

Table 1.8 presents the results of this simulation exercise. Across all four illness condition, 57% of drug sellers would see their profits reduced when selling the most profitable adequate treatment (that does not raise their clients' expenditure). Conversely, 43% of drug sellers would be able to maintain their profits when treating clients correctly. Relative to the 31% of drug sellers that provided adequate treatment to mystery clients (although potentially also selling unnecessary or harmful drugs to them, as discussed in section 1.3.1 and table 1.4), this suggests that profit incentives are not a binding constraint to better treatment. They appear to be, however, a constraint to the *universal* provision of adequate treatment, as 57% of drug sellers would see their profits reduced when strictly adhering to the IMCI medical guidelines.

Substantial qualitative differences in drug sellers' ability to provide adequate treatments (that are at least equally profitable and at most as expensive as the treatment they provided to mystery clients) exist between the four illnesses; diarrhoea stands out in particular. When only providing the diarrhoea treatment foreseen by the IMCI, 71% of drug sellers would see their profits reduced; specifically, profits from treating diarrhoea would on average be 33% lower in such a scenario. For malaria, respiratory infections, and anaemia, on the other hand, fewer drug sellers (roughly one in two), would see their profits reduced when providing the treatment specified by the IMCI. These differences originate from two sources: firstly, the benchmark (in terms of profitability and expenditure) that drug sellers set for themselves (through their behaviour in mystery clients visits) was differently demanding between illness conditions; secondly, the availability of differently priced and differently profitable drugs that drug sellers could sell to treat each illness (thereby allowing them to generate profits as high as possible without exceeding the expenditure mystery clients incurred) varied between illness conditions.[19]

[19]For the treatment of malaria, a wide range of differently priced (and differently profitable) ACTs was available to drug sellers and could be procured from wholesalers. Nevertheless, a substantial number of cases when drug sellers could not sell an equally profitable treatment in the simulation exist; these are generally cases in which drug sellers had sold mystery clients several drugs together (often an antimalarial, vitamin syrup, and a pain killer), leading to larger expenditure (and subsequently higher profits). The profits obtained in such interactions were then difficult to attain when only being allowed to sell an ACT. For respiratory infections, the problem was generally similar: in a substantial number of cases, drug sellers had sold mystery clients rather expensive cough syrups, or had combined them with painkillers, co-trimoxazole, or an antimalarial (commonly quinine, an outdated monotherapy). This increased the cost (and profitability) of

Generally, for malaria and respiratory infections, drug sellers who had sold mystery clients a bundle of several (costly but also profitable) drugs saw their profits reduced when only being allowed to sell the appropriate treatment in the simulation; this occurred as ACTs and antibiotics were generally cheaper than the bundles drug sellers had sold before and hence left less room for profits. For anaemia, on the other hand, drug sellers who had sold inadequate (but profitable) vitamin tonics to mystery clients saw their profits reduced when they had to sell haematinic liquids with lower margins instead (cheaper haematinic liquids that would have potentially allowed to sell another product in addition, generating overall higher profits without increasing expenditure above the mystery client benchmark, were available in some cases, but strict adherence to the IMCI did not allow drug sellers to sell these additional products). For diarrhoea, the principal problem (and cause) of the widely experienced decrease in profits lay in the cheap price (and low absolute profits) of ORS and zinc, the correct treatment. Sold as dissoluble powder in little sachets, one sachet of ORS is sold (at the median) for GhC 1 by drug sellers; a blister of zinc tablets was similarly cheap. Even when selling the maximum amount of ORS and zinc that made sense from the perspective of the IMCI (three sachets of ORS and one blister of zinc tablets), the treatment for diarrhoea only yielded GhC 4 in revenue (and about GhC 2.66 in profits) to drug sellers; this was simply insufficient, in a large majority of cases, to match the profits that drug sellers had attained from (generally inadequately) treating mystery clients.

---

these portfolio to an extent that was not possible to obtain by selling the appropriate (modestly priced and modestly profitable) antibiotic. For anaemia, on the other hand, the situation was somewhat different: here, drug sellers who had treated mystery clients inadequately had often sold them relatively cheap (but profitable) vitamin tonics that did not contain any (or sufficient) iron for the treatment of anaemia. However, haematinic liquids were more expensive than vitamin tonics at wholesalers. This resulted in a large number of cases in which (in the simulation) drug sellers would have to sell a relatively cheap haematinic liquid in order to maintain expenditure at a sufficiently low level (relative to mystery client visits), which then in turn left them with little possibility to earn sufficient profits. Diarrhoea was the most problematic illness condition to treat adequately (and equally profitably, compared to mystery clients); this was due to the nature of the appropriate treatment. Although they carry high profit margins, both ORS and zinc are very cheap. Even when selling the maximum amount of ORS sensible from the perspective of the IMCI, drug sellers hence commonly could not obtain the same level of profits that they had obtained by selling unnecessary or inadequate (both more expensive and profitable) treatments to mystery clients; given the low retail price of ORS and zinc, they simply generated too little revenue on which drug sellers could earn profits.

Results from the simulation exercise are hence two-fold: firstly, more than 7 out of 10 drug sellers would see their profits reduced when treating diarrhoea as recommended by the IMCI guidelines. Thus, simply put, adequate treatment of diarrhoea is relatively unprofitable to drug sellers. This provides an intuitive explanation for the large differences between drug sellers' knowledge and practice in treating diarrhoea. Secondly, for the three other illnesses, for which no difference between drug sellers knowledge and practices existed, however, a substantial share of drug sellers, roughly 1 in 2, would also see their profits reduced when strictly adhering to IMCI medical guidelines. Drug sellers placing sufficient value on their clients' health outcomes (relative to their own profits) might be willing to forego such profits. However, those not willing to do so have reason to not provide better treatment: adherence to the IMCI guidelines is costly in terms of foregone profits.

Yet, such strict adherence to the treatment guidelines might be an unnecessarily demanding benchmark in the first place: a policy maker might primarily be interested in whether drug sellers have financial (dis-)incentives to provide adequate treatment, while still accepting that drug sellers might also sell medically unnecessary but harmless drugs, such as cough syrups, vitamin tonics, and pain killers. Rather than strict adherence to the IMCI guideline, a more lenient definition might hence consider any treatment that a drug seller provides as adequate if it includes the drugs indicated by the IMCI algorithm and no harmful drugs, while still allowing unnecessary drugs or items to be sold in addition. In the next section, I will hence relax the definition of appropriate treatment accordingly and conduct a second simulation exercise to investigate whether, when allowed to also sell unnecessary but harmless drugs, drug sellers would face similar financial incentives discouraging better treatment.

1.4.3. Scenario 2: Provision of adequate treatment

Following the discussion of the previous section, I repeat the simulation exercise, but now allow drug sellers to also sell unnecessary but harmless items in addition to the appropriate treatments for each illness condition. Intuitively, being able to sell additional items allows drug sellers to increase the revenue on which they can earn profits (unless the expenditure constraint is binding); this in turn should increase the profits which drug sellers are able to earn, while still treating each illness appropriately in accordance with IMCI guidelines.

A definition of adequate treatment in which drug sellers sell medically unnecessary, but harmless items, such as cough syrups, pain killers, and vitamin tonics to clients, might at first appear unusual. It is important to understand, however, that all treatments provided that adhere to this definition are medically adequate and thereby pareto improvements; their only drawback is that they might be somewhat more expensive to clients. Put formally, the combinations of drugs that drug sellers are now allowed to sell must fulfil three conditions to be pareto improvements: (1) the quality of treatment provided must be higher compared to the status quo[20] (2) without reducing drug sellers' profits or (3) increasing clients' cost of treatment. Pareto improvements are hence incentive-compatible: drug sellers would not see their profits reduced and would hence be willing to sell such pareto-superior portfolios of drugs to clients, while clients would not see their expenditure increased and would hence be willing to buy such a portfolio. On the other hand, pareto improvements do not require that drug sellers exclusively sell drugs that are appropriate for the treatment, even when sold alongside unnecessary drugs, providing the adequate treatment

This second simulation exercise hence provides a compelling benchmark to test whether financial incentives are impediments to the provision of better treatment by drug sellers: if drug sellers could sell a bundle of drugs that (1) includes the adequate drug(s) for the treatment of each illness condition (without including harmful drugs), (2) is at least as profitable to drug sellers (as the status quo), but (3) does not cost more to clients than the status quo, then financial incentives are no impediments to adequate treatment. Applying the same approach as in the first simulation exercise, I hence select, for each drug seller and illness, the most profitable portfolio (if at least one exists) that satisfies these three conditions. The only modification, relative to the last exercise, is in allowing drug sellers to also sell unnecessary but harmless items in addition to the appropriate treatment. Namely, for all illness conditions, drug sellers are now also allowed to sell a vitamin

[20] In this case, my definition is more demanding: treatments provided need to be fully adequate from a medical perspective, rather than just "better".

syrup (choosing among 20 vitamin syrups that were sold to mystery clients and for which wholesale prices were available); furthermore, for malaria, drug sellers are also allowed to sell a paracetamol (choosing among 12 brands), and for respiratory infections a cough syrup (choosing among 11 brands).

Table 1.9: Expenditures and profits: adequate treatments

	all	Malaria	Diarr.	Resp. Inf.	Anaem.
Benchmark expenditure	9.49	12.31	6.88	9.18	9.69
Altern. portfolio: expenditure	8.27	10.74	5.92	7.46	9.10
- of which: beneficial drugs	5.97	5.77	4.62	4.88	9.10
- of which: unnecessary drugs	2.29	4.97	1.30	2.59	0.00
Benchmark profit	4.31	5.02	3.85	4.06	4.31
Altern. portfolio: profit	4.79	5.97	4.27	4.73	4.08
- of which: beneficial drugs	3.31	2.82	3.44	3.02	4.08
- of which: unnecessary drugs	1.47	3.13	0.83	1.71	0.00
Observations	283	75	77	69	62
share of visits: lower profits	0.08	0.00	0.01	0.03	0.31

As in the previous table, each column in the table summarises a portfolio of drugs constructed for the treatment of one illness condition. Vertically, the table is split into two halves: the top half considers client expenditure and disaggregates this (for each portfolio) into spending on drugs prescribed by treatment guidelines and on unnecessary drugs (on which spending is zero by definition in this simulation exercise); the bottom half of the table similarly disaggregates drug sellers' profits into profits from drugs prescribed by treatment guidelines and from unnecessary drugs (where profits can now be positive, as the simulation permits drug sellers to also sell unnecessary drugs). The rows 'benchmark expenditure' and 'benchmark profits' display the clients' average expenditure on drugs and the drug sellers' average profits from the drugs sold for each illness condition as observed through mystery client visits. Drug portfolios for each illness conditions are constructed so that they strictly follow the IMCI guidelines, using the average prices (retail and wholesale) of each drug.

Table 1.9 displays the results of this simulation exercise: in 92% of mystery client visits, financial incentives were no impediment to adequate treatment; as the table displays, in these visits drug sellers could have provided an adequate treatment that would have generated the same level of profit to them and cost at most as much to clients. Malaria, diarrhoea, and respiratory infections stand out in particular; here drug sellers could in almost all cases have provided medically fully adequate treatment without lowering their profits or increasing the expenditure of their clients. Being allowed to sell unnecessary but harmless vitamin tonics in this scenario, drug sellers were now also able to treat diarrhoea adequately without reducing their profits. While the sale of unnecessary drugs has little benefit to clients' health, it is of large benefit for aligning drug sellers' incentives and ensuring that the provision of adequate treatments does not result in a reduction in profits. Unfortunately, this benefit was slightly muted for anaemia. Here, drug sellers would only be able to maintain their profit levels in 70% of mystery client visits; still, this is a substantial improvement over the quality of treatment observed in mystery client visits (where 25% of drug sellers provided adequate treatment and 17% referred the client). Furthermore, the 31% of mystery client interactions for anaemia in which drug sellers were not able to provide adequate treatments in the simulation exercise were overwhelmingly cases in which the (incorrect) treatments that drug sellers had provided to mystery clients had been relatively cheap but included profitable vitamin tonics.[21] Since haematinic liquids, the appropriate treatment, had a higher wholesale price and was hence less profitable (when sold to mystery clients at prices similar to vitamin tonics), the reduction in drug sellers' profits in the simulation exercise is unsurprising. Nevertheless, although pareto improvements are not possible for these drug sellers (as this does not allow for raising clients' expenditure), drug sellers could also provide adequate treatment in these cases anaemia when (slightly) raising expenditure to clients to an average level.

For an overwhelming majority of mystery client interactions, financial incentives are hence not a constraint to the provision of adequate treatment by drug sellers: taking as benchmark the profits that drug sellers earned when treating mystery clients, drug sellers would be able to provide adequate medical treatment

[21]Mean client expenditure in this subsample of interactions was low at 6.68 GhC (compared to 9.69 GhC in the unconditional sample for anaemia). Since the wholesale price of haematinic liquids was higher than the wholesale price of vitamin tonics, drug sellers then failed to maintain profit levels when having to sell adequate haematinic liquids that did not cost more to clients but were more expensive to procure from wholesalers. For these 31% of drug sellers, profits then dropped by 49% at the median (and 12% and 75% at the 10th and 90th percentile, respectively) when providing adequate treatment.

without reducing their profits (in fact, profits would be 11% higher) in 92% of interactions.

1.5. Summary and Conclusion

Pneumonia, diarrhoea, and malaria are infectious diseases which are relatively easy to diagnose and to treat. Yet, they are the deadliest childhood diseases in sub-Saharan Africa, accounting for 44% of deaths of under-five children (Liu et al., 2012). Across many developing countries, private pharmacists and chemical/drug sellers are important sources of medical advice and treatment, as trained medical providers are often not consulted in cases of illness (Sudhinaraset et al., 2013; Peters and Bloom, 2012). However, surprisingly little is known about the quality of medical care that drug sellers provide for childhood illness and about the determinants of this quality. I hence (1) provided empirical evidence on the quality of diagnosis and treatment that drug sellers provide for common childhood illnesses, (2) studied the extent to which drug sellers have knowledge of appropriate diagnostic and treatment procedures, and (3) investigated to what extent adverse financial incentives prevent drug sellers' from providing adequate treatment. To do so, I collected an original dataset on drug sellers' treatment practices, knowledge, and financial incentives in Northern Ghana.

The quality of treatment for childhood illness that drug sellers provide was overwhelmingly inadequate and violated treatment guidelines; when selling drugs, drug sellers were twice as likely to sell inadequate or harmful drugs as to sell the drugs recommended by treatment guidelines. However, substantial differences existed across illness conditions. Knowledge was found to be a principal bottleneck to better treatment for three of the four illness conditions, although some significant differences between drug sellers' self-reported and actual treatment practices existed: when drug sellers were asked (rather than secretly observed) how they would treat a child presenting symptoms of one of the four illness conditions, drug sellers were more likely to refer the child to a formal provider for all illness conditions and provided substantially better treatment for diarrhoea. However, for the remaining three illness conditions - malaria, respiratory infections, and anaemia - drug sellers did not provide better treatment. Financial incentives do not impede the provision of better treatments: for each illness conditions, drug sellers could sell bundles of drugs to clients, which would provide adequate treatment to the child, while still yielding similar profits to drug sellers and not costing more to clients as the largely inadequate treatments currently sold.

1.6. Appendix

Table 1.A1: Questions asked (mystery client visits)

	all	Malaria	Diarrhoea	Resp. Inf.	Anaemia
questions: total	2.86	3.03	2.71	2.71	2.99
questions: relevant	2.23	2.58	2.22	2.07	2.03
Observations	292	76	77	70	69

Average number of questions asked. Sample: interactions in which drugs were sold (with or without referral). This covers 91% of actual interactions.

Table 1.A2: Questions asked: combined

	actual	stated	difference (se)
Seeing the child			
general: see child	0.09	0.19	-0.10 (0.04)*
General questions			
general: age	0.88	0.83	0.05 (0.04)
danger signs: vomits	0.30	0.26	0.04 (0.04)
danger signs: does not drink	0.16	0.09	0.07 (0.03)*
danger signs: convulsions	0.02	0.04	-0.02 (0.02)
danger signs: unconsc./letharg.	0.00	0.03	-0.02 (0.01)
Malaria: illness-specific questions			
fever: duration	0.53	0.56	-0.03 (0.08)
sign of measles: cough	0.12	0.06	0.06 (0.05)
fever: runny nose	0.12	0.03	0.09 (0.04)*
bacterial cause: abdom pain	0.03	0.01	0.01 (0.02)
bacterial cause: tenderness	0.01	0.00	0.01 (0.01)
bacterial cause: sores	0.01	0.01	-0.00 (0.02)
sign of measles: red eyes	0.00	0.04	-0.04 (0.02)
sign of measles: rash	0.00	0.00	0.00 (0.00)
fever: stiff neck	0.00	0.01	-0.01 (0.01)

Continued on next page

Table 1.A2 – continued from previous page

	actual	stated	difference (se)
fever: measles in prev 3 months	0.00	0.01	-0.01 (0.01)
bacterial cause: swelling & boils	0.00	0.00	0.00 (0.00)
bacterial cause: avoids limb	0.00	0.00	0.00 (0.00)
Diarrhoea: illness-specific questions			
diarrhoea: duration	0.52	0.55	-0.03 (0.08)
diarrhoea: blood in stool	0.14	0.16	-0.02 (0.06)
dehydration: drinks eagerly	0.10	0.05	0.05 (0.04)
dehydration: sunken eyes	0.05	0.11	-0.06 (0.04)
dehydration: pinched skin slow	0.01	0.07	-0.05 (0.03)
dehydration: restless & irritable	0.00	0.07	-0.07 (0.03)*
Respiratory Infection: illness-specific questions			
cough: duration	0.44	0.53	-0.09 (0.09)
cough: breathing fast	0.21	0.20	0.01 (0.07)
cough: noisy breathing	0.20	0.07	0.13 (0.06)*
cough: chest indrawing	0.00	0.07	-0.07 (0.03)*
Anaemia: illness-specific questions			
pale: where pale	0.61	0.50	0.11 (0.19)
pale: dewormed recently	0.22	0.08	0.14 (0.07)*
Observations	292	257	

*Number of questions asked. Sample: observations in which drugs were sold (with or without referral of the child). This covers 91% (mystery clients) and 87% (vignettes) of observations, respectively. Significance of differences: * for p <.05, ** for p <.01, and *** for p <.001*

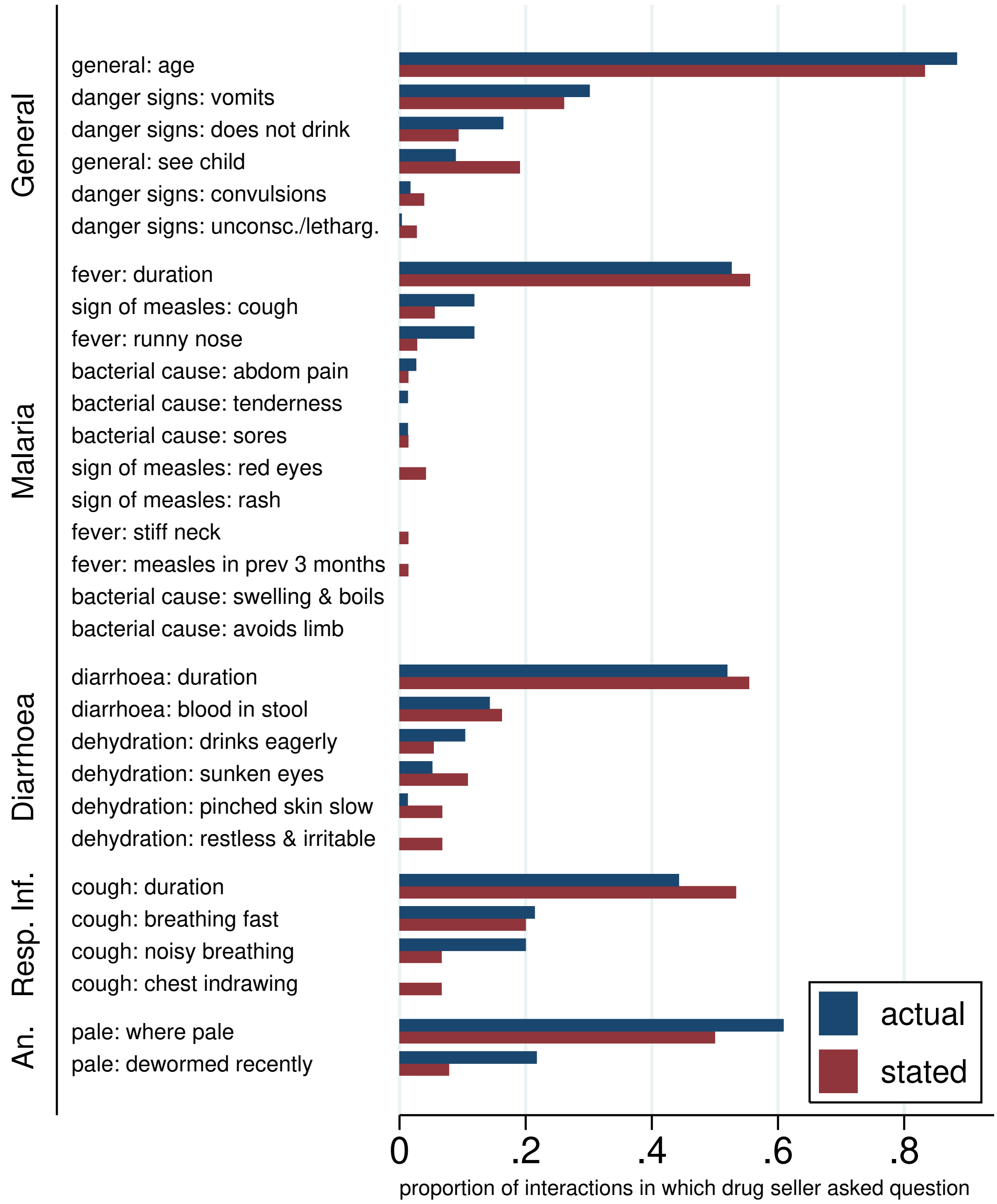


Figure 1.A1: Questions asked by drug sellers

Table 1.A3: Treatments provided: by illness condition

	actual	stated	difference (se)
Overall: all conditions combined			
adequate treatment	0.31	0.40	0.10 (0.04)*
referred	0.05	0.23	0.18 (0.03)***
inadeq. treatm. & no referral	0.64	0.36	-0.27 (0.04)***
sold harmful drugs	0.34	0.16	-0.18 (0.03)***
sold unnecessary drugs	0.64	0.55	-0.09 (0.03)**
Malaria			
adequate treatment (ACT)	0.46	0.46	-0.00 (0.08)
referred	0.01	0.10	0.08 (0.04)*
inadeq. treatm. & no referral	0.53	0.44	-0.08 (0.08)
- of which: sold monotherapies	0.39	0.29	-0.10 (0.08)
sold harmful drugs	0.04	0.03	-0.01 (0.03)
sold unnecessary drugs	0.93	0.96	0.02 (0.04)
- of which: painkillers	0.89	0.96	0.06 (0.04)
Diarrhoea			
adequate treatment (ORS)	0.34	0.85	0.51 (0.07)***
- and also sold zinc	0.18	0.77	0.59 (0.07)***
- but also sold harmful drugs	0.22	0.19	-0.03 (0.07)
referred	0.00	0.03	0.03 (0.02)
inadeq. treatm. & no referral	0.66	0.12	-0.54 (0.07)***
sold harmful drugs	0.88	0.31	-0.57 (0.07)***
- of which: antibiotics	0.86	0.30	-0.56 (0.07)***
sold unnecessary drugs	0.12	0.03	-0.09 (0.04)*
Respiratory Infection			
adequate treatment (Amoxicillin)	0.17	0.18	0.01 (0.06)
- but also sold harmful drugs	0.03	0.03	-0.00 (0.03)
referred	0.03	0.33	0.30 (0.06)***
inadeq. treatm. & no referral	0.80	0.49	-0.31 (0.08)***
sold harmful drugs	0.30	0.12	-0.17 (0.07)*

Continued on next page

Table 1.A3 – continued from previous page

	actual	stated	difference (se)
sold unnecessary drugs	0.94	0.74	-0.20 (0.06)***
- of which: cough syrups	1.00	0.93	-0.07 (0.03)*
Anaemia			
adequate treatment (Iron)	0.25	0.10	-0.15 (0.06)*
- and also provided dewormer	0.04	0.03	-0.01 (0.03)
referred	0.17	0.49	0.32 (0.07)***
inadeq. treatm. & no referral	0.57	0.41	-0.17 (0.08)*
sold harmful drugs	0.15	0.19	0.04 (0.06)
sold unnecessary drugs	0.57	0.48	-0.10 (0.08)
Observations	299	288	

*Proportion of interactions. Sample: interactions in which drugs were sold (with or without referral) or the child was referred (without drugs sold). This covers 93% of actual interactions and 90% of stated interactions. Significance of differences: * for p <.05, ** for p <.01, and *** for p <.001*

Table 1.A4: Expenditure on drugs: by illness condition

	actual	stated	difference (se)
Overall: all conditions combined			
Total expenditure	9.43	7.72	-1.72 (0.44)***
- on beneficial/adequate drugs	2.24	2.45	0.21 (0.34)
- on unnecessary drugs	4.30	3.71	-0.58 (0.25)*
- on harmful drugs	2.23	1.09	-1.14 (0.25)***
- on outdated antimalarials	0.66	0.46	-0.20 (0.11)
Malaria			
Total expenditure	12.34	11.70	-0.64 (0.89)

Continued on next page

Table 1.A4 – continued from previous page

	actual	stated	difference (se)
- on beneficial/adequate drugs	3.89	4.49	0.59 (0.84)
- on unnecessary drugs	5.71	5.45	-0.26 (0.54)
- on harmful drugs	0.21	0.13	-0.09 (0.15)
- on outdated antimalarials	2.53	1.63	-0.89 (0.49)
Diarrhoea			
Total expenditure	6.88	3.34	-3.53 (0.44)***
- on beneficial/adequate drugs	0.86	1.79	0.92 (0.18)***
- of which ORS	0.52	0.98	0.46 (0.12)***
- of which zinc	0.34	0.81	0.47 (0.10)***
- on unnecessary drugs	0.60	0.08	-0.52 (0.21)*
- on harmful drugs	5.41	1.47	-3.94 (0.45)***
- of which antibiotics	5.12	1.43	-3.70 (0.44)***
Respiratory Infection			
Total expenditure	9.14	8.41	-0.72 (0.87)
- on beneficial/adequate drugs	0.87	1.33	0.46 (0.43)
- on unnecessary drugs	6.16	5.76	-0.40 (0.49)
- of which cough syrup	5.67	5.32	-0.35 (0.34)
- on harmful drugs	2.11	1.32	-0.79 (0.66)
Anaemia			
Total expenditure	9.38	7.62	-1.76 (0.71)*
- on beneficial/adequate drugs	3.36	1.87	-1.49 (0.82)
- on unnecessary drugs	4.99	4.13	-0.86 (0.78)
- of which vitamins	4.93	3.91	-1.02 (0.76)
- on harmful drugs	1.03	1.62	0.59 (0.50)
Observations	292	257	

*Expenditure in Ghana Cedis. Sample: observations in which drugs were sold (with or without referral of the child). This covers 91% (actual) and 87% (stated) of observations, respectively. Significance of differences: * for p <.05, ** for p <.01, and *** for p <.001*

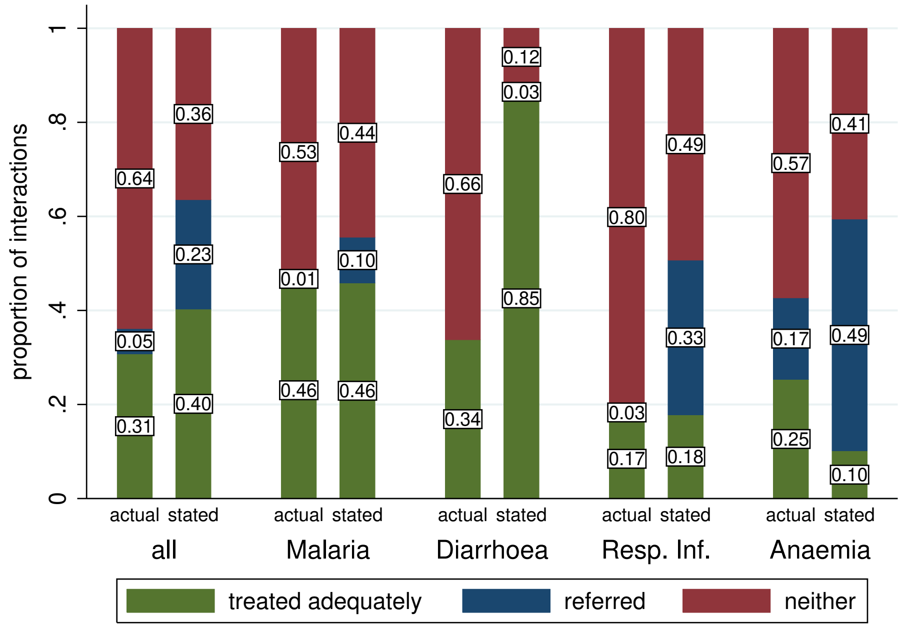


Figure 1.A2: Treatments provided (actual vs. stated). Identical to figure 1.2, with numerical labels

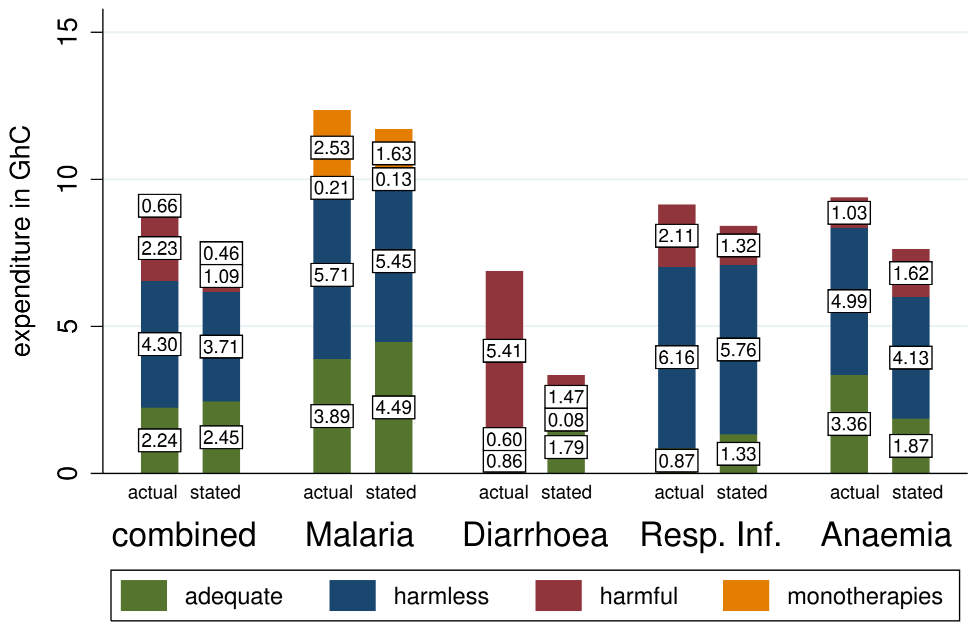


Figure 1.A3: Expenditure on treatment by condition (actual vs. stated). Identical to figure 1.3, with numerical labels

Table 1.A5: Profits from drugs: by illness condition (including referrals)

	all	Malaria	Diarr.	Resp. Inf.	Anaem.
Total profits	4.18	5.07	3.85	3.96	3.83
- from appropriate drugs	0.98	1.42	0.63	0.52	1.32
- from unnecessary drugs	1.83	2.57	0.31	2.49	2.02
- from harmful drugs	1.13	0.12	2.92	0.95	0.49
- from outdated antimalarials	0.24	0.96	0.00	0.00	0.00
Observations	299	76	77	71	75

Profits in Ghana Cedis. Sample: observations in which drugs were sold (with or without referral of the child) or the child was referred (without drugs sold). This covers 92% (actual) of observations. The only observations excluded are if the drug seller was out of stock, refused to proceed without seeing the child, or was not around.

Chapter 2

IS RURAL-URBAN MIGRATION SORTING? (AND HOW STABLE IS IT?) EVIDENCE FROM TANZANIA

2.1. Introduction

Across much of the developing world, there is an urban-rural gap: individuals' productivity, wages, and consumption tend to be higher in urban than in rural areas (Young, 2013; Gollin et al., 2014), suggesting that urban residents lead better lives than their rural counterparts. Considerable debate exists about the cause of these urban-rural differences; identifying this may have large implications for policy. One class of explanations suggests that cities *cause* these differences due to their better organisation, more efficient labour markets, or other channels (Duranton and Puga, 2004; Behrens et al., 2014): urban residents earn more than their rural counterparts per unit of labour *because* they live in cities. A second class of explanations attributes the observed urban-rural differences to unobserved individual heterogeneity in skill and selective migration, or sorting, of individuals between rural and urban areas: higher-skill individuals have sorted to urban areas and earn more than their rural counterparts per unit of labour because of their higher (unobserved) skill (Lagakos and Waugh, 2013). In this case, moving to an urban area would only be beneficial for high-skilled rural residents and migration should be selective (based on skill).

In this paper I hence study migration flows from rural households to urban areas to investigate whether migration between rural and urban areas is indeed selective. Finding evidence of such sorting, I examine how changes in the welfare of rural households affect the out-migration of its members to urban areas.

While explanations relying on unobserved forces may not be particularly appealing, sorting based on unobserved skill is a potentially compelling explanation of the rural-urban gap: it can not only explain higher wages in urban areas and migration flows *to* urban areas, but can also make sense of the (substantial) migration flows *from* urban to (lower-earning) rural areas (Young, 2013). Furthermore, it is a testable theory: using data that contains information on both individuals' skills (proxied by education) and migration behaviour, I analyse whether migration, especially from rural to urban areas is indeed selective. Moreover, drawing on data on the economic situation of individuals' households, I attempt to cast light on the circumstances in which such migration takes place.

Using data from Tanzania, I re-purpose a particularly rich and nationally representative household panel survey to study the selectivity of domestic migration and investigate (1) the extent to which there is rural-urban sorting by (observable) skill, (2) the extent to which migrants to urban and rural areas differ on a wide range of individual-level characteristics, and (3) the extent to which changes in the welfare of agricultural households affect rural-urban sorting and out-migration. I construct measures of individuals' flows out of and into households from rosters of the repeatedly surveyed households and take advantage of the tracking conducted by the survey to determine the subsequent location of individuals leaving the household. The analysis yields three sets of results. There is a large amount of movement and migration out of and into households in Tanzania: in a given two-year period (the spacing of survey rounds), one out of two households experiences the departure or arrival of a new household member and 10-12% of Tanzanians leave their previous household. Even at first sight, there is evidence of selective migration out of agricultural households: the more educated half of individuals (with seven or more years of education)[1] are 30% more likely than their less educated peers to leave their household (13.3% vs. 10.4%) and, staggeringly, are three times as likely to move to an urban area as their less educated peers (3.3% vs. 1.1%). This causes the group of individuals who leave agricultural households and move to urban areas to be heavily selected: 30% of these movers have nine or more years of education (compared to 11% in the general population), while only 14% have four or less years of education (compared to 37%).

Considering a wide range of covariates, I then examine how individuals leaving agricultural households and moving to urban areas differ from their peers who stay behind or move to rural areas. In line with the results on selection by education, I find movers to urban areas to be a specific group of individuals who differ significantly from both movers to rural areas and the general population. Prior to moving, movers to urban areas are twice as likely to earn a salary (which is substantially higher than that of movers to rural areas) when working; however, they are also more likely to report being unemployed at their agricultural households, suggesting a potential lack of commensurate employment opportunities. Furthermore, urban movers report better subjective health (which corresponds to objective indicators) but expressed more dissatisfaction with their job and financial situation at their agricultural households. Movers to rural areas, on the other hand, do not exhibit such significant differences to the general population.

Last, I investigate how sensitive the out-migration of (differently educated) individuals is to changes in the economic situation of these agricultural households. To do so, I employ data on shocks to the economic welfare of households collected by the survey, which asks households about the major events that have impacted their welfare since the previous survey wave. Focussing my attention on the most commonly reported and plausibly exogenous shocks (droughts/floods, increases in the price of food, decreases in the sale price of agricultural crops, and increases in the price of agricultural inputs), I find that changes in households' economic conditions have a substantial impact on the sorting of their more educated members to urban areas. Falls in the sale prices of harvested crops more than halve the

[1] Here and subsequently, I am restricting the analysis to individuals who are at least 14 years or older and thereby could have obtained up to nine years of education when starting school at age five

proportion of households that have a more educated member move to an urban area (according to point estimates from 6.3% to 3%), whereas increases in the prices of agricultural inputs significantly increase the incidence of such migration (by 70%, from 6.3% to 10.8%). These results suggest that sending households' economic conditions have a significant impact on the out-migration of individuals to urban areas. Given the different nature and timing of shocks, I interpret my findings to suggest that being able to finance the migration of their household members to urban areas might be an important pre-condition (and bottleneck) to the sorting of more educated individuals to urban areas.

This paper relates to several recent studies, which have proposed and evaluated skill-based sorting as an explanation for the urban-rural or agricultural productivity gap in developing countries. Economists have long noted large differences in productivity between agricultural and non-agricultural sectors, present particularly in developing countries (Lewis (1954), Kuznets (1971)). Carefully working with household surveys from 10 countries, Gollin et al. (2014) establish that large (approximately two-fold) differences in productivity between agricultural and non-agricultural sectors are a robust finding: these differences are not just present in national accounts data, but also when taking advantage of more detailed micro-level data and carefully adjusting for sectoral differences in labour inputs, most importantly hours worked and human capital per worker.

Taken at face value, these findings suggest that re-allocating workers from agricultural sectors to modern, non-agricultural sectors would substantially raise aggregate productivity and income. The lack of such large-scale re-allocation is hence the puzzle that recent studies of the agricultural productivity gap aim to explain. Lagakos and Waugh (2013) are the first to do so through a theory of skill-based sorting. In their model, closed economies with two sectors (agriculture and non-agriculture) differ in their overall productivity. Labour is provided by individuals, who are subject to a subsistence requirement (i.e. they need to consume a minimum amount of the agricultural good) and choose the sector they work in. Sorting is generated by individual heterogeneity: each individual draws (positively correlated) productivities of labour in the agricultural and non-agricultural sector (however, non-agricultural labour productivities have a higher variance). These three components are sufficient to generate an agricultural productivity gap in poor countries that persists in equilibrium.[2]

Young (2013) similarly asks whether sorting *could* explain urban-rural inequality, drawing on a large amount of microdata to conduct a less stylised analysis. Examining 170 DHS surveys covering 65 countries and 2.1 million households, Young (2013) first shows that large migration flows between rural and urban areas exist *in both directions*: 20-25% of individuals raised in rural areas move to cities as young adults (where they earn more than their rural peers who stayed behind), while a similar proportion of individuals raised in urban areas moves to rural areas, where they earn less than their urban peers who stayed behind. If cities causally raised their residents' incomes, these urban-to-rural migrants would seem to choose a worse life; a model of sorting, on the other hand, could make sense of their choices. Young then shows that a model in which individuals sort geographically by skill can explain these stylised facts.[3] Starting from the assumptions that industries located in urban areas are more skill-intensive and that more educated individuals are more likely to be skilled, Young builds a model that attributes higher urban incomes solely to unmeasured skill and argues that his model maintains a sufficient distance between assumptions and results. He readily admits that explaining the urban-rural gap through (unmeasured) skill might be unsatisfying; understanding his paper as a "thought experiment" that leaves room for further examination of what this skill is might hence be an appropriate interpretation.

Hicks et al. (2017) take a different approach to investigating whether individual skill could account for the urban-rural gap: rather than relying on cross-sections, the authors use panel data to study the impact of rural-to-urban migration

[2]Low overall productivity requires a large share of the population to work in agriculture to satisfy the subsistence requirement, and only the highest-productivity individuals work in the non-agricultural sector. Since these individuals have higher non-agricultural than agricultural productivities (as they are drawn from a distribution with higher variance), a productivity gap between the two sectors naturally emerges. In richer countries with higher overall productivity, this gap is smaller: there, the subsistence requirement can be fulfilled by a smaller number of individuals working in the agricultural sector, and a larger (and "less elite") share of individuals works in the non-agricultural sector.

[3]The author establishes through a decomposition exercise that within-country inequality is driven to a substantial extent by inequality between rural and urban areas whereas, within areas, education accounts for little inequality: sorting by education can hence not explain the observed urban-rural differences

on (moving) individuals' productivity and consumption, which allows them to explicitly control for individual skill through fixed effects. Their estimates of the urban-rural gap, identified by the individuals moving from rural to urban areas, are over 80% smaller than estimates derived from cross-sectional data, which suggests that, if one is willing to extrapolate from the local nature of their estimates, unobserved individual characteristics are major causes of the urban-rural gap. Skill might indeed be one such characteristic, as the authors find that rural-to-urban migrants have significantly higher cognitive ability as measured by Raven's z-score (which is 0.3 standard deviations higher for this group).

For Tanzania, Beegle et al. (2011) investigate the impact of migration on individuals' consumption by implementing an impressive 2004 follow-up (including tracking) to a 1991/94 household survey in north-western Tanzania. Migration, they find, is associated with large (36 ppts) increases in consumption relative to non-movers; increases are particularly stark for individuals who moved out of their rural communities and subsequently worked in the non-agricultural sector. Using panel data, the authors can credibly identify the consumption gains associated with migration for moving individuals; however, given the self-selection of individuals into migration (and lacking exogenous variation or convincing instruments), generalising their findings to non-migrating individuals takes a leap of faith. Kubik and Maurel (2016) also analyse migration in Tanzania, using climate data to study the impact of weather shocks on the incidence of individuals leaving agricultural households; however, methodological difficulties make it difficult to interpret their findings.[4]

In contribute to this strand of literature by conducting an empirical study of rural-urban migration in the Tanzanian context. This paper is organised as follows: section 2 presents the data used in the analysis; section 3 then presents my findings on the incidence of migration and its selectivity. Section 4 considers the impacts of shocks to household welfare on migration and section 5 concludes.

[4]The authors find that a weather-induced 1 percent reduction in agricultural income leads to a 13 percentage point increase in the probability of migration in the following year. However, their climate data stems from largely interpolated weather datasets, which are unlikely to capture local variation in households' climatic conditions.

2.2. Data

2.2.1. The Tanzania National Panel Survey

The Tanzania National Panel Survey (TZNPS) is the main data source employed in my analysis. Part of the World Bank's Living Standards Measurement Study (LSMS) program, the TZNPS is of high quality and provides me with extensive information on the situation and living conditions of a nationally representative sample of households across Tanzania. Furthermore, it includes a "shock module", which asks households about recent shocks to welfare that they experienced (Heltberg et al., 2015). As a panel, the TZNPS is well-suited to study the flows of individuals into and out of households and track changes in household composition over time. Furthermore, since the survey aims to track and re-survey all individuals (who were 13 years or older during the previous wave) who move out of a surveyed household, I also have information on the destination of their movement.

2.2.2. Constructing measures of migration

The individual-level panel structure of the TZNPS makes it easy to study the flows of individuals into and out of households (which I will subsequently refer to as "migration"). Each wave of the survey contains a household roster and detailed information on each household member. Therefore, by combining the different waves, I can identify out-migrating members; individuals who are listed in the household roster of one, but not the subsequent wave.[5] Similarly, I identify in-migrating individuals as those who are listed in the household roster of one wave but were not present in the previous wave (accounting for births, using data on age). Since in both cases the individual has been part of the household in one wave, I have detailed data on individual characteristics.

Using data from the three waves available, I examine patterns of migration into and out of households over two periods of time, between waves 1 and 2 (2008/09 - 2010/11) and between waves 2 and 3 (2010/11 - 2012/13) and code household-level dummy variables, indicating whether the household experienced

[5] Data from death rosters allows me to account for individuals who died.

the outflow or the inflow of an individual during each period of time. I use the TZNPS' rural/urban classification to categorise areas as rural or urban. My results are not driven by any potential particularities of this classification, however; in section 3, I show in robustness tests that migration patterns are similar when categorising households as agricultural/non-agricultural or as rural/urban. When I examine urban-rural sorting by education and consider how different amounts of education (seven or more years versus six or fewer years) affect out-migration and its destination, I restrict my attention to individuals over the age of 13 to ensure that every individual could have obtained at least seven years of education (when starting school at age six).

2.2.3. Analysis sample and sample size

I construct an individual-level panel by combining the three available waves of the TZNPS, 2008/09, 2010/11, and 2012/13. During the first wave, 3,265 households (containing 16,709 individuals across 409 enumeration areas) were surveyed, which comprise the basis of my panel. Since wave-1 households which moved in their entirety, as well as all individuals (over the age of 15) who left their wave-1 household were tracked and re-interviewed (alongside all members of their new "splitoff" household), the sample size of the second wave increased to 3,924 households (containing 20,559 individuals). Tracking again all wave-2 households who moved and all individuals (over the age of 15) who left their wave-2 household led to a further increase in the sample size of the third wave, which surveyed 25,412 individuals in 5,010 households.

I am forced exclude households who could not be tracked from the analysis, since I lack data on changes to their household composition; furthermore, I also exclude a small number households who disintegrated between two waves. My final sample consists of 3,112 households that I observe between waves 1 and 2, and 3,666 households that I observe between waves 2 and 3.[6] In the second part of my analysis, I consider how shocks to household welfare affect the inci-

[6]These sample sizes arise as follows: of the 3,265 households surveyed in wave 1, 56 had disintegrated by the time of wave 2 and 97 could not be found again. Of the 3,924 households surveyed in wave 2, 120 had disintegrated by the time of wave 3 and 138 could not be found again. Household attrition rates are 4.7% and 6.6%.

dence and direction of migration into and out of affected households. Since the most commonly reported shocks to welfare are primarily of agricultural nature, I restrict my attention to households engaging in agriculture in the second part. As I am interested in how changes in the economic circumstances of agricultural households affect the incidence and destination of out-migration of its members, I exclude households who moved in their entirety from the analysis. I do this to ensure that my results on the impact of shocks are not driven by the wider changes that moving households might experience.[7] My final sample in this part of the analysis hence consists of 2,024 households that I observe between waves 1 and 2, and 2,283 households that I observe between waves 2 and 3.[8]

Not all individuals present in a household at the time of interview will be household members; similarly not all household members will usually be present at the time of the interview. In defining household membership, I follow the definition of the TZNPS: all individuals who normally live and eat their meals together in the surveyed location are considered household members (as long as they have been present for at least 3 months in the previous year, or are new household members, newly born, or boarding school students).

Table 2.1 provides summary statistics on the households in my sample: columns 1 and 3 describe the full sample of households, columns 2 and 4 describe the sample of agricultural households; the left half of the table does so at the time of first wave in 2008/09, the right half at the time of the second wave in 2010/11. On average, agricultural households are larger, less educated, and poorer, both by income and incidence of stunting, than their non-agricultural counterparts. Notably, between seven and eight in ten agricultural households list agriculture as their main source of income.

[7] Reassuringly, regression specifications including these households confirm that my results are robust to the inclusion of these households.

[8] Of the 3,112 households that I observe between waves 1 and 2, 2,231 engage in agriculture, of those 207 moved in their entirety and are hence excluded. Of the 3,924 households surveyed in wave 2, 2,540 engage in agriculture, of those 217 moved in their entirety and are hence excluded.

Table 2.1: Descriptive statistics: households

	wave 1	wave 1	wave 2	wave 2
Demographics				
# of HH members	5.06	5.40	5.29	5.78
# of deps. (chldn & grndchldn of HH head)	2.40	2.65	2.46	2.77
HH has a female head	0.25	0.24	0.24	0.23
Years of education of HH head	5.56	4.71	5.61	4.71
Years of education of members 15-65	5.89	5.16	6.06	5.37
Consumption and welfare				
Cons. (ann., per ad. eq., 1000s real TSH)	752	554	852	627
HH worried about suff. food in prev. week	.	.	0.34	0.35
HH has a wasted child (0-4 years)	0.04	0.05	0.10	0.11
HH has a stunted child/adolesc. (0-18 yrs)	0.63	0.70	0.56	0.64
HH has access to piped water	0.43	0.34	0.38	0.29
Agricultural operations				
HH names agric. as a main source of inc.	0.56	0.77	0.51	0.72
Acres of land available to HH	3.93	5.58	4.46	6.39
Location characteristics				
Urban area	0.36	0.14	0.32	0.15
Cost of travel to district HQ (1000s TSH)	2.39	2.98	2.40	2.91
EA has primary school	0.85	0.92	0.87	0.92
EA has secondary school	0.36	0.39	0.35	0.41
Observations	3112	2024	3649	2283
Sample	all	agric.	all	agric.

2.2.4. Attrition

As with any survey data, (non-random) attrition may threaten the internal validity of findings derived from the data. However, in the context of this study, attrition is a lesser concern for two reasons: Firstly, the TZNPS attempts to track (and devotes significant resources to doing so) all households who moved (within Tanzania) and all household members who moved away from their original household (if they were older than 13 years at the time of the previous wave). Attrition in the TZNPS (particularly at the household level) is hence very low compared to other surveys: in wave 2, more than 97% of households interviewed in wave

1 are successfully re-interviewed and in wave 3, more than 96% of households interviewed in wave 2 are successfully re-interviewed. Secondly, a large part of my analysis is unaffected by individual-level attrition: since I use household rosters to determine whether individuals have migrated out of or into households and have individual-level information on migrating individuals from household surveys (from the round before out-migration for leaving individuals or from the round after in-migration for arriving individuals), individuals who were not successfully tracked are still part of my analysis. However, it is worth noting that my estimates on the *direction* of out-migration may be affected by individual-level attrition, as I only know the destination of migration for individuals who were successfully tracked. In this analysis, attrition is larger, although still relatively low compared to other national household surveys, due to the extensive tracking conducted by the TZNPS: in wave 1, individual-level attrition is 3.9% (as 60.1% of individuals leaving households are successfully re-surveyed), in wave 2, individual-level attrition is 4.6% (as 62.5% of individuals leaving households are successfully re-surveyed). Table 2.A1 presents data on individual-level attrition for the full sample of TZNPS households (excluding households that disintegrated or moved in their entirety).[9]

2.2.5. Shock data

In the second part of my analysis, I study how changes in the welfare of households affect the flow (and destination) of individuals leaving these households. To do so, I employ both self-reported data on shocks to household welfare (collected by the shock modules of the TZNPS) and meteorological data on climatic conditions during the agricultural growing season. I discuss these data sources in turn in this section.

The TZNPS collects data on shocks to welfare that households suffered in the years prior to the survey through "shock modules;" these modules have been

[9]Individual-level attrition is calculated by dividing the number of non-tracked individuals (the "leavers", who left the survey) by the total number of individuals in the wave, excluding deaths. The proportion of individuals leaving a household who are successfully tracked and re- surveyed is calculated by dividing the number of individuals who leave households and are successfully tracked and interviewed (the "movers") by the number of *all* individuals leaving a household, whether successfully tracked or not (i.e., the sum of the "leavers" and the "movers").

found to perform well in many settings (Heltberg et al., 2015). During the interview, households are presented with a list of nineteen events (with negative impact on household welfare, such as various price shocks, disasters, employment shocks, asset/crop losses, illness/death, or crime)[10] and asked whether they were "severely affected" by any (or several) of these events in the previous five years. For the three shocks that the household declared to be the "most severe", the year and month in which the shock was experienced are also collected. I restrict my analysis to these "most severe" shocks, both because of their economic impact and because I need to know their date, as I want to consider only shocks experienced since the previous wave.

In my analysis in section 2.4, I consider the impact of the four most common shocks that are not household-specific and plausibly exogenous: large rises in the price of food, droughts and floods, large falls in the sale prices for crops, and large rises in the prices of agricultural inputs (I follow Heltberg et al. (2015) in categorising shocks as specific or not specific to households). As I do not detect any impact of self-reported drought or flood shocks in my analysis, I also tested regression specifications that include severe water shortage shocks, but did not detect any impact of these either.[11]

Compared to objectively measured shocks, self-reported shocks have a peculiar feature: instead of measuring how severe a shock was (as objective data would), they measure how severe an impact a shock had on a household (as felt by the household). An increase in food prices of a given size, for example, might be reported as a shock by poorer households (who spend a larger share of their income on food), but not by richer households. Hence, even without recall bias, two households experiencing the same objective event might not both report it as a shock: self-reported shock data is thereby also "subjective" shock data, as households report shocks that they *felt* affected them. Since I am analysing the reaction of households to such shocks, I find this a desirable feature. However, this also makes it difficult to understand the impact that an objectively measured shock of a given size has on households. In a second part of the analysis, I there-

[10] See table 2.A2 for a full list of events.

[11] Similarly, I did not detect any impact when also including the sixth-most common shock in this category, crop disease or crop pests.

fore analyse the impact of objectively measured climate shocks on migration out of households.

A large share of households in Tanzania engage in agriculture; different location-specific climatic conditions during the agricultural season hence generate a natural source of variation in the agricultural production and welfare of households. I employ the Standardised Precipitation-Evapotranspiration Index (SPEI) to measure drought and excessive rainfall, climatic conditions relevant to agriculture. The SPEI provides a measure of climatic water balance (the difference between monthly precipitation and monthly evapotranspiration) and has been shown to measure weather/drought conditions for agriculture well and better than traditional indices (such as the Standardised Precipitation Index or the Palmer Drought Severity Index) or rainfall and temperature alone (Vicente-Serrano et al., 2010; Beguería et al., 2014). My choice of the SPEI is consistent with a number of other studies investigating the impact of drought (La Ferrara and Harari, 2018; Kubik and Maurel, 2016; Mueller et al., 2014).[12] Using the GPS location of each household, I extract data on local climatic conditions during the growing season from the SPEI dataset: for households in areas with a unimodal rainfall pattern, I use the 4-month SPEI from January to April (for all complete agricultural seasons since the previous interview), whereas for households in areas with a bimodal rainfall pattern, I use the 3-month SPEI from April to June as well as the 2-month SPEI from November to December (for all complete agricultural growing seasons). Figure 2.A1 in the appendix illustrates the geographic and temporal variation in climatic conditions captured by the SPEI (the left panel displays climatic conditions across Tanzania for the 2010 Msimu season, the right panel does so for the 2012 Msimu season).[13]

[12] Vicente-Serrano et al. provide geo-referenced SPEI time series which are calculated from the CRU TS 3.23 weather dataset. While the CRU TS dataset has a long historical coverage, it is less reliable for sub-Saharan Africa as it is based on scarce weather data and substantial interpolation. Following La Ferrara and Harari (2018), I therefore use the scripts provided by Beguería and Vicente-Serrano (2017) to calculate the SPEI based on the ERA-Interim Reanalysis dataset, which has a more accurate coverage for sub-Saharan Africa. I use monthly rainfall and evapotranspiration (as estimated by the Penman equation using monthly means of daily maximum temperature, daily minimum temperature, windspeed, bright sunshine hours, total cloud cover, dewpoint temperature, and atmospheric pressure, and altitudes) as inputs and use a Gaussian kernel as recommended.

[13] Tanzania has two rainfall regimes, making for different growing seasons across the country: northern, and north eastern, and eastern Tanzania (within a corridor of roughly 200km extending

2.3. How much sorting is there in migration?

I begin my analysis with a detailed examination of the movement of individuals in Tanzania. I first examine the incidence of individuals' movement and present data on how common it is for individuals to move and leave their previous household (to join an existing household or establish their own). I next investigate the geographic patterns of individuals' movement and the extent to which individuals move within and between rural and urban areas. Last, I examine whether there is any evidence for the sorting of individuals between rural and urban areas by education or other characteristics.

My analysis reveals a large amount of movement and migration in Tanzania: in any given year, one in four households in Tanzania experiences the departure or arrival of a member. While the majority of moving individuals stay within the rural/urban setting they previously lived in, substantial amounts of migration between settings also take place in both directions: one in six individuals leaving a rural household moves to an urban area; while about one in four individuals leaving an urban household moves to a rural area. When considering individuals leaving agricultural households, I find that education stands out as an important determinant of their choice of destination. I interpret this as evidence for sorting: individuals with seven or more years of education are three times as likely to move to an urban area as individuals with six or less years of education. Overall, individuals leaving agricultural households and moving to urban areas are substantially better educated than their peers moving to rural areas (or not moving at all); furthermore, they also tend to report higher wages when working (but are unemployed more commonly) than both their rural-moving peers and the general population. However, for individuals leaving non-agricultural households, I find no evidence of sorting.

---

south from Tanzania's northern border and extending further in the East to include Morogoro, Dar es Salaam and the lower-lying Pwani region) have a bimodal rainfall regime, with a long rainy season (Masika) during April, May, and June, and a short rainy season (Vuli) during November and December. Rainfall in the rest of the country (southern, south western, central, and western Tanzania) is unimodal, where the Msimu rains occur in January, February, March, and April (Gommes and Houssiau, 1982; Yohan et al., 2006; FAO, 2017).

Table 2.2: Flow of individuals (seen from wave of origin)

	2008-2010		2010-2012		Total
	#	%	#	%	%
stayed	14058	89.3	16778	86.7	87.8
left	1541	9.8	2367	12.2	11.1
died	152	1.0	209	1.1	1.0
Total	15751	100.0	19354	100.0	100.0

2.3.1. How commonly (and where) do Tanzanians move?

Incidence of movement. There is a large amount of movement and migration in Tanzania, as tables 2.2 and 2.3 show. During the two-year periods between subsequent survey waves, 10%-12% of all individuals left the household they previously lived in, translating into a 5%-6% departure rate per year (while 0.5% of Tanzanians died).[14] In-migration was similarly common; at the end of each two-year period, approximately 9%-11% of household members were new arrivals who had moved into the household since the previous wave (while 5% of household members were infants born since the previous wave). When examined at the household level, the common incidence of individual movement translates into a large number of affected households, as table 2.4 shows. Over each two-year period, approximately 50% of households experienced the departure or arrival of a household member (one quarter of these households experienced both). Experiencing changes in household composition hence is a common feature of life in Tanzania.

Direction of movement. I next examine *where* individuals move. The TZNPS tracks (with considerable success)[15] all individuals (13 years and older) who leave a household. This allows me to analyse the movement of individuals between rural and urban areas.[16] Table 2.5 presents these results: most individuals (about 80%), move within their setting: 60% of the moves take place within rural areas and 20%

[14] The demographic structure of Tanzania makes for a low mortality rate: in 2015, 72% of Tanzanians were below the age of 30.

[15] Section 2.2.4 provides a discussion of attrition.

[16] I follow the rural-urban classification of the TZNPS.

Table 2.3: Flow of individuals (seen from wave of arrival)

	2008-2010		2010-2012		Total
	#	%	#	%	%
stayed	14037	83.5	16754	86.3	85.0
arrived	1843	11.0	1644	8.5	9.6
born	923	5.5	1022	5.3	5.4
Total	16803	100.0	19420	100.0	100.0

Table 2.4: Households affected by migration

	2008-2010		2010-2012		Total
	#	%	#	%	%
Unaffected	1595	51.3	1820	49.9	50.5
Affected: Lost member	482	15.5	798	21.9	18.9
Affected: Received member	667	21.4	573	15.7	18.3
Affected: both	368	11.8	455	12.5	12.2
Total	3112	100.0	3646	100.0	100.0

Table 2.5: Direction of migration

	Destination		
Origin	urban	rural	Total
urban	20.5	8.1	28.6
rural	12.2	59.2	71.4
Total	32.7	67.3	100.0

within urban areas. However, among the 20% of individuals who change setting, I find - perhaps surprisingly - substantial movement in both directions: 12% of moves lead from rural to urban areas (accounting for 17% of the movement of individuals originally from rural areas), but 8% of moves also lead from urban to rural areas (accounting for 28% of the movement of individuals originally from urban areas). My findings hence align with those of Young (2015): rural-urban migration is a common occurrence, but so is urban-rural migration.

2.3.2. Who moves?

Tanzanians commonly move, and evidently this movement takes place in all directions. Taking advantage of the richness of my dataset, I now turn to investigating whether sorting (between rural and urban areas) based on individual skill or other characteristics might be a driver of such movement, as suggested by Lagakos and Waugh (2013) and Young (2013). To do so, I study the pool of migrants from rural areas and conduct three complimentary analyses (I discuss results for migrants from urban areas at the end of the section). First, I examine how the probability of leaving their previous home and moving to an urban/rural destination differs for different groups of individuals, defined by education, sex, and age; this enables me to understand how various characteristics affect the propensity to leave. I then study and compare who the individuals moving to urban and rural areas are: specific groups of individuals differ not only in their propensity of moving to an urban/rural area but also in their size (depending on how common this characteristic is in the general population). In this analysis, I hence consider how these factors interact to lead to substantial selection and thus bring about large differences between the individuals flowing to urban and to rural locations. Third, I examine a wide range of individual characteristics and compare individuals moving to urban areas to those staying behind or moving to rural areas; this allows me to understand the extent of selection along these characteristics.

In his exploration, Young (2013) emphasizes the role of individual "skill" (correlated with, but different from, education) in sorting.[17] However, he cannot observe this skill and instead infers it from education. I am similarly unable to observe individual skill in my data and hence focus on education (although I consider a wide range of individual characteristics beyond education potentially correlated with skill).

Below, I present my analysis in terms of agricultural and non-agricultural household to ensure consistency with the following section, in which I consider how shocks to the welfare of explicitly agricultural households affect the out-movement and destination of individuals. Empirically, however, this largely overlaps with the distinction between rural and urban households; I show in robustness checks in the appendix (tables 2.A11 and 2.A12, which correspond to tables 2.6 and 2.7) that results when categorising households as rural and urban (rather than agricultural and non-agricultural) are quantitatively similar and qualitatively the same.[18]

Who leaves agricultural households? I first examine *who* leaves agricultural households and consider the incidence of leaving by age, sex, and education groups; then I consider whether destination choices differ across education groups. Table 2.A3 presents my results. Rather than considering means, I have divided my sample into groups defined by age, sex, and education, and examine the incidence of leaving for each group separately. Unsurprisingly, these probabilities vary widely for individuals with different demographic characteristics. Only 1 in 25 individuals aged 35 and older leaves their household in a given two-year

[17]Young sees skill as a wider measure of human capital, of which education is only one component. Thinking of skill, rather than education alone, as the determinant of individuals' productivity and wages then also explains in Young's account why (on average more skilled) urban resident earn higher wages, even when controlling for education.

[18]Indeed, Lagakos and Waugh (2013) and Young (2013) also seem to think of them as largely interchangeable; Lagakos and Waugh conduct their analysis in terms of sorting between agricultural and non-agricultural sectors, while Young does so between rural and urban locations.

Table 2.6: Probability of leaving agricultural HHs (by education)

	Probability of leaving to		
	anywhere	urban	rural
Education (if 14+ yrs)			
7+ yrs edu	0.133	0.033	0.072
6- yrs edu	0.104	0.011	0.068
Education (if 14+ yrs)			
9+ yrs edu	0.164	0.060	0.062
7/8 yrs edu	0.124	0.025	0.075
5/6 yrs edu	0.146	0.019	0.089
4- yrs edu	0.092	0.009	0.062
Unconditional probability	0.119	0.023	0.070

Sample: agricultural households.
Results are similar for rural households (Table 2.A11).

period, while approximately one in three females aged 15-24 does so;[19] in comparison, only one in six males does so. Above 24, this gender gap narrows, as the probability of moving out falls for both females and males. Considering any of these dimensions separately, instead of in respective groups, yields comparable results.

In table 2.6, I examine differences in the probability of leaving a household and moving to urban or rural areas between more and less educated individuals.[20] Analysing the incidence (and destinations) by years of education, I arrive at my most interesting finding: more educated individuals are a staggering three times as likely (at 3.3% compared to 1.1%) to move to an urban area as individuals with less six or less years of education. I interpret this as convincing evidence of sorting between urban and rural areas based on education. Furthermore, when I define education groups more finely (as in the bottom half of the table), I find that the probability of moving to an urban area monotonically increases with in-

[19]Tanzania is a patrilineal society; upon getting married the bride generally moves to the family or village of the groom.

[20]Distinguishing between individuals with seven or more and individuals with six or less years of education is appealing, because seven years mark the completion of primary school; furthermore, this approximately divides my sample evenly.

dividuals' education (while no clear pattern is visible in individuals' probability of moving to a rural area). The large differences between more and less educated individuals appear to be driven by a very low probability (0.9%) of moving to an urban area for the sizeable group of individuals with four or less years of education, a higher probability (2.5%) for the substantial group of individuals with seven or eight years of education, and a comparatively large probability (6.0%) for the smaller group of individuals with nine or more years of education (who are virtually as likely to move to urban as to rural areas). I hence find compelling evidence of education-based sorting between rural and urban areas in the outmigration of individuals from agricultural households.

Flows to urban and rural areas. A complimentary approach to appreciate the extent of sorting is to examine and compare the groups of individuals that move to urban and to rural areas. Comparing the probability of moving to urban areas among differently educated individuals, I found that more educated individuals are more likely to *leave* agricultural households and move to urban areas. I now examine the cumulative impact of this selection and examine how the pool of individuals that moves to urban areas differs (in terms of education) from the pool of individuals that moves to rural areas. In other words, while before I looked at the probability of moving (urban or rural) given an individual's level of education, I now look at movers' levels of education, given their choice of destination.

Table 2.7 presents this analysis. Flows of individuals to urban and rural areas differ substantially in their composition: individuals with 7 or more years of education account for 77% of individuals moving to urban areas, but only for 55% of the individuals moving to rural areas. When I again examine education levels more finely, I find that this difference is driven by relatively highly educated (and relatively poorly educated) individuals: individuals with nine or more years of education account for 30% of the share of movers to urban areas, but only for 10% of the pool of movers to rural areas. On the other hand, individuals with four or less years of education account for 32% of movers to rural areas, but only for 14% of movers to urban areas. Education-based selection in the destination of outmigration hence results in substantially different flows of individuals to urban and to rural areas.

How do movers differ? Last, I examine how individuals moving to urban

Table 2.7: Composition of flows leaving agricultural HHs (by education)

	Share among movers to		
	anywhere	urban	rural
Education (if 14+ yrs)			
7+ yrs edu	0.596	0.774	0.551
6- yrs edu	0.404	0.226	0.449
Education (if 14+ yrs)			
9+ yrs edu	0.156	0.303	0.100
7/8 yrs edu	0.440	0.471	0.451
5/6 yrs edu	0.120	0.084	0.125
4- yrs edu	0.284	0.141	0.324
# of observations	1569	297	922

Sample: agricultural households.
Results are similar for rural households (Table 2.A12).

and to rural areas differ on a wide range of individual characteristics: Table 2.8 presents this analysis. Using data from the survey wave before individuals moved (when they were still part of and surveyed in their previous agricultural household), I compare movers to urban and rural areas.[21]

In line with my previous results, I find substantial differences between urban and rural movers, not only in terms of education, but also in their work and employment histories, life satisfaction, demographic background, and anthropometric outcomes. Although better educated (with two additional years of education, on average) and healthier, urban movers were 25% less likely to be employed (before moving, while residents of their agricultural households) and reported significantly lower satisfaction with their job and financial situation. However, those employed were twice as likely to earn a wage and this wage was substantially higher than that of their peers moving to rural areas. Overall, urban movers hence appear to have higher human capital (as they are better educated and earn more when employed) but seem to lack employment opportunities commensurate with

[21]Given the timing of the surveys, this captures the situation of individuals 0-24 months prior to their move (as the survey rounds are two years apart and individuals left the households sometime between two survey rounds).

Table 2.8: Movers to urban vs. movers to rural areas

	rural	urban	difference (se)	# to r	# to u
Demographic Characteristics					
age	25.4	23.9	-1.48 (0.70)*	1001	338
male	0.36	0.42	0.057 (0.031)	1001	338
father alive	0.71	0.71	-0.0049 (0.029)	1000	338
father in HH	0.34	0.38	0.031 (0.030)	1000	338
mother alive	0.82	0.84	0.018 (0.023)	1001	338
mother in HH	0.42	0.53	0.11 (0.031)***	1001	338
currently married	0.24	0.11	-0.13 (0.022)***	1000	338
spouse or partner lives in HH	0.31	0.11	-0.20 (0.022)***	1000	338
# months absent in past year	1.07	1.51	0.44 (0.18)*	1001	338
Education					
years of education	5.36	7.35	1.99 (0.20)***	998	338
ever been to school	0.79	0.93	0.13 (0.019)***	1001	338
completed year 4	0.74	0.91	0.17 (0.021)***	998	338
completed year 7	0.57	0.77	0.21 (0.028)***	998	338
completed year 9	0.11	0.31	0.21 (0.027)***	998	338
completed year 11	0.034	0.15	0.12 (0.020)***	998	338
completed more than year 11	0.0090	0.047	0.038 (0.012)**	998	338
currently in school	0.15	0.22	0.064 (0.025)*	1001	338
education spending (if curr in school)	119.8	227.3	107.5 (62.6)	155	74
Health					
visited healthcare prov in past month	0.095	0.13	0.035 (0.021)	1000	338
hospitalised in past year	0.047	0.027	-0.020 (0.011)	1001	338
physically handicapped (w1 only)	0.024	0.022	-0.0013 (0.015)	338	134
slept under bednet last night	0.48	0.54	0.058 (0.031)	1001	338
gave birth in past 2 years (if f/12-49)	0.26	0.19	-0.067 (0.034)	593	183
Work and Employment					
worked in last 7 days	0.55	0.40	-0.16 (0.031)***	995	336
worked in last 7 days or will resume	0.78	0.58	-0.20 (0.030)***	995	335
unemployed: could but did not work	0.020	0.072	0.052 (0.015)***	995	335
earned wage in last 7 days (if w/ job)	0.13	0.23	0.094 (0.033)**	775	193
monthly wage (if earned wage)	211.0	743.8	532.7 (549.7)	102	43
- winsorized at top 1%	211.0	428.4	217.3 (248.3)	102	43
hrs unpaid non-ag HH work (prev wk)	6.99	7.11	0.12 (0.87)	989	333
hrs unpaid ag HH work (prev wk)	14.0	7.23	-6.82 (0.97)***	996	335
Life Satisfaction: 1 (highest) - 7 (lowest)					
life sat: health	2.36	2.08	-0.28 (0.12)*	731	229
life sat: financial situation	4.26	4.58	0.32 (0.16)*	718	223
life sat: housing	3.14	3.10	-0.043 (0.15)	725	229
life sat: husband/wife (w1 only)	1.63	1.81	0.18 (0.25)	91	21
life sat: job	3.10	3.66	0.56 (0.18)**	611	164
life sat: overall	3.85	3.93	0.080 (0.16)	723	228
Anthropometric outcomes					
BMI for age (z-score)	-0.20	-0.052	0.15 (0.076)*	769	240
height for age (z-score)	-1.28	-1.49	-0.21 (0.077)**	769	239
BMI for age z-score <-2	0.040	0.021	-0.019 (0.012)	769	240
height for age z-score <-2	0.24	0.31	0.069 (0.034)*	769	239

Sample: agricultural households.

their skill at their original locations in agricultural households, as their rate of joblessness is substantially higher. I discuss these differences in turn.

- **Minor differences in terms of demographic characteristics:** Movers to urban areas are on average 1.5 years younger than movers to rural areas, but similar in gender composition: for both destinations, around 4 in 10 are male. However, those moving to urban areas are less likely to be married or living together with their spouse; on average they have also been absent for slightly longer periods in the previous year.

- **Major differences in terms of education:** Notably and in line with previous results, those moving to urban areas are substantially better educated than their peers who move to rural areas: on average, urban movers have 7.4 years of schooling, while rural movers have 5.4. Looking beyond average years of education, urban movers are 14 percentage-points more likely to have completed year 4, and 20 percentage-points more likely to have completed year 7 and year 9 (making them three times as likely to have completed year 9 as rural movers). At 5%, they are also five times as likely as rural movers to have more than 11 years of education.

- **Major differences work and employment histories:** Regarding work and employment, urban movers differ from rural movers in three ways. First, fewer urban movers had a job they worked at in the past 7 days or would return to in the future: while 55% of rural movers did so, this was only 40% for urban movers. However, when working, urban movers were significantly more likely to earn a wage (at 23%, relative to 13% for rural movers), and, on average, earned twice as much as rural movers (although this difference is not statistically significant). Last, urban movers spent substantially (and significantly) fewer hours doing unpaid agricultural work for the household than their peers moving to rural areas.

- **Differences in terms of subjective well-being and anthropometric outcomes:** Urban movers reported slightly higher subjective satisfaction with their health (which corresponds to objective measures, a lower hospitalisation rate in the previous year and a higher utilisation of bed nets), but ex-

> pressed less satisfaction with their financial situation and their job than rural movers. Urban movers also report slightly higher body weights, although they were slightly shorter on average.[22]

Urban movers hence differ significantly from their rural-moving peers: better educated and earning more when working, they nevertheless appear less satisfied with their lives in agricultural households, reporting, alongside lower satisfaction with their job and financial situation, not having a job twice as often as rural movers (only 22% of rural movers report not having worked in the previous 7 days or having a job they will return to, whereas 42% of urban movers do so).

Movers to urban areas also differ significantly from the general population (whereas this is less obvious for movers to rural areas). Table 2.A7 presents this analysis:[23] urban movers are younger, better educated, and healthier (which is perhaps unsurprising given their younger age). Importantly, however, urban movers differ from the general population in the same way they differ from rural movers: despite their better education, they are both more likely to be unemployed at their agricultural households and, when working, earn higher wages. Compared to the characteristics of the general population, urban movers are hence a specifically selected sample. This is also notable as, when similarly comparing *rural* movers to the general population (tables 2.A9 and 2.A10), few such differences between rural movers and the general population in terms of education, work, and employment exist aside from being younger and somewhat healthier. In short, individuals moving to urban areas hence appear to be a particularly selected sample when examining a range of individual characteristics, whereas individuals moving to rural areas are not.

No education-based sorting when leaving non-agricultural households. I now turn to examine the movement of individuals out of *non*-agricultural households to investigate whether education-based sorting also takes place in this direction, thus leading less educated individuals to leave these households and move to rural areas. Table 2.A5 presents this analysis and shows that this is not the case:

[22]However, as only household members personally present at the time of the survey were measured and asked about their subjective life satisfaction, the surveyed sample might be selected.

[23]These differences are robust and also present when I restrict my sample to only households who have an out-mover, as table 2.A8 shows.

among urban residents, having less education does not appear to increase the probability of moving to a rural area. Consequently, the flow of individuals moving from urban to rural areas does not appear to be selected in terms of education, as table 2.A6 shows. These results need not imply that there is no selection by skill in urban-to-rural migration, however: if the low skill that leads urban individuals to move to rural areas is not well captured by years of education, then I am not able to detect urban-to-rural sorting by low skill.[24] Robustness tests reported in tables 2.A13 and 2.A14 confirm that these results hold whether I consider urban or non-agricultural households.

2.3.3. Discussion

Overall, I find that education has a large impact on the destination of out-migration for individuals leaving agricultural households. More educated individuals are more likely to move to urban areas and the magnitude of this impact is substantial: individuals with seven or more years of education are three times as likely to move to an urban area as individuals with six or fewer years of education.[25] These education-based differences in individuals' rates of movement to urban and rural areas in turn cause the pool of urban migrants to be substantially better educated: 55% of movers to rural areas but 77% of movers to urban areas have at least 7 years of education; on average, urban movers have obtained two more years of education than rural movers. Comparing urban and rural movers on other characteristics, further differences become obvious: likely due to their better education, urban movers both earn more when working (pre-move at their agricultural locations) but are also more commonly without a job, possibly as there is a lack of employment opportunities commensurate with their education at agricultural locations. Compared to the general population, urban movers also appear to be a particularly selected group, as they are better educated, better-paid when working (but also more likely to be unemployed), and healthier. Movers to rural

[24]This might be the case, for example, if schools teach well to students of higher skill but fail to teach low-skill students effectively (who nevertheless stay in school).

[25]The effect is even larger when comparing individuals with particularly high or low amounts of education (nine or more years and four or fewer years), whose probability of moving to an urban area differs by a factor of six, while their probability of moving to a rural area is similar.

areas, on the other hand, do not appear to be a particularly selected group, barely differing (except for their younger age) from the general population.

2.4. Sorting and the sending households

I now investigate how stable migration is in response to shocks to households' welfare. In the previous section, I documented that a substantial amount of migration exists in Tanzania and that more educated individuals sort from agricultural households to urban areas. Now, I examine the circumstances under which this migration and rural-urban sorting occur and investigate how changes in the economic conditions of households affect migration: are negative shocks to the welfare of agricultural households push factors leading individuals to out-migrate and, if so, do they affect the rural-urban sorting of individuals by education?

The Roy model provides useful guidance and intuition for such thinking about migration (Lagakos and Waugh, 2013; Dustmann and Glitz, 2011; Borjas, 1987). In its essence, there are two locations and individuals who differ in their skill endowments in the model. As skills are rewarded differently between the two locations, the same individual will be paid differently in the two locations. Migration is costly, however, and an individual only moves if their net expected earnings at the destination (or expected utility, depending on the formulation of the model) are greater than earnings (or utility) at their current location.

Negative shocks to the agricultural production and welfare of households can hence in principle have two effects: firstly, they affect the relative return of migration: as the status quo at the agricultural household worsens, the difference in earnings between urban and agricultural work, the relative return to migration, increases (assuming that the direct effects of shocks on the household are larger than any potential indirect effects on urban consumers). Secondly, however, negative shocks may also affect households' ability to finance the cost of migration (if households are credit-constrained) or their ability to bear the financial risk of failed migration (if households are sufficiently poor).[26]

In my analysis, I find suggestive evidence of both effects: overall, the sorting

[26] Bryan et al. (2014) demonstrate this in an experimental study in Bangladesh.

of more educated individuals to urban areas is highly sensitive to the economic conditions of sending agricultural households as shocks have a large impact on the movement of individuals. Shocks that plausibly have a direct negative effect on households' ability to finance out-migration cause a large reduction in the out-migration of educated individuals to urban areas, whereas shocks that reduce potential returns to agriculture, but do not immediately affect households' cashflows do not and cause large increases instead. More specifically, I find that educated urban migration falls in response to falls in the sale price of harvested crops but increases in response to rises in the price of agricultural inputs.[27] These two findings suggest that more educated individuals leave agricultural households (and agriculture) and move to urban areas when the economic returns to agriculture decrease, but that their migration might often be difficult to finance, as tighter household budget constraints (induced by lower sale prices for harvested crops) reduce such migration.

I also consider the impact of shocks on the movement of individuals from and to households more generally: here, I find, in line with my previous results, that rises in agricultural input prices also reduce migration into agricultural households. Objectively measured adverse weather shocks, droughts and large amounts of rainfall, similarly lead to a large reduction in migration into the household, without a significant impact on migration out of the affected household. These findings provide further evidence for the potential importance of household liquidity: worse household circumstances reduce the flow of individuals into households but plausibly do not lead to greater outflows if these are difficult to finance. Moreover, I find that food price rises also reduce out-migration from households; interestingly, however, they also increase in-migration. I discuss and present my analysis in the subsections that follow.

[27]Plausibly, an (unexpected) fall in the sale prices of harvested crops reduces household earnings at the time of harvest and leaves little scope for households to adapt their economic activities. A rise in the price of agricultural inputs, on the other hand, does allow for the re-optimisation of economic activity: rather than purchasing inputs at a higher price, households can invest in alternative strategies, such as financing the out-migration of a member.

2.4.1. Econometric framework

My econometric approach is intuitive and builds on my previous analysis: by comparing household rosters between adjacent survey waves, I detect flows of (differently educated) individuals into and out of households (and to rural or urban destinations) and code corresponding dummy variables; these are my outcome variables. I use the shock module of surveys to code for each shock an indicator variable equal to 1 if the household experienced such a shock. When I examine the impact of objectively measured weather shocks, I proceed similarly, but use shock measures derived from the SPEI during the household's previous agricultural season. I then regress my various binary outcomes on these shock dummies in a logistic regression model, controlling for household characteristics (measured at the time of the earlier wave).

The regression model I estimate is

$$p_i \equiv \Pr[y_i = 1|\mathbf{x}_i, \mathbf{c}_i] = \frac{e^{\mathbf{x}_i^\intercal \beta_1 + \mathbf{c}_i^\intercal \beta_2}}{1 + e^{\mathbf{x}_i^\intercal \beta_1 + \mathbf{c}_i^\intercal \beta_2}}$$

where y_i is the outcome dummy, $\mathbf{x}_i$ is a vector of variables, indicating whether the household reported a specific shock, and $\mathbf{c}_i$ is a vector of control variables.[28] In my estimation, I pool both periods (i.e. between waves 1 and 2 and between waves 2 and 3) and cluster standard errors at the level of the enumeration area.

To understand how shocks to household welfare affect the rural-urban sorting of individuals by education, I consider how these shocks affect the probability that the household experiences the out-migration of a member with a given amount of education to a given location. This approach leads me to run, for each group of individuals with a specific amount of education, one regression for each outcome,

[28] The control variables I include in regressions are the number of household members; the mean age of household members; the mean years of education of household members; the mean years of education of household members aged 15-65; the years of education of the most educated household member; a dummy if the household head is female; the amount of agricultural land available to the household (in acres and by rank); the monthly consumption of the household (in real terms, geo-temporally adjusted, and per adult-equivalent household member in Tanzanian Shillings); a dummy if the household has good flooring (concrete, cement, tiles, or timber), a dummy if agriculture is the household's principal source of income; and three measures of remoteness (a dummy whether an elementary school exists in the community of the household, a dummy whether a secondary school exists in the community of the household, and the cost of travel to the district capital in Tanzanian Shillings).

namely out-migration to an urban area, and out-migration to a rural area. As such migration is only posible if the household has a member with that amount of education in the first place, I restrict my sample to households that had such a member at the start of the period. To examine the *differential* impact of a shock on migration to urban versus rural areas, I need to compare the estimated coefficients on the impact of a shock on outmigration to urban and to rural areas. A straightforward way for doing so, testing for the equality of estimated coefficients, is to combine the two regression models into one by "stacking" them and to then perform a Wald test. I do so and report the p-value of a Wald test for the equality of individual coefficients between each pair of groups.

When I examine more generally how shocks to household welfare affect the flows of individuals out of and into households, I estimate the same regression models, but now consider as outcomes whether the household experienced the outflow of a member to any destination and whether the household experienced the inflow of a member.

2.4.2. Shocks and urban-rural sorting

In my analysis in section 2.3, I found a large amount of sorting based on education. I now investigate the conditions under which this takes place. Specifically, I ask how sensitive this sorting, the selective out-migration to urban or rural areas by education, is to changes in the welfare of households (induced by shocks). To do so, I employ the regression model discussed in section 2.4.1.

Table 2.9 presents my results; I find that falls in the sale prices of harvested crops and rises in the prices of agricultural inputs substantially impact the out-migration of individuals from households. Each column in the table presents my findings on the impact of shocks on outmigration for a different subgroup and destination, making for four columns, since I consider two subgroups (individuals with at least seven years of education and individuals without) and two destinations (urban and rural areas). Columns 1 and 2 present the results of logit regressions of shock dummies (and controls) on a dummy variable equal to 1 if an individual *with at least seven years of education* leaves the household and moves to an *urban* area (column 1) or a *rural* area (column 2). Columns 3 and 4

Table 2.9: Marginal effects: shocks and sorting (by education)

	edu 7+ urb	edu 7+ rur	edu 6- urb	edu 6- rur
1: drought/flood	-0.011	0.013	0.004	-0.002
	(-0.99)	(0.80)	(0.65)	(-0.17)
1: crop price fall	-0.033	0.014	-0.003	-0.028
	(-2.82)***	(0.79)	(-0.34)	(-2.09)**
1: food price rise	0.007	0.007	-0.003	-0.024
	(0.72)	(0.46)	(-0.60)	(-2.10)**
1: ag in price rise	0.045	0.007	0.001	0.006
	(2.50)**	(0.34)	(0.16)	(0.29)
drought/flood: p(equal)		0.238		0.508
crop price fall: p(equal)		0.0220		0.702
food price rise: p(equal)		0.759		0.755
ag in price rise: p(equal)		0.0426		0.988
chi-squared	234.4	212.0	106.4	300.3
# of observations	3112	3112	3386	3386
% HHs with leaver	6.304	13.09	1.823	10.63

present the results of similar logit regressions for individuals with six or less years of education. The bottom row in the table ("% HHs with leaver") presents the unconditional probabilities of moving for both education groups and both destinations, which are in line with my previous individual-level analysis of outmigration (Table 2.6): since individuals with six or less years of education rarely move to urban areas, the lack of any significant effects in column 3 is unsurprising.

As previewed above, falls in the sale prices of harvested crops and rises in the prices of agricultural inputs have very large impacts on the sorting of out-moving individuals between urban and rural areas: falls in the sale prices of crops reduce out-migration from households both for more and less educated individuals, and do so particularly for migration to urban areas. Increases in the prices of agricultural inputs, on the other hand, significantly increase the out-migration of households' more educated individuals to urban areas but have little effect otherwise. I discuss both in turn below.

My most noteworthy finding on crop price falls is their large impact on the movement of (more) educated individuals to urban areas: households that report a fall in crop prices are 50% (or 3.3 percentage points, relative to an unconditional probability of 6.3%) less likely to have a member with seven or more years of education leave and move to an urban area. In other words, only one out of two households who would have a member with completed primary school education leave to an urban area does so when experiencing a shock. This a very large impact. On the other hand, there is no significant impact of such shocks on the probability that individuals with such education move to rural areas. Small positive but insignificant point estimates suggest that substitution effects towards rural areas in the destination choice of movers are small at most. Falls in crop prices also have a large negative impact on out-migration for less educated individuals; these become 2.8 ppts, a 25% reduction, less likely to move to a rural area. Since less educated individuals very rarely move to urban areas in the first place, the insignificant 0.3 ppts reduction (15% relative to an unconditional probability of 1.8%) is unsurprising. Together these results suggest that, although falls in the sale price of harvested crops reduce out-migration for both more and less educated individuals, their impact on urban-rural sorting is substantial.

I next examine increases in the prices of agricultural inputs that households report as shocks and find a similarly significant impact here. Increases in the prices of agricultural inputs cause a large increase in the movement of more educated individuals to urban areas. This impact stands out because of its size: households that report a shock are 4.5 ppts more likely to have a more educated individual move to an urban area. Relative to an unconditional probability of 6.3%, this is a 70% increase. On the other hand, input price increases have no impact on the migration of more educated individuals to rural areas or the migration of less educated individuals; for these groups, I estimate relatively precise zeros. Rises in the prices of agricultural inputs thereby have a drastic impact on urban-rural sorting: they appear to be strong "push" factors on more educated individuals, causing them to leave agricultural households and move to urban areas.

Last, I examine the impact of increases in food prices that households report as shocks. Individuals with seven or more years of education appear to be unaffected by rises in food prices, as I estimate relatively precise zeros for their impact on the out-movement of more educated individuals. Interestingly, the out-movement of less educated individuals, on the other hand, is affected by increases in the price of food to a substantial extent: households reporting a shock are 2.4 ppts less likely to have an individual with six or less years of education move to a rural area, which is a 20% decrease relative to the unconditional probability of 10.6%. For less educated individuals, food price increases thereby have impacts similar to crop price falls, substantially decreasing their migration to rural areas. Conceptually, this seems intuitive as both shocks tighten households' budget constraints; crop price falls by reducing income from crops and food price increases by increasing expenditure on food, a main consumption item.

I also examine the impact of objectively measured rainfall shocks on the out-migration and sorting of individuals. To do so, I use the SPEI as an objective measure of drought/flood shocks, which I calibrate (as discussed in the next section) to match the incidence of objective drought/flood shocks to the incidence of self-reported drought/flood shocks. As with self-reported measures, I find that there is also no (significant) impact of objectively measured drought/flood shocks on the out-movement of individuals from the household (table 2.A15).

My take-away from this analysis is hence threefold.

First, I find that the out-migration of (more) educated individuals from agricultural households to urban areas is highly sensitive to the local conditions of sending households. Falls in the sale prices of harvested crops and increases in the prices of agricultural inputs both have very large impacts on the incidence of out-migration to urban areas, decreasing it by 50% or increasing it by 70% respectively. This suggests that the economic situation of households has a large influence on the outflow of (more) educated individuals to urban areas.

Second, I regard my results as plausible evidence for (or to be at least highly compatible with) the existence of financial/liquidity constraints for households that create barriers to the out-migration of household members. The differential impacts of these shocks, particularly agricultural input price increases and crop sale price falls, suggests that out-migration is costly to finance and that households are less able or willing to do so when budget constraints are tighter. Falls in the sale prices of crops, which directly reduce households' earnings and liquidity lead to a large decrease in out-migration. Increases in the prices of agricultural inputs, on the other hand, have a different timing and not necessarily a direct negative effect on households' cashflows: rather than to purchase these costlier inputs when informed of their higher prices, households can engage in different economic strategies and substitute inputs, reduce their agricultural operations and switch to other sectors, or finance the migration of household members elsewhere. I interpret the shocks' large impact on the out-migration of more educated members to urban areas as suggestive evidence that exactly this might be the case; facing higher costs of agricultural production (relying on inputs), households appear to send out their more educated members to urban areas instead.

Thirdly, when subject to adverse shocks to their circumstances, households appear to preferably send out their more educated members: all three shocks that have a significant impact on out-migration from the household reduce the out-migration of less educated individuals more than that of more educated ones. This suggests that households are selective in sending out members of households in response to shocks, preferring relatively more educated members who are likely to earn more.

2.4.3. Shocks and Migration

Having considered the impact of shocks on urban-rural sorting and out-migration by education and destination, I turn to also examine how they affect the overall flows of individuals out of and into households. In this section, I estimate the impact of exogenous shocks to the agricultural production and living conditions of agricultural households on the general movement of individuals out of and into households. Table 2.10 presents these results (as marginal effects). As discussed in section 2.4.1, I estimate similar regression models as in the previous section, but now use dummy variables indicating whether the household experienced the outflow of any household member to any location or the inflow of an individual into the household, respectively. I consider the movement of individuals *out of* households in the two columns on the left; the two columns on the right consider the movement *into* households. Results are robust to the omission of a full set of control variables, as shown in columns 2 and 4, although precision decreases.

Examining the impact of shocks on out-migration, I find that falls in crop prices also have a significant impact on out-migration more generally: while the

Table 2.10: Marginal effects: shocks and migration (out and in)

	out-migr.	out-migr.	in-migr.	in-migr.
1: drought/flood	0.014	-0.010	0.025	0.006
	(0.74)	(-0.57)	(1.19)	(0.27)
1: crop price fall	-0.033	-0.050	0.052	0.040
	(-1.35)	(-2.07)**	(2.00)**	(1.51)
1: food price rise	-0.021	-0.033	0.056	0.042
	(-1.35)	(-2.08)**	(3.24)***	(2.41)**
1: ag in price rise	0.018	0.019	-0.053	-0.046
	(0.67)	(0.74)	(-2.18)**	(-1.81)*
controls	No	Yes	No	Yes
chi-squared	4.403	292.1	21.82	189.9
# of observations	4307	4100	4307	4100
% HHs with leaver/arriver	28.33	28.33	29.07	29.07

unconditional probability of experiencing the out-migration of a household member is 28% for agricultural households, marginal effects estimates suggest that this is 5 percentage-points lower for households who experience a fall in crop prices (translating into an 18% decrease). For increases in the prices of agricultural inputs, point estimates are positive but not significant, suggesting that their principal impact is on the sorting of better-educated individuals to urban areas, rather than out-migration more widely (which also includes the out-movement of women who get married, for example). Notably, food price rises also appear to affect out-migration, reducing it by 3 percentage-points (a 12% fall). As before, I do not find an impact of droughts and floods on out-migration.

Considering in-migration, I find a large negative impact of increases in the prices of agricultural inputs, significant at the 10%-level: the probability of receiving an in-migrant is 5 percentage-points lower in affected households, a 15% decrease relative to an unconditional probability of 29%. There is no significant impact of crop price falls on in-migration. However, food price rises again have a notable impact: households reporting such increases are 4 percentage-points more likely to experience the inflow of a new member. In line with previous results, droughts and floods do not appear to impact in-migration.

As these results show, shocks to households' welfare and agricultural production do not only impact urban-rural sorting, but also the in- and out-movement of individuals more generally. Importantly, these impacts appear to be consistent with the existence of financial constraints that make it difficult for households to finance out-migration in adverse circumstances: falls in earnings from the sale of crops, tightening households' budget constraints, reduce overall out-migration without a significant impact on in-migration. Increases in the prices of agricultural inputs, that do not immediately tighten financial constraints, on the other hand, do not cause such decreases in out-migration; the lack of a general effect on out-migration beyond the surge in the migration of more educated individuals to urban areas suggests that households primarily send their more educated members out. However, such rises in the prices of agricultural input prices substantially reduce in-migration, which is plausible if households' economic prospects worsen. The impact of a rise in food prices on households appears to be surprisingly symmetric: food prices increases both reduce out-migration and increase in-migration. Given the way I construct my measures, this is not a mechanical effect. Instead, one reading of this result might be that households adjust their size in response to reduced real incomes and "crowd together." However, the existence of returns to scale in household size remains disputed, with puzzling empirical evidence in fact suggesting decreasing returns to household size (Deaton and Paxson, 1998) and an unresolved debate around this puzzle (Gan and Vernon, 2003; Deaton and Paxson, 2003). Heterogenous impacts of food price increases, depending on whether households are net producers or net consumers of food, obscured by my examination of averages might be an alternative explanation. However, since my primary interest is in urban-rural sorting, I do not investigate this issue further.

Somewhat surprisingly, I do not to find any evidence for the impact of self-reported weather shocks, specifically droughts and floods, on the movement of individuals out of and into households. Given that agriculture is overwhelmingly rainfed in Tanzania (irrigation is used by only three percent of farming households), I expected a substantial impact of adverse climatic conditions on migration, as Kubik and Maurel (2016) found. Fortunately, meteorological observation and climate modelling provides me with another source of data (discussed in section 2.2.5) on local climatic conditions, which I can use to verify the results obtained using self-reported data; thereby I can also cast light on the reliability of self-reported shock data.

I hence repeat my analysis of the impact of shocks on flows out of and into households using the standardised precipitation-evapotranspiration index (see section 2.2.5) as my measure of drought and rainfall. The SPEI has been demonstrated to measure droughts well (Vicente-Serrano et al., 2010; Beguería et al., 2014) and can be easily calibrated to detect adverse conditions of different intensities, as it is a normally distributed index (with mean zero and unit standard deviation within each cell across time). This allows me to easily consider different definitions of "dry" conditions (for example, defining dry conditions as a SPEI value <0 would, for each cell, classify all seasons drier than the mean season in that cell as "dry").

I first calibrate the SPEI-based measure of drought so that the number of households that experience a drought shock according to the SPEI is approximately equal to the number of households that self-report a rainfall shock. This is the case if I define SPEI values <-1.1 as drought; by this definition, 562 households experience a drought, which is close to the 582 households that self-reported having experienced a drought. Using this measure of drought, I estimate the impact of drought shocks on migration out of and into households; table 2.A18 in the appendix presents my results. As in my analysis using self-reported measures, I do not detect a significant impact of drought on in- or out-migration from households either, which validates my results relying on self-reported shocks.

I then consider the impact of more serious drought conditions. In my previous analysis, I defined a drought as an SPEI <-1.1; by this threshold, drought conditions were relatively common, experienced by approximately one in eight households. However, it is easy to imagine that the set of adverse climatic conditions that affect migration out of and into the household is narrower, restricted to only the *worst* conditions. I hence adapt my definition of drought (and flood), and restrict it to an SPEI of <-1.5 (and >1.5).[29] Tables 2.A19 and 2.A20 present my results for both specifications (in columns 1 and 2). Focussing my attention

[29]The SPEI happens to be distributed somewhat asymmetrically; only 54 households in my sample experienced an SPEI >1.5 in their last growing season, while 202 households experienced an SPEI <-1.5 in their last growing season. Practically, there is hence relatively little difference in whether I consider only very dry conditions (droughts) or combine them with very wet conditions.

Table 2.11: Marginal effects: droughts and migration (out and in)

	out-migr.	out-migr.	in-migr.	in-migr.
1: abs(SPEI) >1.5	-0.047	-0.049	-0.099	-0.106
	(-1.18)	(-1.21)	(-3.18)***	(-3.34)***
controls	No	Yes	No	Yes
chi-squared	1.378	265.3	10.19	131.0
# of observations	4304	4097	4304	4097
% HHs with leaver/arriver	28.30	28.30	29.07	29.07

on these more serious adverse conditions, I detect a substantial impact on the movement of individuals *into* households, while point estimates for *out*-migration are large, but not significant: households that experienced adverse climatic conditions are 10 percentage-points less likely to receive a new member into their household. Given that the unconditional probability that a household receives a new member is 29%, this is a sizeable decrease of 34%. Out-migration might also be affected by climate shocks during the growing season; point estimates are substantial (around 5%), but insignificant. Alternative specifications of climate shocks (derived from the SPEI) yield comparable results, suggesting that my findings are not driven by a particular specification.[30]

Overall, adverse weather conditions during the agricultural season hence translate into large, 30% reductions in inflows into affected households, but do not generate outflows from affected households which might be expected in response to worse conditions. These impacts of weather shocks are also consistent with

[30]For in-migration, my results are similar whether I consider only drought conditions (SPEI <-1.5) or drought *and* wet conditions (SPEI <-1.5 and SPEI >1.5) as shock (columns 1 and 2). They are also quantitatively the same (although power decreases) when I include the same measure of drought for the previous growing season (columns 3 and 4). I also obtain quantitatively similar results (although again without significance) when I use the proportion of months in the previous growing season (column 5) or in all growing seasons since the previous survey wave (column 6) in which the SPEI was below <-1 as my measure of shock. Lastly, I also consider the impact of less extraordinary and severe conditions on migration into the household (columns 7 and 8), considering SPEI values <-1 and >1 as shocks. Reassuringly, when defining my shocks this loosely, my estimates of their impact shrink towards zero. For out-migration, my results are also similar across specifications: here, I obtain insignificant and small results for the various shock definitions.

explanations emphasizing the importance of households' ability to finance out-migration, which is plausibly reduced by adverse weather shocks. Considering weather shocks of different intensity, I furthermore reconcile the different impacts of self-reported and objectively measured weather shocks: when matching the prevalence of objectively measured weather shocks to that of self-reported weather ones, I do not find a substantial impact of objectively measured weather shocks either, suggesting that only the more severe of self-reported weather shocks impact the movement of individuals.

2.5. Discussion and summary

In developing countries, urban residents tend to lead better lives than their rural peers, enjoying higher productivity, earnings, and consumption. However, the source of these differences, or urban-rural gap, remains disputed. Recent research has suggested that urban-rural differences in individuals' skill, caused or amplified by the selective migration, or sorting, of more skilled individuals to urban areas might cause the urban-rural gap. In this case, migration from rural to urban areas would be selective and individuals moving to urban areas should be more skilled.

In this paper I study domestic migration flows to examine whether this is indeed the case. To do so, I re-purpose a particularly rich and nationally representative household panel survey from Tanzania to investigate (1) the extent to which there is rural-urban sorting by (observable) skill, (2) the extent to which migrants to urban and rural areas differ on a wide range of individual-level characteristics, and (3) the extent to which changes in the welfare of agricultural households affect rural-urban sorting and out-migration.

I document a considerable extent of selection in migration out of agricultural households: educated individuals (with seven or more years of education) are three times as likely to leave their household and move to an urban area as their peers with less education (at 3.3% vs. 1.1%). This leads migration flows from agricultural households to urban areas to be heavily selected and substantially more skilled: 30% of such movers have nine or more years of education (compared to 11% in the general population) and 47% have seven or eight years of education (compared to 40%), while only 14% have four or fewer years of education (compared to 37%).

Examining a wide range of covariates, I also find other large differences between individuals leaving agricultural households and moving to urban areas and their peers who either stay behind or move to rural areas: beyond having on average two more years of education, urban movers are prior to moving both twice as likely to work for salary and to report being unemployed, which suggests higher participation in formal labour markets and a potential lack of commensurate employment opportunities for these more educated individuals at their agricultural households. Furthermore, subsequent urban movers report better health but lower satisfaction with their work and financial situation.

Last, I investigate how sensitive this sorting is to changes in the economic situation of sending agricultural households. Examining the impact of commonly reported and plausibly exogenous shocks to the welfare and agricultural production of households, I find that changes in households' economic conditions have a substantial impact on the sorting of their more educated members to urban areas: income-reducing falls in the sale prices of harvested crops more than halve out-migration rates of more educated individuals to urban areas for affected households, whereas increases in the prices of agricultural inputs, which plausibly decrease the potential returns to future agricultural production, increase the incidence of such migration. I interpret my findings to suggest that sending households' economic conditions have a significant impact on the out-migration of individuals to urban areas; being able to finance the migration of their household members to urban areas might be an important pre-condition and bottleneck to the sorting of more educated individuals to urban areas.

2.6. Appendix

Table 2.A1: Flow of individuals (seen from wave of origin)

	2008-2010		2010-2012		Total	
	#	%	#	%	#	%
stayed	14058	89.3	16778	86.7	30836	87.8
resurveyed (13 or older)	734	4.7	1207	6.2	1941	5.5
resurveyed (under 13)	204	1.3	272	1.4	476	1.4
attrited (13 or older)	296	1.9	414	2.1	710	2.0
attrited (under 13)	307	1.9	474	2.4	781	2.2
died	152	1.0	209	1.1	361	1.0
Total	15751	100.0	19354	100.0	35105	100.0

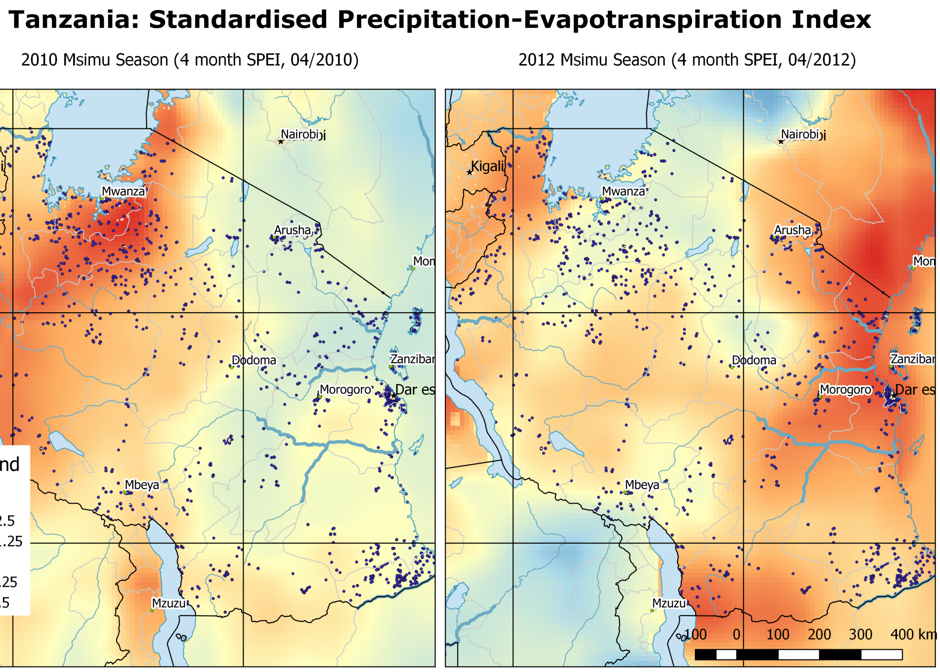


Figure 2.A1: SPEI: spatial and temporal variation

Table 2.A2: Descriptive statistics: shocks

	had shock	exog.	# w/ shock
Self-reported shocks not specific to household			
Large rise in price of food	0.23	0.78	927
Drought or floods	0.14	0.77	582
Large fall in sale prices for crops	0.09	0.82	365
Large rise in agricultural input prices	0.08	0.80	340
Severe water shortage	0.08	0.85	338
Crop disease or crop pests	0.08	0.76	319
Self-reported shocks specific to household			
Death of other family member	0.15	0.12	600
Livestock died or stolen	0.07	0.25	306
Death of a member of household	0.06	0.04	234
Hijacking/robbery/burglary/assault	0.03	0.02	113
Illness or accident of HH member	0.03	0.02	106
Break-up of the household	0.02	0.03	98
Loss of land	0.01	0.14	59
Fire	0.01	0.03	36
Lost salaried job or was not paid salary	0.01	0.12	26
Household business failure (non-agric.)	0.01	0.08	26
Dwelling damaged or destroyed	0.00	0.08	12
Jailed	0.00	0.00	10
other	0.01	0.23	40
Climate shocks			
$\|SPEI\|>1.5$ in previous season	0.06		252
SPEI $<$-1.5 in previous season	0.05		195
Observations			4100

Table 2.A3: Probability of leaving agricultural HHs (by various characteristics)

	Probability of leaving to			Number of
	anywhere	urban	rural	observations
Groups				
child (0-7): girl	0.100	0.009	0.038	3044
child (0-7): boy	0.100	0.006	0.041	2897
child (8-14): girl	0.097	0.008	0.030	2446
child (8-14): boy	0.065	0.003	0.016	2416
adult (15-24): girl, 6- yrs edu	0.285	0.028	0.198	723
adult (15-24): girl, 7+ yrs edu	0.270	0.064	0.151	1340
adult (15-24): boy, 6- yrs edu	0.161	0.020	0.103	805
adult (15-24): boy, 7+ yrs edu	0.165	0.042	0.080	1371
adult (25-34): girl, 6- yrs edu	0.134	0.015	0.092	524
adult (25-34): girl, 7+ yrs edu	0.140	0.038	0.083	847
adult (25-34): boy, 6- yrs edu	0.092	0.009	0.075	346
adult (25-34): boy, 7+ yrs edu	0.142	0.040	0.079	758
adult (35+ years)	0.042	0.006	0.026	5880
Age				
0-7 years old	0.100	0.008	0.040	5941
8-14 years old	0.081	0.005	0.023	4862
15-24 years old	0.224	0.045	0.128	4715
25-34 years old	0.137	0.032	0.083	2493
35+ years old	0.042	0.006	0.026	5880
Sex				
female	0.125	0.019	0.065	12325
male	0.094	0.014	0.044	11566
Education (if 14+ yrs)				
7+ yrs edu	0.133	0.033	0.072	7054
6- yrs edu	0.104	0.011	0.068	6121
Education (if 14+ yrs)				
9+ yrs edu	0.164	0.060	0.062	1495
7/8 yrs edu	0.124	0.025	0.075	5559
5/6 yrs edu	0.146	0.019	0.089	1298
4- yrs edu	0.092	0.009	0.062	4823
Unconditional probability	0.106	0.015	0.053	

Sample: agricultural households.

Table 2.A4: Composition of flows leaving agricultural HHs (by various characteristics)

	Share among entire sample	Share among movers to anywhere	urban	rural
Groups				
child (0-7): girl	0.130	0.122	0.078	0.093
child (0-7): boy	0.124	0.117	0.050	0.096
child (8-14): girl	0.105	0.096	0.053	0.060
child (8-14): boy	0.103	0.063	0.020	0.031
adult (15-24): girl, 6- yrs edu	0.031	0.083	0.056	0.115
adult (15-24): girl, 7+ yrs edu	0.057	0.145	0.240	0.163
adult (15-24): boy, 6- yrs edu	0.034	0.052	0.045	0.067
adult (15-24): boy, 7+ yrs edu	0.059	0.091	0.159	0.088
adult (25-34): girl, 6- yrs edu	0.022	0.028	0.022	0.039
adult (25-34): girl, 7+ yrs edu	0.036	0.048	0.089	0.056
adult (25-34): boy, 6- yrs edu	0.015	0.013	0.008	0.021
adult (25-34): boy, 7+ yrs edu	0.032	0.043	0.084	0.048
adult (35+ years old)	0.251	0.099	0.095	0.122
Age				
0-7 years old	0.249	0.225	0.116	0.179
8-14 years old	0.204	0.150	0.066	0.086
15-24 years old	0.197	0.401	0.533	0.460
25-34 years old	0.104	0.130	0.199	0.159
35+ years old	0.246	0.094	0.086	0.116
Sex				
female	0.516	0.586	0.591	0.612
male	0.484	0.414	0.409	0.388
Education (if 14+ yrs)				
7+ yrs edu	0.535	0.596	0.774	0.551
6- yrs edu	0.465	0.404	0.226	0.449
Education (if 14+ yrs)				
9+ yrs edu	0.113	0.156	0.303	0.100
7/8 yrs edu	0.422	0.440	0.471	0.451
5/6 yrs edu	0.099	0.120	0.084	0.125
4- yrs edu	0.366	0.284	0.141	0.324
# of observations	23891	2631	396	1312

Sample: agricultural households.

Table 2.A5: Probability of leaving non-agricultural HHs (by education)

	Probability of leaving to		
	anywhere	urban	rural
Education (if 14+ yrs)			
7+ yrs edu	0.112	0.052	0.022
6- yrs edu	0.090	0.033	0.020
Education (if 14+ yrs)			
9+ yrs edu	0.108	0.051	0.025
7/8 yrs edu	0.116	0.052	0.020
5/6 yrs edu	0.068	0.030	0.011
4- yrs edu	0.098	0.034	0.023
Unconditional probability	0.108	0.048	0.022

Sample: non-agricultural households.
Results are similar for urban households (Table 2.A13).

Table 2.A6: Composition of flows leaving non-agricultural HHs (by education)

	Share among movers to		
	anywhere	urban	rural
Education (if 14+ yrs)			
7+ yrs edu	0.828	0.858	0.808
6- yrs edu	0.172	0.142	0.192
Education (if 14+ yrs)			
9+ yrs edu	0.359	0.384	0.414
7/8 yrs edu	0.469	0.475	0.394
5/6 yrs edu	0.037	0.037	0.030
4- yrs edu	0.136	0.105	0.162
# of observations	493	219	99

Sample: non-agricultural households.
Results are similar for urban households (Table 2.A14).

Table 2.A7: Movers to urban vs. rest

	stay	urban	difference (se)	# stay	# to u
Demographic Characteristics					
age	35.2	23.9	-11.3 (0.59)***	13624	338
male	0.48	0.42	-0.065 (0.027)*	13624	338
father alive	0.54	0.71	0.17 (0.025)***	13623	338
father in HH	0.22	0.38	0.15 (0.027)***	13623	338
mother alive	0.68	0.84	0.16 (0.020)***	13624	338
mother in HH	0.29	0.53	0.24 (0.027)***	13624	338
currently married	0.45	0.11	-0.34 (0.018)***	13618	338
spouse or partner lives in HH	0.49	0.11	-0.38 (0.017)***	13619	338
# months absent in past year	0.50	1.51	1.02 (0.16)***	13624	338
Education					
years of education	5.19	7.35	2.16 (0.18)***	13576	338
ever been to school	0.76	0.93	0.16 (0.015)***	13615	338
completed year 4	0.71	0.91	0.21 (0.016)***	13578	338
completed year 7	0.54	0.77	0.24 (0.023)***	13578	338
completed year 9	0.12	0.31	0.20 (0.025)***	13578	338
completed year 11	0.049	0.15	0.10 (0.020)***	13578	338
completed more than year 11	0.014	0.047	0.034 (0.012)**	13578	338
currently in school	0.16	0.22	0.057 (0.023)*	13614	338
education spending (if curr in school)	138.1	227.3	89.1 (59.8)	2209	74
Health					
visited healthcare prov in past month	0.12	0.13	0.0059 (0.019)	13618	338
hospitalised in past year	0.057	0.027	-0.030 (0.0090)***	13618	338
physically handicapped (w1 only)	0.050	0.022	-0.028 (0.013)*	6101	134
slept under bednet last night	0.53	0.54	0.012 (0.027)	13621	338
gave birth in past 2 years (if f/12-49)	0.29	0.19	-0.097 (0.030)**	5440	183
Work and Employment					
worked in last 7 days	0.56	0.40	-0.16 (0.027)***	13594	336
worked in last 7 days or will resume	0.77	0.58	-0.19 (0.027)***	13592	335
unemployed: could but did not work	0.016	0.072	0.056 (0.014)***	13592	335
earned wage in last 7 days (if w/ job)	0.14	0.23	0.083 (0.030)**	10477	193
monthly wage (if earned wage)	426.2	743.8	317.6 (549.0)	1511	43
- winsorized at top 1%	283.4	428.4	145.0 (237.4)	1511	43
hrs unpaid non-ag HH work (prev wk)	8.03	7.11	-0.91 (0.77)	13520	333
hrs unpaid ag HH work (prev wk)	15.4	7.23	-8.21 (0.81)***	13590	335
Life Satisfaction: 1 (highest) - 7 (lowest)					
life sat: health	2.59	2.08	-0.51 (0.10)***	11069	229
life sat: financial situation	4.63	4.58	-0.056 (0.14)	10800	223
life sat: housing	3.36	3.10	-0.26 (0.13)*	11048	229
life sat: husband/wife (w1 only)	1.58	1.81	0.23 (0.24)	3044	21
life sat: job	3.35	3.66	0.32 (0.16)*	9269	164
life sat: overall	4.08	3.93	-0.15 (0.14)	10960	228
Anthropometric outcomes					
BMI for age (z-score)	-0.21	-0.052	0.16 (0.068)*	11609	240
height for age (z-score)	-1.43	-1.49	-0.063 (0.068)	11593	239
BMI for age z-score $<$-2	0.057	0.021	-0.036 (0.0095)***	11609	240
height for age z-score $<$-2	0.28	0.31	0.026 (0.030)	11593	239

Sample: agricultural households.

Table 2.A8: Movers to urban vs. rest. (Robustness: only HHs with leaver)

	stay	urban	difference (se)	# stay	# to u
Demographic Characteristics					
age	34.2	23.9	-10.3 (0.62)***	5491	338
male	0.47	0.42	-0.056 (0.028)*	5491	338
father alive	0.56	0.71	0.15 (0.026)***	5490	338
father in HH	0.26	0.38	0.11 (0.027)***	5490	338
mother alive	0.69	0.84	0.15 (0.021)***	5491	338
mother in HH	0.33	0.53	0.20 (0.028)***	5491	338
currently married	0.39	0.11	-0.28 (0.018)***	5488	338
spouse or partner lives in HH	0.44	0.11	-0.33 (0.018)***	5489	338
# months absent in past year	0.66	1.51	0.85 (0.17)***	5491	338
Education					
years of education	5.31	7.35	2.04 (0.18)***	5468	338
ever been to school	0.77	0.93	0.16 (0.015)***	5488	338
completed year 4	0.71	0.91	0.20 (0.017)***	5470	338
completed year 7	0.54	0.77	0.23 (0.024)***	5470	338
completed year 9	0.13	0.31	0.19 (0.026)***	5470	338
completed year 11	0.057	0.15	0.093 (0.020)***	5470	338
completed more than year 11	0.017	0.047	0.030 (0.012)**	5470	338
currently in school	0.17	0.22	0.046 (0.023)*	5487	338
education spending (if curr in school)	175.8	227.3	51.5 (62.8)	950	74
Health					
visited healthcare prov in past month	0.12	0.13	0.010 (0.019)	5489	338
hospitalised in past year	0.056	0.027	-0.029 (0.0093)**	5490	338
physically handicapped (w1 only)	0.047	0.022	-0.024 (0.014)	2094	134
slept under bednet last night	0.54	0.54	-0.0024 (0.028)	5491	338
gave birth in past 2 years (if f/12-49)	0.25	0.19	-0.055 (0.031)	2230	183
Work and Employment					
worked in last 7 days	0.54	0.40	-0.14 (0.028)***	5478	336
worked in last 7 days or will resume	0.76	0.58	-0.19 (0.028)***	5477	335
unemployed: could but did not work	0.017	0.072	0.055 (0.014)***	5477	335
earned wage in last 7 days (if w/ job)	0.15	0.23	0.079 (0.031)*	4168	193
monthly wage (if earned wage)	477.1	743.8	266.7 (553.5)	615	43
- winsorized at top 1%	332.6	428.4	95.8 (240.7)	615	43
hrs unpaid non-ag HH work (prev wk)	6.89	7.11	0.22 (0.78)	5434	333
hrs unpaid ag HH work (prev wk)	14.1	7.23	-6.84 (0.83)***	5477	335
Life Satisfaction: 1 (highest) - 7 (lowest)					
life sat: health	2.52	2.08	-0.44 (0.10)***	4306	229
life sat: financial situation	4.55	4.58	0.033 (0.14)	4212	223
life sat: housing	3.21	3.10	-0.11 (0.13)	4291	229
life sat: husband/wife (w1 only)	1.59	1.81	0.22 (0.24)	910	21
life sat: job	3.23	3.66	0.44 (0.16)**	3550	164
life sat: overall	3.97	3.93	-0.043 (0.14)	4260	228
Anthropometric outcomes					
BMI for age (z-score)	-0.18	-0.052	0.13 (0.069)	4547	240
height for age (z-score)	-1.36	-1.49	-0.13 (0.069)	4549	239
BMI for age z-score <-2	0.054	0.021	-0.033 (0.0098)***	4547	240
height for age z-score <-2	0.27	0.31	0.040 (0.031)	4549	239

Sample: agricultural households.

Table 2.A9: Movers to rural vs. rest.

	stay	rural	difference (se)	# stay	# to r
Demographic Characteristics					
age	35.6	25.4	-10.3 (0.43)***	12961	1001
male	0.49	0.36	-0.13 (0.016)***	12961	1001
father alive	0.53	0.71	0.19 (0.015)***	12961	1000
father in HH	0.22	0.34	0.13 (0.015)***	12961	1000
mother alive	0.67	0.82	0.15 (0.013)***	12961	1001
mother in HH	0.29	0.42	0.13 (0.016)***	12961	1001
currently married	0.46	0.24	-0.22 (0.014)***	12956	1000
spouse or partner lives in HH	0.49	0.31	-0.18 (0.015)***	12957	1000
# months absent in past year	0.48	1.07	0.59 (0.085)***	12961	1001
Education					
years of education	5.23	5.36	0.12 (0.11)	12916	998
ever been to school	0.76	0.79	0.028 (0.013)*	12952	1001
completed year 4	0.71	0.74	0.028 (0.015)	12918	998
completed year 7	0.54	0.57	0.026 (0.016)	12918	998
completed year 9	0.12	0.11	-0.017 (0.010)	12918	998
completed year 11	0.053	0.034	-0.018 (0.0061)**	12918	998
completed more than year 11	0.015	0.0090	-0.0058 (0.0032)	12918	998
currently in school	0.16	0.15	-0.0095 (0.012)	12951	1001
education spending (if curr in school)	142.6	119.8	-22.8 (23.3)	2128	155
Health					
visited healthcare prov in past month	0.13	0.095	-0.032 (0.0097)**	12956	1000
hospitalised in past year	0.057	0.047	-0.0096 (0.0070)	12955	1001
physically handicapped (w1 only)	0.051	0.024	-0.028 (0.0088)**	5897	338
slept under bednet last night	0.53	0.48	-0.049 (0.016)**	12958	1001
gave birth in past 2 years (if f/12-49)	0.29	0.26	-0.030 (0.019)	5030	593
Work and Employment					
worked in last 7 days	0.56	0.55	-0.0031 (0.016)	12935	995
worked in last 7 days or will resume	0.77	0.78	0.014 (0.014)	12932	995
unemployed: could but did not work	0.017	0.020	0.0031 (0.0046)	12932	995
earned wage in last 7 days (if w/ job)	0.15	0.13	-0.013 (0.013)	9895	775
monthly wage (if earned wage)	450.7	211.0	-239.7 (109.0)*	1452	102
- winsorized at top 1%	292.7	211.0	-81.7 (82.4)	1452	102
hrs unpaid non-ag HH work (prev wk)	8.08	6.99	-1.09 (0.45)*	12864	989
hrs unpaid ag HH work (prev wk)	15.3	14.0	-1.29 (0.58)*	12929	996
Life Satisfaction: 1 (highest) - 7 (lowest)					
life sat: health	2.60	2.36	-0.24 (0.063)***	10567	731
life sat: financial situation	4.66	4.26	-0.40 (0.081)***	10305	718
life sat: housing	3.37	3.14	-0.23 (0.076)**	10552	725
life sat: husband/wife (w1 only)	1.58	1.63	0.046 (0.088)	2974	91
life sat: job	3.37	3.10	-0.27 (0.080)***	8822	611
life sat: overall	4.09	3.85	-0.24 (0.081)**	10465	723
Anthropometric outcomes					
BMI for age (z-score)	-0.21	-0.20	0.0070 (0.037)	11080	769
height for age (z-score)	-1.44	-1.28	0.16 (0.039)***	11063	769
BMI for age z-score <-2	0.057	0.040	-0.017 (0.0074)*	11080	769
height for age z-score <-2	0.29	0.24	-0.046 (0.016)**	11063	769

Sample: agricultural households.

Table 2.A10: Movers to rural vs. rest. (Robustness: only HHs with leaver)

	stay	rural	difference (se)	# stay	# to r
Demographic Characteristics					
age	35.3	25.4	-9.88 (0.49)***	4828	1001
male	0.49	0.36	-0.13 (0.017)***	4828	1001
father alive	0.53	0.71	0.18 (0.016)***	4828	1000
father in HH	0.25	0.34	0.092 (0.016)***	4828	1000
mother alive	0.67	0.82	0.15 (0.014)***	4828	1001
mother in HH	0.32	0.42	0.099 (0.017)***	4828	1001
currently married	0.40	0.24	-0.16 (0.015)***	4826	1000
spouse or partner lives in HH	0.44	0.31	-0.13 (0.016)***	4827	1000
# months absent in past year	0.64	1.07	0.43 (0.089)***	4828	1001
Education					
years of education	5.44	5.36	-0.086 (0.12)	4808	998
ever been to school	0.77	0.79	0.017 (0.014)	4825	1001
completed year 4	0.72	0.74	0.015 (0.015)	4810	998
completed year 7	0.55	0.57	0.011 (0.017)	4810	998
completed year 9	0.15	0.11	-0.040 (0.011)***	4810	998
completed year 11	0.069	0.034	-0.035 (0.0068)***	4810	998
completed more than year 11	0.021	0.0090	-0.012 (0.0036)**	4810	998
currently in school	0.18	0.15	-0.025 (0.013)*	4824	1001
education spending (if curr in school)	190.2	119.8	-70.4 (31.6)*	869	155
Health					
visited healthcare prov in past month	0.13	0.095	-0.031 (0.010)**	4827	1000
hospitalised in past year	0.056	0.047	-0.0090 (0.0075)	4827	1001
physically handicapped (w1 only)	0.049	0.024	-0.026 (0.0097)**	1890	338
slept under bednet last night	0.55	0.48	-0.073 (0.017)***	4828	1001
gave birth in past 2 years (if f/12-49)	0.24	0.26	0.021 (0.021)	1820	593
Work and Employment					
worked in last 7 days	0.53	0.55	0.024 (0.017)	4819	995
worked in last 7 days or will resume	0.74	0.78	0.034 (0.015)*	4817	995
unemployed: could but did not work	0.020	0.020	-0.000037 (0.0049)	4817	995
earned wage in last 7 days (if w/ job)	0.16	0.13	-0.023 (0.014)	3586	775
monthly wage (if earned wage)	546.5	211.0	-335.5 (141.1)*	556	102
- winsorized at top 1%	362.3	211.0	-151.3 (94.4)	556	102
hrs unpaid non-ag HH work (prev wk)	6.89	6.99	0.11 (0.48)	4778	989
hrs unpaid ag HH work (prev wk)	13.6	14.0	0.44 (0.61)	4816	996
Life Satisfaction: 1 (highest) - 7 (lowest)					
life sat: health	2.52	2.36	-0.16 (0.067)*	3804	731
life sat: financial situation	4.60	4.26	-0.35 (0.085)***	3717	718
life sat: housing	3.22	3.14	-0.076 (0.080)	3795	725
life sat: husband/wife (w1 only)	1.59	1.63	0.035 (0.091)	840	91
life sat: job	3.27	3.10	-0.17 (0.084)*	3103	611
life sat: overall	3.99	3.85	-0.14 (0.086)	3765	723
Anthropometric outcomes					
BMI for age (z-score)	-0.17	-0.20	-0.034 (0.040)	4018	769
height for age (z-score)	-1.38	-1.28	0.10 (0.041)*	4019	769
BMI for age z-score <-2	0.055	0.040	-0.014 (0.0080)	4018	769
height for age z-score <-2	0.28	0.24	-0.037 (0.017)*	4019	769

Sample: agricultural households.

Table 2.A11: Probability of leaving rural HHs (by education)

	Probability of leaving to		
	anywhere	urban	rural
Education (if 14+ yrs)			
7+ yrs edu	0.129	0.023	0.077
6- yrs edu	0.102	0.008	0.069
Education (if 14+ yrs)			
9+ yrs edu	0.148	0.042	0.062
7/8 yrs edu	0.124	0.018	0.082
5/6 yrs edu	0.143	0.018	0.087
4- yrs edu	0.091	0.006	0.064
Unconditional probability	0.116	0.016	0.074

Robustness check. Sample: rural households.
Main analysis in table 2.6.

Table 2.A12: Composition of flows leaving rural HHs (by education)

	Share among movers to		
	anywhere	urban	rural
Education (if 14+ yrs)			
7+ yrs edu	0.578	0.747	0.546
6- yrs edu	0.422	0.253	0.454
Education (if 14+ yrs)			
9+ yrs edu	0.146	0.304	0.097
7/8 yrs edu	0.431	0.443	0.449
5/6 yrs edu	0.127	0.119	0.122
4- yrs edu	0.296	0.134	0.332
# of observations	1414	194	895

Robustness check. Sample: rural households.
Main analysis in table 2.7.

Table 2.A13: Probability of leaving urban households (by education)

	Probability of leaving to		
	anywhere	urban	rural
Education (if 14+ yrs)			
7+ yrs edu	0.120	0.063	0.023
6- yrs edu	0.101	0.040	0.022
Education (if 14+ yrs)			
9+ yrs edu	0.124	0.066	0.027
7/8 yrs edu	0.118	0.060	0.020
5/6 yrs edu	0.090	0.032	0.029
4- yrs edu	0.105	0.043	0.020
Unconditional probability	0.116	0.058	0.023

Robustness check. Sample: urban households.
Main analysis in table 2.A5.

Table 2.A14: Composition of flows leaving urban HHs (by education)

	Share among movers to		
	anywhere	urban	rural
Education (if 14+ yrs)			
7+ yrs edu	0.812	0.848	0.786
6- yrs edu	0.188	0.152	0.214
Education (if 14+ yrs)			
9+ yrs edu	0.332	0.357	0.365
7/8 yrs edu	0.480	0.491	0.421
5/6 yrs edu	0.043	0.031	0.071
4- yrs edu	0.145	0.121	0.143
# of observations	648	322	126

Robustness check. Sample: urban households.
Main analysis in table 2.A6.

Table 2.A15: Marginal effects: droughts and sorting (by education)

	edu 7+ urb	edu 7+ rur	edu 6- urb	edu 6- rur
1: abs(SPEI) >1.5	-0.011	0.001	0.000	0.021
	(-0.53)	(0.02)	(.)	(1.07)
SPEI: p(equal)		0.498		0.251
chi-squared	172.1	192.0	77.29	227.4
# of observations	3110	3110	3160	3383
% HHs with leaver	6.304	13.09	1.823	10.63

Table 2.A16: Out- and In-Migration

	out-migr.	out-migr.	in-migr.	in-migr.
1: drought/flood	0.067	-0.054	0.119	0.029
	(0.75)	(-0.56)	(1.20)	(0.27)
1: crop price fall	-0.169	-0.284	0.242	0.196
	(-1.31)	(-1.96)*	(2.06)**	(1.55)
1: food price rise	-0.104	-0.181	0.265	0.210
	(-1.33)	(-2.04)**	(3.31)***	(2.47)**
1: ag in price rise	0.087	0.102	-0.270	-0.245
	(0.68)	(0.75)	(-2.07)**	(-1.73)*
HH size		0.229		0.085
		(9.23)***		(5.22)***
Mean age		0.019		0.013
		(5.05)***		(3.36)***
Mean yrs. education		0.120		0.112
		(3.22)***		(2.85)***
Mean yrs. educ. (age 15-65)		-0.135		-0.079
		(-4.51)***		(-2.37)**
Max yrs. education		0.118		-0.010
		(4.52)***		(-0.44)
Plot size (acres)		-0.003		0.006
		(-1.79)*		(1.55)
Plot size (rank)		0.076		0.052
		(1.24)		(0.85)
Consumption (TSH)		-0.472		0.213
		(-2.55)**		(1.46)
Consumption (rank)		0.106		0.061
		(1.32)		(0.85)
1: HH has female head		0.506		0.338
		(5.97)***		(3.57)***
1: HH has good floor		0.124		0.088
		(1.30)		(0.86)
1: HH main income src. is ag.		-0.079		-0.033
		(-0.90)		(-0.38)
1: EA has primary school		0.170		0.252
		(1.15)		(1.66)*
1: EA has second. school		0.038		0.040
		(0.49)		(0.50)
EA distance to district HQ		-0.020		0.000
		(-1.48)		(0.01)
chi-squared	4.403	292.1	21.82	189.9
# of observations	4307	4100	4307	4100

Table 2.A17: Out- and In-Migration

	out-migr.	out-migr.	in-migr.	in-migr.
main				
1: abs(SPEI) >1.5	-0.233	-0.260	-0.479	-0.534
	(-1.17)	(-1.21)	(-3.19)***	(-3.35)***
HH size		0.225		0.102
		(9.57)***		(6.17)***
Mean age		0.019		0.013
		(5.01)***		(3.36)***
Mean yrs. education		0.127		0.116
		(3.43)***		(2.95)***
Mean yrs. educ. (age 15-65)		-0.129		-0.085
		(-4.27)***		(-2.58)***
Max yrs. education		0.101		-0.018
		(3.90)***		(-0.73)
Plot size (acres)		-0.004		0.003
		(-1.97)**		(1.17)
Plot size (rank)		0.183		0.217
		(3.24)***		(3.35)***
Consumption (TSH)		-0.432		0.227
		(-2.38)**		(1.56)
Consumption (rank)		0.088		0.082
		(1.10)		(1.12)
1: HH has female head		0.520		0.418
		(6.10)***		(4.35)***
1: HH has good floor		0.041		-0.026
		(0.42)		(-0.26)
1: HH main income src. is ag.		-0.101		-0.129
		(-1.16)		(-1.54)
1: EA has primary school		0.219		0.328
		(1.45)		(2.17)**
1: EA has second. school		-0.003		0.019
		(-0.04)		(0.23)
EA distance to district HQ		-0.019		0.001
		(-1.36)		(0.10)
chi-squared	1.378	265.3	10.19	131.0
# of observations	4304	4097	4304	4097

Table 2.A18: Marginal Effects: Drought and Migration (Out and In)

	out-migr.	out-migr.	in-migr.	in-migr.
1: SPEI <-1.1	-0.003	-0.006	-0.039	-0.034
	(-0.15)	(-0.27)	(-1.71)*	(-1.60)
controls	No	Yes	No	Yes
chi-squared	0.0236	264.2	2.943	130.2
# of observations	4307	4100	4307	4100
% HHs with leaver/arriver	28.33	28.33	29.07	29.07

Table 2.A19: Out-Migration: Different specifications for rainfall shocks (marginal effects)

	abs >1.5	z<-1.5	abs >1.5	z<-1.5	pbm(y1)	pbm(all)	abs(z)>1	z<-1
shock in season t-1	-0.049	-0.028	-0.032	0.042	-0.045	0.035	-0.011	0.000
	(-1.21)	(-0.60)	(-0.57)	(0.49)	(-0.81)	(0.45)	(-0.65)	(0.02)
shock in season t-2			0.053	0.056				
			(1.81)*	(1.72)*				
chi-squared	265.3	263.8	154.1	155.2	263.7	260.5	262.9	261.8
# of observations	4097	4097	2595	2595	4097	4097	4097	4097
# HHs with leaver	1218	1218	777	777	1218	1218	1218	1218
# HHs with shock	266	202	308	249			880	761
# HHs with both	64	55	97	79			241	219

Table 2.A20: In-Migration: Different specifications for rainfall shocks (marginal effects)

	abs >1.5	z<-1.5	abs >1.5	z<-1.5	pbm(y1)	pbm(all)	abs(z)>1	z<-1
shock in season t-1	-0.106	-0.086	-0.104	-0.084	-0.073	-0.101	-0.026	-0.005
	(-3.34)***	(-2.82)***	(-1.85)*	(-1.22)	(-1.40)	(-1.40)	(-1.33)	(-0.23)
shock in season t-2			0.055	0.056				
			(2.09)**	(1.91)*				
chi-squared	131.0	127.0	90.73	86.90	124.5	123.9	124.3	122.2
# of observations	4097	4097	2595	2595	4097	4097	4097	4097
# HHs with arriver	1251	1251	700	700	1251	1251	1251	1251
# HHs with shock	266	202	308	249			880	761
# HHs with both	55	44	101	81			235	215

Chapter 3

PROXY MEANS TESTING REVISITED

3.1. Introduction

Many social programs in developing countries specifically aim to benefit poor households whose income, consumption, or wealth is below a given threshold (depending on the definition of poverty). However, large shares of informal and self-employment in developing countries generally make it difficult to observe income, while consumption (and wealth) are also difficult to observe directly. Proxy means testing (PMT) promises to nevertheless identify poor households using information on household assets and demographic characteristics that are correlated with poverty but can easily be collected and verified through surveys. Using these proxies the probability that a household is poor is then estimated through a statistical model.[1] PMT can thereby be employed to identify and target poor households or to estimate poverty rates.

Recently, however, the predictive power of PMT has become subject to debate: Brown et al. (2016) argue that PMT frequently fails to identify poor households. When used to target the bottom 20% (40%) of households, the authors find that their PMT specifications predict a large share of poor households to be non-poor,

[1] Other specifications estimate consumption or income and then predict a household to be poor if these are below the poverty line.

generating a high rate of exclusion errors: "standard proxy-means testing helps filter out the nonpoor, but excludes many poor people." "PMT allows a substantial reduction in the rate of inclusion errors; ... [however], this success at avoiding leakage to the nonpoor comes with seemingly weak coverage of poor people - a high rate of exclusion errors. In other words, the methods do not reliably reach the poorest." The authors hence note that "[o]ne can understand why many of those accepted or rejected might be tempted to believe that econometric targeting is something like a random lottery, or maybe even divine intervention."

In this paper, I revisit PMT and the results of Brown et al. (2016) to examine the causes of its poor performance. As the authors suggest that PMT is "particularly deficient in reaching the poorest [households]," I carefully examine the performance of PMT by income quantiles to understand how misclassification (exclusion errors for the poor and inclusion errors for the rich) varies by income and affects the performance of PMT. This allows me to understand the gravity of PMT misclassifications: a wrongfully excluded household just below the poverty line might be a lesser exclusion error than a wrongfully excluded household far below the poverty line; similarly a wrongfully included household just above the poverty line might be a lesser inclusion error than a wrongfully included much richer household. I also examine the performance of Poverty Score Cards (also referred to as the Poverty Probability Index), which are a popular and commonly implemented version of proxy means testing (Innovations for Poverty Action, 2018). The version of PMT that Brown et al. (2016) find and criticise to perform relatively poorly uses proxies that are chosen ad-hoc. Poverty Score Cards, on the other hand, are carefully designed to use the proxies with the highest predictive power for a given country. I hence analyse their performance to understand whether the critique of Brown et al. (2016) is more widely relevant when proxies are carefully chosen.

I find that poor calibration is a major, mechanical driver of the poor performance of PMT in several of the specifications examined by Brown et al. (2016). In a two-step procedure, the authors first predict the consumption of each household using OLS (or a version thereof) and then classify households whose predicted consumption is below a given consumption decile as poor. However, their OLS specifications tend to poorly predict the consumption of households, which leads to large under- or over-estimates of the share of poor households: when setting the poverty line at the 20th consumption percentile, OLS only predicts 7.5% of households to have a consumption below this 20th percentile. Mechanically, this results in a large number of incorrectly excluded households; it should also bewilder any user of PMT who knows that by their definition 20% of households are poor. For a poverty line at the 20th consumption percentile, PMT as examined by Brown et al. (2016) hence performs poorly not because the wrong households are classified as poor, but because far too few households are classified as poor; widely inaccurate OLS estimates of the share of poor households are a major driver of poor PMT performance and high exclusion or inclusion error rates. For a poverty line at the 40th consumption percentile, OLS tends to estimate the share of poor households more accurately; however, the share of households predicted to be poor on average still differs by 6 percentage-points from 40%, the share defined to be poor.

I hence examine the performance of PMT when calibrating, for a given poverty line, the predicted poverty rates to be equal to actual poverty rates. To do so, I estimate a logit model on a dummy variable indicating whether the household is below the poverty line. I then use the model to estimate the probability of being below the poverty line for each household and predict all households above a certain probability cut-off to be poor, and households below this cut off to be non-poor. I choose this probability cut-off so that the proportion of households that I predict to be poor is equal to the proportion of households that are actually poor. Unlike Brown et al. (2016), I use a $1.90 poverty line rather than coding the bottom 20% (40%) of households in every country as poor to account for the different wealth of countries.

Calibrating my model to match actual poverty rates (and using country-specific poverty rates) allows for an appreciation of the realistic performance of PMT without mechanically arising exclusion errors. I find that on average across countries, PMT correctly classifies approximately two thirds of poor households, and thereby substantially out-performs chance, which would classify 40% of poor households correctly. Nevertheless, this also means that PMT misses one third of poor households. Substantial variation exists between countries: for Malawi, PMT adequately classifies 82% of poor households, while for Ethiopia, PMT classifies only 51% of poor households. However, this variation appears to be largely induced by differences in poverty rates between countries; since in countries with higher poverty rates PMT classifies a larger share of households as poor, it also accurately classifies a larger share of the poor. Nevertheless, while by this standard PMT performs worst in absolute numbers for Ethiopia, the country with the lowest poverty rate, its gains relative to chance are largest, as PMT outperforms chance by a factor of almost 2.

I next investigate which poor households are most likely to be erroneously classified as non-poor (and, conversely, which non-poor households are most likely to be classified as poor). To do so, I examine the probability of being classified as poor by consumption decile and find that the poorer a household is, the more likely it is to also be classified as poor. In other words, unlike Brown et al. (2016) suggest, the poorest households are the ones that are also *most* likely to be classified as poor (and correctly targeted by PMT). Nevertheless, misclassification as non-poor does occur even for the poorest households, and, in some cases, to a substantial extent.

Last, I explore the performance of Poverty Score Cards, a popular implementation of proxy means testing. Poverty Score Cards are readily available for many countries, can be used by non-technical users, and include fewer items as proxies, which have been chosen to maximise predictive power and classification performance; this makes them an attractive tool. However, I find that Poverty Score Cards on average perform no better, and if anything slightly worse, than the relatively ad-hoc PMT regressions that I have been considering. While interesting in itself, this also suggests that my results do not stem from a poor choice of proxies.

Ultimately, any verdict on the performance of PMT will depend on one's benchmark and its purpose. PMT performs substantially better than chance, but its performance is also substantially short of perfect: on average across the five countries I study, only two out of three poor households are also classified as poor. Given the speed and ease with which information on proxies can be collected (particularly compared to the data required to estimate household consumption), the performance of PMT might be seen as impressive in some settings and PMT might serve as a useful rule of thumb for whether a household is likely to be poor. Beyond uses as rule of thumb, however, PMT's error rates are likely to make it a sub-optimal tool for restricting or granting access to social programs or transfers.

My analysis draws on and contributes to a recent literature on targeting and proxy means testing. Unsurprisingly, it is most closely related to Brown et al. (2016), whose results I re-examine. Providing an overview of recent literature on targeting and proxy means testing, the authors also note a surprising gap in research: although used in research and policy alike (Grosh et al., 2008; del Ninno and Mills, 2015), little research beyond individual country studies (Cnobloch and Subbarao, 2015; Pop, 2015; Stoeffler et al., 2015) assesses the performance of proxy means testing. In doing so, Brown et al. (2016) hence make an important contribution and provide results comparable across countries.

A number of other research papers are also of wider relevance: Coady et al. (2004) provide a comprehensive review of targeting in 122 poverty interventions in 48 countries and examine targeting effectiveness (as well as potential correlates) in these interventions. For the interventions they study, targeting (means testing, geographic targeting, and/or targeting involving self-selection) at the median transfers 25% more resources to the poor (by their definition the lower 40% of the income distribution) than universal allocation would. However, this metric masks wide variation: while the most effective intervention transfers four times as many resources to the poor as uniform allocation would, a quarter of the examined programs are regressive, transferring less to the poor than universal allocation would. Richer countries (presumably with higher administrative capacity), countries with more accountable governments, and countries with more inequality (making it potentially easier to identify poor people and increasing the gains from targeting) perform better on average.

Aiming to update the review of Coady et al. (2004), Devereux et al. (2017) emphasize that there are important political and social dimensions to targeting, which might affect its performance: involving the transfer of economic resources from the state to its citizens, the authors emphasize that targeting does not take place in a political vacuum. Local arrangements and contexts furthermore appear to be important factors in the optimal choice of targeting method, as no one-size-fits-all method is apparent to the authors.

Alatas et al. (2012) empirically assess the performance of three targeting methods in a field experiment in Indonesia and similarly highlight that there might be important non-technical aspects to targeting: comparing proxy means testing, community targeting (whereby villagers themselves rank local households by wealth), and a hybrid, the authors find that community targeting performs slightly worse than PMT (compared to objective measurements of wealth), but yields substantially higher satisfaction with villagers and generates fewer complains. A slightly different, but locally widely shared definition of poverty might explain these findings, the authors suggest.[2] Elite capture, on the other hand, did not appear to be a threat to effective community targeting in their study; however, this finding might be specific to smaller, one-time transfers (in this case USD 3, approximately the daily wage of a manual labourer in the setting).

Similar in their approach to this study, but with a different research question, Diamond et al. (2016) examine and compare the performance of regression-based PMT methods (OLS, WLS, Logit, and Lasso Regression) and Poverty Score Cards in predicting poverty rates for a given area (rather than evaluating their performance in predicting the poverty status of a given household). While they find little difference in performance between the methods (using the same explanatory variables) when estimating poverty rates at the national level, Poverty Score Cards perform measurably worse than regression models for subnational populations. Since regression models can be trained for specific target populations, while Poverty Score Cards can not, this might not be a surprising finding; however, it does highlight that the simplicity of Poverty Score Cards comes at the potential cost of lower predictive power when used for populations that are not nationally representative.

This paper is organised as follows: section 3.2 describes data and methodology, section 3.3 revisits the findings of Brown et al. (2016), and section 3.4 presents my analysis of the performance of proxy means testing using country-specific poverty rates. Section 3.5 concludes.

[2] Comparing the targeting performance of PMT and community-based targeting in Cameroon, Stoeffler et al. (2016) similarly find that communities appear to apply a different definition of poverty, selecting households with low human and physical capital, rather than low consumption.

Table 3.1: Data and Surveys

Country	Survey	HHs
Ethiopia	2013/14 Socioeconomic Survey	5262
Malawi	2013 Integrated Household Panel Survey	4000
Nigeria	2012/13 General Household Panel Survey	4536
Tanzania	2012/13 National Panel Survey	5010
Uganda	2011/12 National Panel Survey	2845

3.2. Data and Methodology

3.2.1. Data Sources

I set up a laboratory for evaluating the performance of PMT by using household surveys that have information on both the proxies commonly used by PMT and on household consumption. This allows me to simulate the performance of proxy means testing: upon training a model, I use only information on households' proxies from the surveys to predict whether a household is poor; I then compare this prediction to households' actual poverty status and examine whether the household was correctly predicted to be (non-)poor by the proxy means test. The World Bank's Living Standards Measurement Studies (LSMS) are a high-quality source of data for this purpose. I employ five such surveys, for Ethiopia, Malawi, Nigeria, Tanzania, and Uganda; table 3.1 lists my data sources.

Each survey contains information on a wide range of household characteristics and also measures household consumption; I use the annual household consumption aggregates calculated by national statistical agencies and provided with the LSMS.[3] In all countries, consumption is locally and temporally adjusted (as households are surveyed at different times of the year and costs of living also vary within countries); my data hence captures real consumption. Figure 3.A1 shows the distributions of consumption for the countries I examine.

[3]For Uganda, only consumption measured *per adult equivalent* is readily available; I hence use this measure. For the remaining four countries, I use consumption per capita.

Table 3.2: Calculations: \$1.90 Poverty Line

Country	2011 PPP	CPI	LCU PL	% BPL
Ethiopia	5.44	1.83	6913.67	84.6
Malawi	78.02	1.45	78306.12	31.1
Nigeria	79.53	1.18	64867.92	44.3
Tanzania	585.52	1.31	530799.17	47.9
Uganda	946.89	0.69	449500.75	48.5

Data on 2011 PPP conversion rates and on the CPI in the survey year (base 2011 = 1) are from PovcalNet. For Uganda consumption is given in 2005/06 prices, hence the CPI is below 1. LCU PL expresses the \$1.90/day poverty line in terms of annual consumption in local currency units and prices at the year of the survey. % BPL states the proportion of households with per-capita (for Uganda, per adult-equivalent) consumption below this poverty line.

3.2.2. Poverty Lines

Proxy Means Testing aims to predict whether a household, as measured by its consumption, is poor (or not). This requires a definition of poverty or a poverty line that specifies the amount of consumption below which a household is considered poor. Almost by definition, all poverty lines are somewhat arbitrary; nevertheless "a dollar a day" (Ravallion et al., 2009) has gained widespread acceptance as a poverty line in policy circles and is thought to provide a reasonable measure of extreme poverty. Accounting for inflation, a dollar a day (when the poverty line was first proposed in 1990) approximately equals \$1.90 a day in 2011 prices (see Ferreira et al., 2015a,b), which I use as poverty line.

I employ two approaches to draw the \$1.90 consumption poverty line. Firstly, the PovcalNet database maintained by the World Bank (The World Bank, 2018) provides me with information on the share of households with consumption below \$1.90 per day (and member) in 2011 PPP prices. I hence draw the poverty line for each country by coding the corresponding share of households with the lowest consumption as poor; mechanically this causes the poverty rates in my surveys to be equal to the poverty rates reported by PovcalNet.[4] Secondly, rather than

[4]For Ethiopia, Malawi, and Nigeria, the latest available PovcalNet data on poverty rates is three

Table 3.3: Poverty Rates

Country	GDP ($ pc)	PHC WB	WB (year)	PHC data
Ethiopia	571	33.5	2010	80.2
Malawi	333	70.9	2010	25.4
Nigeria	2997	53.5	2009	29.1
Tanzania	902	49.1	2011	39.1
Uganda	648	36.6	2012	41.8

Proportion of individuals below the poverty line according to PovcalNet (PHC WB) and according to survey data when considering households with consumption below $1.90/individual/day as poor (PHC data). WB (year) states the year of the PovcalNet estimate.

relying on poverty rates from PovcalNet, I also examine the consumption of all households and code the households whose consumption (in local currency units and adjusted for inflation) is below $1.90 per day and member as poor. Table 3.2 provides details on my calculations: I convert the $1.90 poverty line (expressed per year in 2011 prices) to local currency units (in 2011 prices) using 2011 PPP conversion rates and then account for local inflation between 2011 and the survey year through CPI indices (using also data from PovcalNet).

Surprisingly, the two approaches lead to relatively large discrepancies as seen in table 3.3: the second column states the proportion of individuals that are poor according to the first approach, when I code the same share of households as poor as reported by PovcalNet: correspondingly, the fourth column states the proportion of individuals that are poor according to the second approach, when I code households whose reported consumption is below $1.90/day/member as poor. Omitted items from consumption calculations might be a major cause of these differences; the LSMS survey for Ethiopia, for example, does not include rent/the imputed value of housing in consumption calculations, causing the survey to understate household consumption (Central Statistical Agency (Ethiopia) and

years older than the surveys I use; for Tanzania, it is one year older. The drawback of this approach is that it does not account for changes in poverty rates that occurred during these periods.

The World Bank, 2015). Conversely, the survey report for Malawi emphasizes that poverty rates estimated from the survey are only a lower bound for actual rates due to the timing of data collection (National Statistical Office (Malawi), 2014). Given these discrepancies (which are substantial for Ethiopia, Malawi, and Nigeria and smaller for Tanzania and Uganda), I rely on my first approach, and code the households with the least consumption as poor such that my poverty rates match those reported by PovcalNet. Still, as a robustness check, I also examine PMT performance when using my second approach and discuss the results in section 3.4.5.

3.2.3. Proxy Means Tests

PMT exploits the relationship between households' assets and demographic characteristics (the proxies) and households' consumption in order to predict whether a household is poor. Simply put, poorer households tend to have fewer assets, worse housing, and often particular demographics. Observing these characteristics hence provides information on expected household consumption. Proxies are chosen such that they are both correlated with household consumption and can easily be collected and verified by enumerators. I follow Brown et al. (2016) in their choice of proxies.[5] Table 3.A1 provides summary statistics on these for all countries. Using these proxies, I estimate a household-level PMT logit model for every country, regressing a dummy variable (equal to 1 if a given household is poor) on the proxies.

Once estimated, I consider how well the PMT model performs by using it to predict the poverty status of every household and then compare households' predicted poverty status to their actual poverty status.[6] The logit model predicts a

[5]Proxies focus on dwelling quality, general household characteristics, and demographic composition. The list of proxies is: (i) Housing characteristics: type of toilet, flooring, walls, roof, and cooking fuel used, (ii) Household characteristics: household size, whether the household head completed primary education, whether the household head completed secondary education, gender of the household head, marital status of the household head, employment status of the household head, religion of the household head, and (iii) Demographic composition of the household: proportion of members female and 0-5 years, proportion of members male and 0-5 years, proportion of members female and 6-14 years, proportion of members male and 6-14 years, proportion of members female and 65+ years, proportion of members male and 65+ years.

[6]I follow Brown et al. (2016) in their approach. Since my logit model estimates relatively

probability of poverty for every household; however, I require a binary prediction. I hence need to choose a threshold probability above which I classify a household as poor and below which I classify as a household as non-poor. In my main specification, I choose this threshold such that the proportion of households predicted to be poor is equal to the proportion of households who are actually poor. In section 3.4.5, I also examine a specification in which I instead classify all households with a predicted probability of poverty equal to or larger than 0.5 as poor.

3.2.4. Poverty Score Cards

I also examine the performance of proxy means testing using one particular and popular implementation of PMT, Poverty Score Cards (PSCs). Developed by Marc Schreiner and now maintained by Innovations for Poverty Action as the Progress Out of Poverty Index (PPI), Poverty Score Cards are ready-to-use survey instruments for proxy means testing that are available for a large number of countries (Innovations for Poverty Action, 2018). Starting from nationally representative household surveys, Schreiner et al. use an iterative procedure to identify the ten indicators most predictive of households' poverty status that can be easily collected and verified by enumerators Diamond et al. (2016). Using these indicators, the proxies, Schreiner then estimates a logit model and re-scales the estimated model coefficients into points to generate a score card, which allows for the easy calculation of poverty probabilities by non-technical users using pen and paper. For every proxy, a household receives a score (for example, if the proxy relates to the household's roof, 0 points if the household has a makeshift roof and 5 points if the household has a proper roof); with the help of a conversion table, the total score of a household is then translated into a probability of poverty.

I evaluate the performance of Poverty Score Cards by simulating their use: taking the Poverty Score Card provided for each country, I score every household surveyed by the LSMS in that country(Schreiner, 2015a,b,c, 2016a,b). I then pre-

few parameters, I also do not split my data into test and training sets. Considering leave-one-out-validation as a thought experiment suggests that this is a reasonable approach: estimating the model from the full dataset except one observation will lead to virtually the same parameter estimates, which in turn will lead the predicted poverty status for the excluded household to be virtually the same as when estimating the model from the full dataset.

dict the lowest-scoring households in that country to be poor, such that the poverty rate matches the one reported for the country.[7] As before, I also test the robustness of my results by alternatively predicting all households with a poverty probability greater than or equal to 0.5 according to the PSC as poor.

Figure 3.1 shows the distribution of poverty scores according to score cards for each country.

3.2.5. Evaluation

Simulating a proxy means test yields two measures for every household: I know whether the household is *truly* poor according to the consumption data from the surveys; second, I know whether the household is *predicted* to be poor according to the proxy means test. This allows me to categorise every household into one of four categories: true positives (predicted to be poor and actually so), false positives (predicted to be poor, but actually non-poor), true negatives (predicted to be non-poor and actually non-poor), and false negatives (predicted to be non-poor, but actually poor).

I measure the performance of PMT by three statistics: $P(+|D)$ states the proportion of poor households which are also predicted to be poor; the higher the proportion of poor households who are also predicted to be poor, the fewer poor households are missed by the test. $P(\neg D|-)$ states the proportion of households predicted to be non-poor that actually are non-poor: the higher this proportion, the lower the rate of false negatives. Lastly, I also report the total proportion of households that are classified correctly; this is the proportion of all households that are true positives or true negatives, $P(+, D) + P(-, \neg D)$.

To better appreciate the performance of proxy means tests, I compare them to the performance of a random coin toss predictor which matches the poverty rate for each country (such that $P(+) = P(D)$). Since the coin's prediction is independent of a household's actual poverty status, its performance is $P(+|D) =$

[7] Due to the discreteness of PSC scores, the fractions of households I predict to be poor (i.e. those with PSC scores below a certain threshold) do not exactly match the poverty rates reported for the countries; however they are very close: for Ethiopia, the poverty rate using PSC scores is 26.2% vs. 26.6% in World Bank data; for Malawi, this is 60.9% vs. 60.9%; for Nigeria 42.3% vs. 42.7%; for Tanzania 38.1% vs. 37.1%; for Uganda 33.4% vs. 34.3%.

$P(+)$ and $P(\neg D|-) = 1 - P(+)$. I report these benchmarks for each country in the tables presenting my results.

Brown et al. (2016) focus on inclusion and exclusion error rates in their discussion of PMT performance. The statistics I report are closely related: the inclusion error rate (IER) is proportion of households predicted to be poor who are not, $P(\neg D|+) = 1 - P(D|+)$; the exclusion error rate (EER) is the proportion of poor households who are not predicted to be poor, $P(-|D) = 1 - P(+|D)$. Since I set $P(+)$ equal to $P(D)$, these are equal. Brown et al. (2016) report this statistic as the total error rate.

When I revisit the results of Brown et al. (2016) in section 3.3, I make use of the following identities.

- To calculate $P(+)$, the proportion of households predicted to be poor, I apply Bayes' Rule: $P(+) = \frac{P(D)P(+|D)}{P(D|+)}$. Noting that $P(+|D) = 1 - EER$ and $P(D|+) = 1 - IER$, I can calculate $P(+)$ given the data reported. Bayes' Rule can also be derived intuitively: $(1 - EER)$ is the fraction of poor that are correctly identified as positive; they are the true positives. However, these only account for the share (1-IER) of all positives. Hence, the total share of positives is $P(D)(1 - EER)/(1 - IER)$, as also obtained by Bayes' Rule.

- To calculate the proportion of households classified correctly, I sum the shares of true positives and true negatives, $P(+|D)P(D)+P(-|\neg D)P(\neg D)$. I note that $P(+|D)$ is $1 - EER$, the complement to the exclusion error rate. For the second term, I note that $P(-|\neg D) = 1 - P(+|\neg D) = 1 - \frac{P(+)P(\neg D|+)}{P(\neg D)}$ and that $P(\neg D|+)$ is the inclusion error rate.

3.3. Specification and performance: Brown revisited

I commence my analysis by revisiting the results of Brown et al. (2016), summarized in table 3.4. Each panel presents the PMT classification performance of a particular approach as reported by the authors in tables 5-8 of their paper:

OLS with poverty lines at the 20th and 40th consumption percentile in panels 1 and 2, quantile regression for a poverty line at the 40th percentile in panel 3, and weighted least squares with zero-weights on non-poor households above the 40th percentile (panel 4) and on households above the 60th percentile (panel 5).[8] Inclusion and exclusion error rates are as reported by the authors; the remainder are my own calculations as outlined in section 3.2.5.

Already at first sight, large differences in classification performance between the different approaches become obvious, particularly when looking at the exclusion error rate: on average, 77% of poor households are excluded (as they are predicted to be non-poor) by basic OLS PMT regressions with a poverty line at the 20th consumption percentile (top panel of table 3.4), while these are only 37% for a poverty line at the 40th consumption percentile (panel 2), 22% in a quantile regression for the 40th consumption percentile (panel 3), and 0% (4%) when setting the weights for all households above the 40th (60th) consumption percentile to zero in a weighted least squares regression (panels 4 and 5). In other words, while the first approach wrongly classifies 3 of 4 poor households as non-poor, the last approach correctly classifies 19 of 20 poor households as poor. These are large and staggering differences, which naturally beg an explanation.

An examination of inclusion error rates provides part of the explanation. Overall, differences in IERs between the approaches are substantially smaller than differences in EERs; however, this does not mean that inclusion error rates play a minor role in understanding the approaches' predictive performance, as they have to be read together with the EERs: The IER for Ethiopia in the top panel (OLS with a poverty line at the 20th consumption percentile) is 0.52, for example; only roughly 50% of positive households (those classified as poor) are true positives (actually poor). At the same time, the EER reveals that only 5% of poor households (the bottom 20% of households in consumption terms are defined as poor) are classified as positives. Hence, 1% of all households are true positives. Since

[8] I do not revisit the "enhanced PMT" regressions examined by the authors, as these also include covariates (such as self-reported information on shocks experienced by the household) that are generally not collected by proxy means tests and are difficult to verify. In any case, the authors find only small gains from such additional information that do not qualitatively change their findings. To be concise, I also do not re-visit (except for OLS) specifications with a poverty line at the 20th consumption percentile, which generally perform worse than those with a poverty line at the 40th consumption percentile.

Table 3.4: Key Results of Brown et al. (2016)

	Ethiopia	Malawi	Nigeria	Tanzania	Uganda	mean
T5, OLS PMT (0.2)						
IER	0.515	0.431	0.332	0.396	0.357	0.406
EER	0.945	0.880	0.548	0.822	0.663	0.772
$P(D\|+)$ = 1-IER	0.485	0.569	0.668	0.604	0.643	0.594
$P(+\|D)$ = 1-EER	0.055	0.120	0.452	0.178	0.337	0.228
P(+)	0.023	0.042	0.135	0.059	0.105	0.073
T5, OLS PMT (0.4)						
IER	0.396	0.333	0.247	0.323	0.350	0.330
EER	0.562	0.451	0.243	0.291	0.294	0.368
$P(D\|+)$ = 1-IER	0.604	0.667	0.753	0.677	0.650	0.670
$P(+\|D)$ = 1-EER	0.438	0.549	0.757	0.709	0.706	0.632
P(+)	0.290	0.329	0.402	0.419	0.434	0.375
T6, QR PMT (0.4)						
IER	0.441	0.383	0.299	0.364	0.407	0.379
EER	0.292	0.304	0.164	0.153	0.172	0.217
$P(D\|+)$ = 1-IER	0.559	0.617	0.701	0.636	0.593	0.621
$P(+\|D)$ = 1-EER	0.708	0.696	0.836	0.847	0.828	0.783
P(+)	0.507	0.451	0.477	0.533	0.559	0.505
T7, poor-only WR PMT (0.4)						
IER	0.598	0.597	0.560	0.588	0.581	0.585
EER	0.000	0.000	0.004	0.001	0.001	0.001
$P(D\|+)$ = 1-IER	0.402	0.403	0.440	0.412	0.419	0.415
$P(+\|D)$ = 1-EER	1.000	1.000	0.996	0.999	0.999	0.999
P(+)	0.995	0.993	0.905	0.970	0.954	0.963
T8, poor plus 20% WR PMT (0.4)						
IER	0.577	0.521	0.428	0.466	0.500	0.498
EER	0.024	0.040	0.051	0.049	0.037	0.040
$P(D\|+)$ = 1-IER	0.423	0.479	0.572	0.534	0.500	0.502
$P(+\|D)$ = 1-EER	0.976	0.960	0.949	0.951	0.963	0.960
P(+)	0.923	0.802	0.664	0.712	0.770	0.774

true positives make for half of all positives, roughly 2% of all households are classified as positive. In other words: although 20% of households are poor by the definition of Brown et al. (2016), their regression only predicts 2% of households to be poor. When I calculate the poverty rates $P(+)$ predicted by PMT for other countries, I find that while Ethiopia fares worst, OLS performs very poorly for a poverty line at the 20th consumption percentile overall: on average only 7% of households are predicted to be poor (while 20% actually are).[9]

Mechanically, classifying too few households as poor causes a high exclusion error rate (in this case, even if the 7% of households predicted to be poor are all actually poor, the exclusion error rate will be 0.65). The two weighted least squares approaches that obtained very low exclusion error rates (panels 4 and 5) lie on the other end of the spectrum: weighted least squares (WLS) as examined by Brown et al. (2016) delivered exclusion error rates of 0.1% and 4%. However, calculating the poverty rates implied by these approaches reveals the cause of low inclusion error rates: instead of predicting 40% of households to be poor, WLS in panel 4 predicts 96% of households to be poor; WLS in panel 5 predicts 77% of households to be poor. Low exclusion error rates are the obvious benefit and consequence of large overpredictions of the share of poor households, but they are much less impressive once one becomes aware that these approaches simply predict every household (or a very large share of them) to be poor.

Large variations in the share of households predicted to be poor thus explain the large variation in inclusion error rates found by Brown et al. (2016). When I analyse their findings and calculate the implied predicted poverty rates, I find large mis-estimates which vary by approach; these explain and mechanically drive the poor performance of PMT that the authors criticise. Of the five approaches, two (both versions of WLS) over-predict the share of poor households by more than 90%, while a third (OLS with a poverty line at the 20th consumption percentile) under-predicts the share of poor households by more than 60%.

However, my analysis also reveals that two approaches perform better in correctly predicting the share of poor households: OLS for a poverty line at the

[9] As the authors themselves note, OLS is a poor predictor of consumption at lower percentiles: while an unbiased predictor of the mean, OLS overpredicts the consumption of poorer households (and under-predicts the consumption of richer households); this leads to an underprediction of the share of households below the poverty line for poverty lines at low consumption percentiles.

40th consumption percentile predicts on average 37.5% of households to be poor; quantile regression (with and for a poverty line at the 40th consumption quantile) predicts on average 50.5% of households to be poor. Better calibrated and with fewer mechanically arising inclusion or exclusion errors, these approaches thereby provide a more realistic picture of the (best) possible performance of PMT: on average, OLS correctly detects between 6 and 7 of every 10 poor households (panel 2); 2 out of 3 households predicted to be poor are actually poor. As it overpredicts the share of poor households, quantile regression has a lower exclusion error rate and correctly detects almost 80% of poor households; however, this comes at the cost of a slightly higher inclusion error rate, as almost 4 out of 10 households predicted to be poor are not.

These better calibrated approaches hence give a glimpse of what the performance of PMT can be when its predicted poverty rates are close to actual ones and thereby do not mechanically generate inclusion or exclusion errors. Still, further improvements are possible and indeed necessary when wanting to eliminate mechanical sources of inclusion and exclusion errors: although performing best, OLS still predicts poverty rates, which on average differ by 6 percentage points from the actual poverty rate of 40%; for quantile regression, this difference is on average 13 percentage-points. In the next section, I hence evaluate the performance of PMT when well-calibrated, such that the predicted poverty rate matches the actual rate.

3.4. PMT, well-calibrated

I now evaluate the performance of well-calibrated proxy means testing, setting the predicted poverty rates equal to the actual ones: as discussed in section 3.2.3, I estimate a logit model to predict whether a household is poor. Rather than giving a binary prediction, the model estimates a probability. I hence choose the probability threshold (above which a household is classified as poor) such that the proportion of households predicted to be poor matches the proportion of households that are actually poor.[10] Instead of using poverty lines at the 20th and 40th

[10]In section 3.4.5, I also present results when classifying all households with a predicted probability of poverty greater than or equal to 0.5 as poor.

consumption percentiles, which are somewhat arbitrary given the different incomes of the examined countries, I use the World Bank's \$1.90/day poverty line, as discussed in section 3.2.2.

Table 3.5 summarises my results on the performance of logit PMT regressions: $P(+|D)$ states the proportion of poor households that are correctly classified as poor.[11] $P(\neg D|-)$ states the proportion of households classified as non-poor that actually also are non-poor.[12]

3.4.1. Overall performance

I begin by examining the overall performance of my calibrated proxy means tests. On average, two out of three poor households are correctly classified as poor across the five countries. However, relatively large differences exist between countries: for Ethiopia, only half of poor households (51%) are classified as poor, while for Malawi, these are more than four out of five poor households (82%). At 72%, 69%, and 60%, Nigeria, Tanzania, and Uganda fall in between. Interestingly, countries with a higher poverty rate also have larger shares of households that are correctly classified as poor.[13] Examining households classified as non-poor, I note that on average 79% of households classified as non-poor are actually also non-poor, while 21% of negatives are falsely so (for four countries, four out of five negatives are also non-poor, while for the fifth, Malawi, seven out of ten negatives are non-poor).

3.4.2. Benchmark: chance

To gain a better quantitative appreciation, I next compare the performance of PMT to that of a predictor oblivious to any information on household characteristics: tossing a coin. Obviously, even a coin tossed to predict whether a household

[11] Since I calibrate the PMT such that $P(+) = P(D)$, $P(+|D) = P(D|+)$ is the complement to the inclusion error rate, $P(\neg D|+)$.

[12] This is equal to the proportion of non-poor households that are correctly identified as non-poor, since $P(\neg D|-) = P(-|\neg D)$.

[13] This is consistent with a scenario in which poor households relatively close to the poverty line are less reliably classified as poor than poor households further below the poverty line: a larger share of households (further) below the poverty line in poorer countries would then lead to a higher $P(+|D)$. In section 3.4.6, I examine whether this is indeed the case.

Table 3.5: Classification Accuracy

	Ethiopia	Malawi	Nigeria	Tanzania	Uganda	avg
PMTs						
P(+|D)	0.51	0.82	0.72	0.69	0.60	0.67
P(¬D|-)	0.82	0.71	0.79	0.82	0.79	0.79
correctly classified	0.74	0.77	0.76	0.77	0.73	0.76
PSCs						
P(+|D)	0.45	0.80	0.72	0.69	0.63	0.66
P(¬D|-)	0.78	0.69	0.79	0.79	0.81	0.77
correctly classified	0.71	0.75	0.76	0.76	0.75	0.75
Random assign.						
P(+|D)	0.27	0.61	0.43	0.37	0.34	0.40
P(¬D|-)	0.73	0.39	0.57	0.63	0.66	0.60
correctly classified	0.61	0.52	0.51	0.53	0.55	0.55
% of HHs poor	0.266	0.609	0.427	0.371	0.343	0.403

is poor would by chance alone classify some households correctly. I hence compare $P(+|D)$ and $P(\neg D|-)$ (the PMT probabilities that a poor household is classified as poor and that a household predicted to be non-poor is also actually non-poor) to those of a random coin toss predictor with $P(+|D) = P(+|\neg D) = P(+)$ and $P(+) = P(D)$. This provides me with a lower bound benchmark against which I can judge the performance of PMT. The bottom panel of table 3.5 presents this benchmark.

Reassuringly, PMT performs substantially better than chance: across the five countries, a coin would correctly classify 40% of households as poor, while PMT does so for 67% of households (classifying 1.67 times as many households correctly as chance). Conversely, 60% of households classified as non-poor by chance also are non-poor, whereas 79% of households are so for PMT. Some variation exists between countries: when measured by $P(+|D)$, PMT performs 1.89 better than chance for Ethiopia (as chance performs relatively poorly there, given low poverty rates), while gains are smallest for Malawi, where PMT performs 34% better. Since higher poverty rates increase $P(+|D)$ for a random coin toss predictor, relative gains from PMT are largest in countries with low poverty rates (where chance performs poorly); conversely, when predicting $P(\neg D|-)$, chance performs worse in countries with higher poverty rates; hence the relative gains from PMT are largest there.

3.4.3. Benchmark: Brown

I also compare my results on classification performance to those of Brown et al. (2016). Since the inclusion and exclusion error rates of PMT (and thereby also $P(+|D)$ and $P(\neg D|-)$) reported by the authors vary widely across their different specifications, the choice of comparison is not obvious. However, as established in the previous section, estimated poverty rates also vary (and deviate from actual poverty rates) substantially, which mechanically drives inclusion and exclusion errors. Choosing the specification of Brown et al. (2016) in which estimated poverty rates match actual poverty rates most closely hence appears to be the most appropriate (and demanding) comparison. These are basic PMT for Nigeria, Tanzania, and Uganda, and Quantile Regression for Ethiopia, and Malawi, all with poverty lines at the 40th consumption percentile.

My results are relatively similar for Nigeria and Tanzania;[14] for Ethiopia, my specification appears to perform worse[15] as it also does to a smaller extent for Uganda.[16] Lastly, for Malawi, I find my specification to perform better.[17] Termed total error rate, Brown et al. (2016) also report (but do not seem to discuss) the classification performance of their models when setting $P(+) = P(D)$. My results appear relatively similar to theirs.[18]

My findings on the performance of PMT hence help to make sense of the wide

[14]For these countries actual poverty rates are close to the 40% that Brown et al. (2016) consider and the poverty rates predicted by the authors' regressions in turn match actual poverty rates closely.

[15]The authors report an EER of 29% and an IER of 44%, while I find 49%. These differences might stem from a smaller share of households below the poverty line in the specification I consider (at 27% relative to 40%) and from an overestimation of the share of poor households in the authors' specification by 10 percentage-points.

[16]Brown et al. (2016) report an EER of 29% and IER of 35% for Uganda, while I obtain 40%.

[17]The authors report an EER of 30% and an IER of 38%, while I find 18%; a substantially higher poverty rate in my specification compared to the 40% used by the authors might drive this finding.

[18]They find an IER=EER of 0.42, 0.36, 0.24, 0.31, and 0.34, for Ethiopia, Malawi, Nigeria, Tanzania, and Uganda, respectively.

range of results reported by Brown et al. (2016): pessimism about the usefulness of PMT stemming from the poor performance of a number of their specifications hence appears unjustified, as my analysis shows that these arise from severe mis-estimates of poverty rates, to which some specifications appear particularly prone. When choosing specifications that estimate poverty rates more accurately (or when calibrating estimated poverty rates to be equal to actual ones) PMT performs substantially better than in the worst specifications of Brown et al. (2016). Nevertheless, a candid appreciation of its limitations and still substantial shares of poor households classified as non-poor is also important.

3.4.4. Poverty Score Cards

Further gains in PMT classification performance might be possible when using more sophisticated approaches: a potential criticism of the proxy means tests that I have examined is that the choice of proxies is relatively ad-hoc; the chosen proxies might therefore not be the ones with the highest predictive power. An inferior performance might hence not stem from the methodology itself, but from a sub-par implementation, which does not use the most informative proxies. As discussed in section 3.2.4, Poverty Score Cards are an implementation of proxy means testing that specifically aims to use the proxies that have the largest explanatory power. I report my findings on their respective performance in the second panel of table 3.5: Poverty Score Cards perform somewhat worse for Ethiopia and somewhat better for Uganda, otherwise their performance is surprisingly similar to the PMT regressions I have been analysing. Poverty Score Cards with carefully chosen proxies hence do not appear to have greater predictive power than the PMT I have been considering; in other words, the ad-hoc choice of proxies in the previous PMT regressions does not appear to understate the potential of PMT. Nevertheless, Poverty Score Cards might have their practical advantages as they use fewer proxies and can be easily deployed in the field.

3.4.5. Robustness tests

In my analysis so far, I have taken two decisions relevant for the specification of my proxy means tests as outlined in sections 3.2.2 and 3.2.3. Firstly, I specified poverty lines such that the poverty headcount in my data matches the poverty headcount reported by the World Bank. Secondly, I chose the cut-off probability (above which households are predicted to be poor) such that the proportion of households predicted to be poor matched the proportion of households who are actually poor. In this section, I consider alternative specifications in turn.

Table 3.A2 presents the classification performance of logit model PMT and Poverty Score Cards when classifying all households with a predicted probability of poverty of 50% or greater as poor. As the table shows, logit model PMT and PSCs predict relatively similar proportions of households to be poor for four countries.[19] Predicted poverty rates (ignoring the PSC for Nigeria) also match actual poverty rates relatively closely for four countries, while for the fifth, Ethiopia, both approaches under-predict poverty rates by more than 50%. Compared to my main specification, this approach detects on average slightly fewer poor households; $P(+|D)$ is 62% (compared to 67% in my main specification), while a similar proportion of negative households are actually also non-poor (78% compared to 79%). As before, Poverty Score Cards perform worse than logit model PMT for all countries except Uganda. PSCs also perform worse compared to my main specification: as their performance is worse for Ethiopia and Nigeria, PSCs detect on average only 49% of poor households in this specification, while 66% are detected in my main specification. 72% of households predicted to be non-poor by PSCs are also non-poor, while 77% are in my main specification. As in previous specifications, I find that $P(+|D)$, the rate at which poor households are detected, increases with the poverty rates of countries.

In table 3.A3, I examine the classification performance when using a different approach to define which households live on less than \$1.90 per member and day. Instead of coding the poorest proportion of households as poor, such that poverty rates match the poverty rates reported by the World Bank, I first convert the \$1.90 poverty line into local currency units at the time of the survey (as explained in section 3.2.2) and then code households whose consumption per member is below this poverty line as poor. As previously discussed, these two approaches produce surprisingly large differences in poverty rates, most notably for Ethiopia (77%

[19]Nigeria is an exception; there the PSC predicts only 10% of households to live on less than \$1.90/day/member, while these are 44% according to World Bank Data.

vs. 27%) and Malawi (23% vs. 61%). Perhaps unsurprisingly, $P(+|D)$ in this specification differs substantially for countries that have different poverty rates. Overall, however, the proportion of poor households classified as poor, $P(+|D)$, increases with the poverty rate of countries, while $P(\neg D|-)$ decreases. Averaged across countries, $P(+|D)$ and $P(\neg D|-)$ are relatively similar (for both my PMT regressions and for PSCs) to my main specification, as is the average poverty rate. As before, PMT and PSCs also perform relatively similarly to each other for this specification.

3.4.6. How bad are the mistakes?

Apart from its frequency, the severity of classification mistakes also affects the performance of proxy means testing, and thereby the degree of confidence one might be willing to have in proxy means testing. In this section, I therefore explore *who* the false negative (and false positive) households are. To a policy-maker, classifying a poor household just below the poverty line as non-poor might be a more acceptable misclassification than classifying a household far below the poverty line as non-poor, especially since any poverty line will always be somewhat arbitrary and consumption data is likely to be subject to measurement error. Similarly, a household just above the poverty line classified as poor might not be deemed a severe inclusion error, whereas including a much richer household far above the poverty line is likely to be a graver inclusion error.

In their analysis, Brown et al. (2016) suggest that PMT particularly misses the poorest households: "The finding that errors tend to be higher using the lower poverty line again suggests that targeting may have difficulty in identifying those who are very poor;" "the point remains that PMT is missing many of the poorest households in all countries" (Brown et al., 2016, pp. 17, 18). However, this assertion appears to stem from PMT's worse performance for lower poverty lines rather than an analysis of classification performance by the degree of poverty (as measured by consumption).

To understand who the households likely to be missed by PMT are, I hence conduct such an analysis and examine the classification performance of PMT by consumption quantile. Figure 3.1 and table 3.6 present my results and cast light on

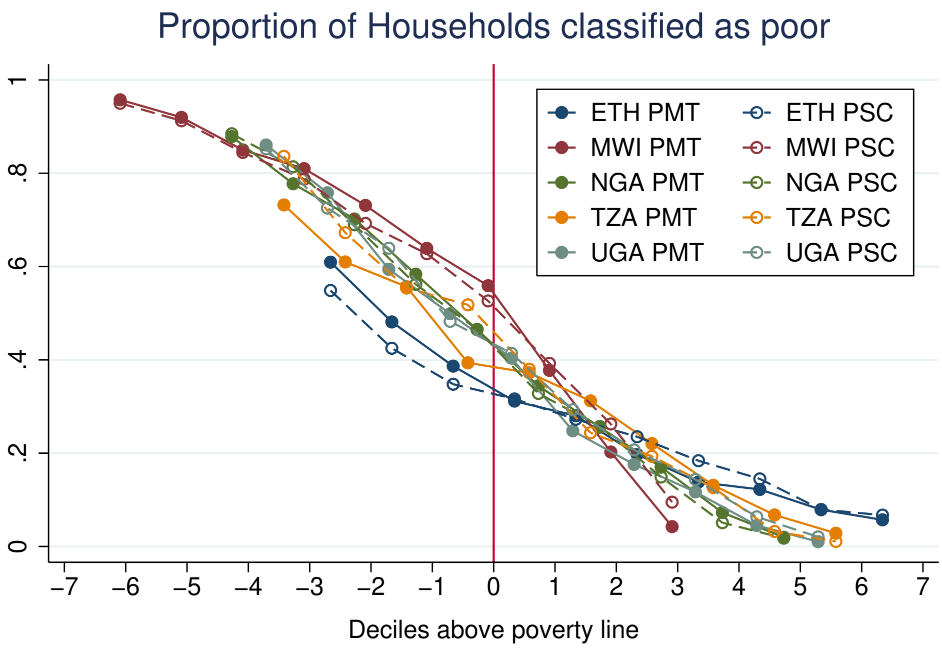


Figure 3.1: Probability of classifying household as poor by consumption decile

which households are likely to be falsely excluded or included by PMT.[20] Contrary to the assertions of Brown et al. (2016), I find that the poorer a household is, the more likely it also is to be identified as poor by proxy means testing.

Graph 3.1 presents this finding compellingly: grouping households by consumption, I show the share of households that are predicted to be poor for each decile.[21] Since poverty lines are not at the same consumption percentile across the countries, I arrange the consumption deciles in the graph as deciles above and below the poverty line to facilitate the comparison between countries. This graph hence allows me to study which households are likely to be falsely excluded and which non-poor households are likely to be falsely included by proxy means testing.

[20]Figures 3.A3-3.A7 in the appendix provide further detail, presenting the classification performance of PMT by consumption for each country.

[21]A perfect PMT would predict all households in deciles below the poverty line to be poor and no households in deciles above the poverty line to be poor (in the consumption decile containing the poverty line, it would only predict a share of households, those below the poverty line, to be poor).

The same finding emerges for all five countries and two versions of proxy means testing: the poorer a household is, the more likely it is to also be classified as poor by PMT (and thereby the less likely it is to be erroneously excluded). In consumption deciles below the poverty line, the share of households that are classified as poor increases as consumption decreases; proxy means testing correctly targets *particularly* the poorest households. Nevertheless, a considerable share of poor households still is classified as non-poor even in the poorest deciles. Similarly, in consumption deciles above the poverty line, the share of households that are classified as poor decreases as consumption increases; inclusion error rates are lower among richer households.

Table 3.6 non-graphically presents the same results, grouping households into quintiles rather than deciles: further below the poverty line, a larger share of poor households is correctly classified as poor, while further above the poverty line, a smaller share of non-poor households is erroneously classified as poor. Figures 3.A3-3.A7 in the appendix present this analysis separately for each country. I graph a histogram of consumption for each country, and display the poverty line and consumption deciles; in each consumption bin, I chart the number of households classified as poor (in blue) and classified as non-poor (in red); a perfect PMT would produce a histogram that is entirely blue to the left of the poverty line and entirely red to the right of the poverty line. Again, the analysis reveals that the proportion of households predicted to be poor increases with poverty (or rather, decreases with consumption); conversely, the proportion of households predicted to be poor increases with consumption. The most severe exclusion errors (excluding very poor households) tend to happen less frequently than lesser exclusion errors (excluding households only somewhat below the poverty line); similarly, more severe inclusion errors occur less frequently than lesser inclusion errors.

3.5. Summary

Poorer households generally live in worse circumstances, have fewer assets, and sometimes exhibit particular demographic characteristics. Proxy means testing (PMT) promises to exploit this relation between households' characteristics (the proxies) and households' consumption and poverty status to identify poor

Table 3.6: Classification Accuracy by Quintile

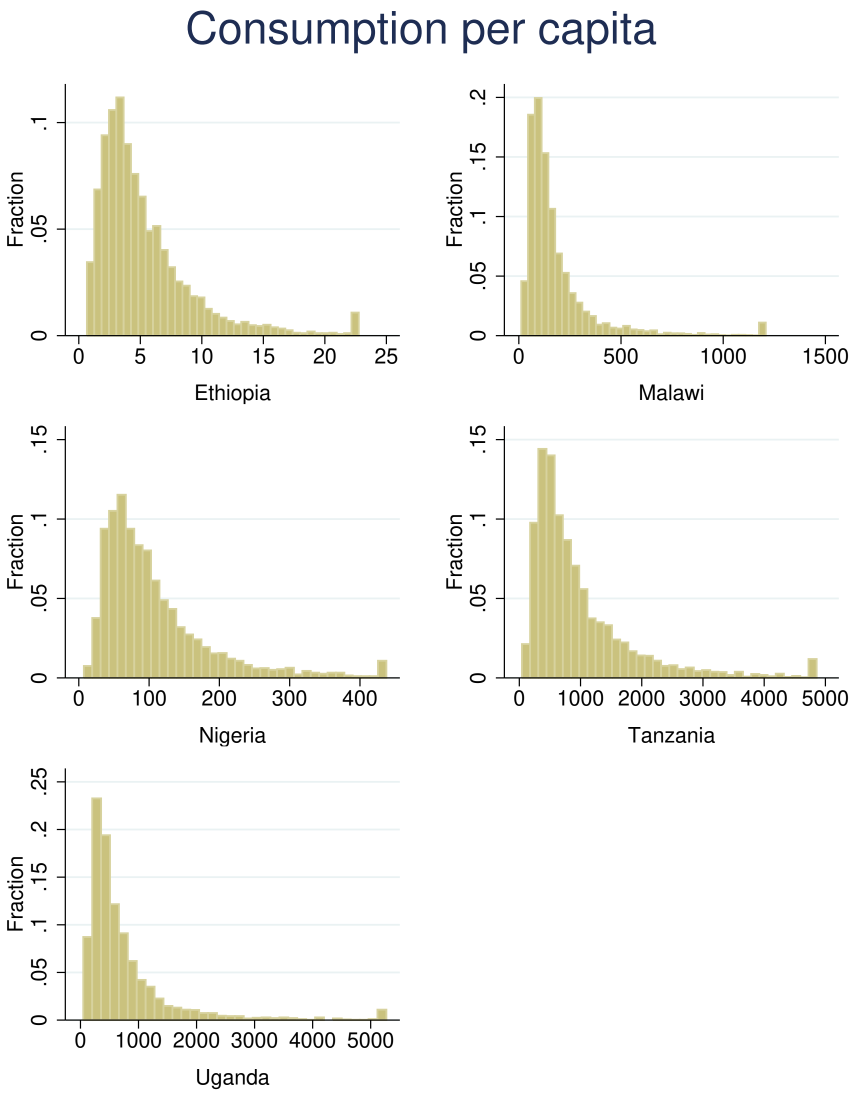
		Proxy Means Test		Poverty Score Card			
	P(D)	P(+|D)	P(+|¬D)	P(+|D)	P(+|¬D)	#D	#¬D
Ethiopia							
Bottom (Q1)	1.00	0.55		0.49		1014	0
20-40% (Q2)	0.33	0.42	0.31	0.36	0.32	336	678
40-60% (Q3)	0.00		0.24		0.25	0	1014
60-80% (Q4)	0.00		0.13		0.16	0	1014
Top (Q5)	0.00		0.07		0.07	0	1015
Malawi							
Bottom (Q1)	1.00	0.94		0.93		799	0
20-40% (Q2)	1.00	0.83		0.82		800	0
40-60% (Q3)	1.00	0.69		0.66		797	0
60-80% (Q4)	0.05	0.64	0.46	0.56	0.45	36	763
Top (Q5)	0.00		0.12		0.18	0	800
Nigeria							
Bottom (Q1)	1.00	0.83		0.85		907	0
20-40% (Q2)	1.00	0.64		0.63		907	0
40-60% (Q3)	0.13	0.54	0.38	0.54	0.37	122	785
60-80% (Q4)	0.00		0.21		0.20	0	907
Top (Q5)	0.00		0.05		0.03	0	908
Tanzania							
Bottom (Q1)	1.00	0.81		0.79		976	0
20-40% (Q2)	0.86	0.56	0.46	0.58	0.46	836	141
40-60% (Q3)	0.00		0.33		0.35	0	976
60-80% (Q4)	0.00		0.15		0.18	0	977
Top (Q5)	0.00		0.03		0.04	0	977
Uganda							
Bottom (Q1)	1.00	0.67		0.75		562	0
20-40% (Q2)	0.71	0.51	0.39	0.55	0.50	402	162
40-60% (Q3)	0.00		0.34		0.31	0	564
60-80% (Q4)	0.00		0.18		0.16	0	563
Top (Q5)	0.00		0.05		0.02	0	563

households from information on households that can easily be collected and verified through surveys. Recently, however, the performance of PMT has become debated: evaluating the accuracy of PMT in targeting the poorest 20% and 40% of households in 9 countries, Brown et al. (2016) find that PMT performs poorly, producing high rates of exclusion errors, as it wrongly predicts a large share of poor households to be non-poor.

I hence revisit PMT and examine the results of Brown et al. (2016) to shed light on the performance of proxy means testing: I find that poor calibration is a major driver of the weak performance of PMT in the analysis of Brown et al. (2016). When I calibrate the econometric model such that the predicted poverty rate matches the population's actual poverty rate, I find that PMT performs substantially better. Across the countries I examine, PMT on average correctly classifies between 6 and 7 of every 10 poor households, while chance would only correctly classify 4 out of 10 poor households. Yet, this also means that PMT misses approximately one third of poor households. Compared to perfect targeting this is a substantial gap. However, poorer households are less likely to be missed by PMT, while the households wrongly predicted to be poor do not tend to be among the richest households: inclusion errors rates decrease with consumption, while exclusion errors rates decrease with poverty (or a lack of consumption).

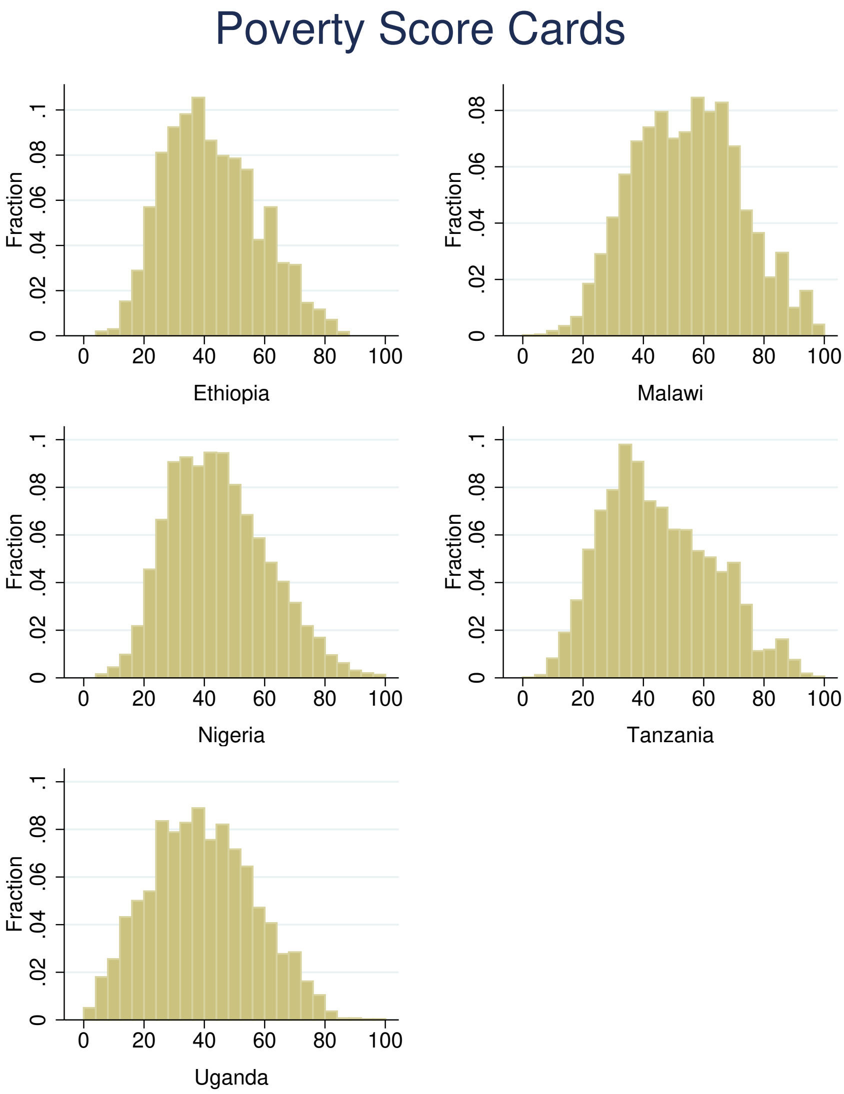
Any verdict on the performance of PMT will depend on one's benchmark: measurement error in consumption might contribute to its imperfect performance, but the nature of PMT itself is likely to be a key cause. Predicting household consumption (or poverty) from a small number of relatively slow-moving proxies is by nature a difficult endeavour and thus likely to produce imperfect results with substantial rates of both false negatives and false positives. PMT is hence a potentially useful but fallible heuristic that can provide an indication of whether a household *might* be poor; using only a small number of indicators, it is not a tool that can reliably and without errors classify households as above or below the poverty line. However, to users aware of its limitations and humble about the confidence of its predictions, PMT might still provide useful information about households' likely poverty status.

3.6. Appendix



Consumption per household member in thousands of local currency (winsorized at the 99th percentile).

Figure 3.A1: Descriptive Statistics: Consumption per capita



Distribution of PSC scores by country.

Figure 3.A2: Descriptive Statistics: Poverty Score Cards

Table 3.A1: Descriptive Statistics: Households

	Ethiopia	Malawi	Nigeria	Tanzania	Uganda
HH total consumption	5.42	188.42	110.98	1005.64	1378.22
Housing					
Toilet: pit	0.61	0.84	0.48	0.67	0.90
Toilet: flush	0.07	0.06	0.16	0.21	0.02
Floor: finished	0.04	0.35	0.03	0.46	0.01
Wall: finished	0.13	0.57	0.50	0.30	0.36
Roof: finished	0.04	0.00	0.07	0.71	0.00
Fuel: electric, gas, kerosene	0.07	0.04	0.23	0.64	0.01
Fuel: charcoal or coal	0.09	0.17	0.01	0.03	0.20
Household characteristics					
Household size	4.98	5.06	5.83	5.10	5.62
Head completed prim educ	0.30	0.36	0.58	0.62	0.41
Head completed secd educ	0.19	0.25	0.39	0.27	0.14
Female head	0.31	0.23	0.15	0.25	0.31
Head divorced or separated	0.10	0.09	0.03	0.11	0.10
Head widow	0.14	0.12	0.14	0.11	0.14
Head works for wage/salary	0.16	0.20	0.16	0.27	0.24
Head self-employed (non-ag)	0.14	0.18	0.38	0.25	0.26
Christian	0.68	0.82	0.54		
Muslim	0.29	0.14	0.44		
Urban	0.37	0.26	0.30	0.35	0.20
Household composition					
female, 0-5 years	0.07	0.09	0.06	0.08	0.09
male, 0-5 years	0.07	0.08	0.07	0.08	0.09
female, 6-14 years	0.10	0.11	0.11	0.09	0.13
male, 6-14 years	0.11	0.11	0.11	0.09	0.13
female, 65+ years	0.04	0.03	0.05	0.04	0.03
male, 65+ years	0.03	0.02	0.05	0.02	0.03
# of Households	5071	3995	4536	4883	2816

Household total consumption is annual per adult equivalent (except for Uganda, where it is per capita), in thousands of local currency, geo-temporally adjusted. Household size follows the definition of the LSMS surveys. All other indicators on household characteristics and housing are binary; reported summary statistics indicate the proportion of households with a given characteristic. Summary statistics on household composition state the proportion of household members belonging to each group.

Table 3.A2: Classification Accuracy (alternative decision rule)

	Ethiopia	Malawi	Nigeria	Tanzania	Uganda	avg
PMTs						
P(+|D)	0.29	0.87	0.74	0.66	0.54	0.62
P(¬D|-)	0.78	0.75	0.80	0.81	0.78	0.78
correctly classified	0.76	0.77	0.77	0.77	0.73	0.76
% predicted poor	0.13	0.67	0.44	0.35	0.30	0.38
PSCs						
P(+|D)	0.23	0.80	0.20	0.64	0.61	0.49
P(¬D|-)	0.74	0.69	0.61	0.77	0.80	0.72
correctly classified	0.73	0.75	0.64	0.76	0.75	0.73
% predicted poor	0.12	0.61	0.10	0.34	0.32	0.30
Random assign.						
P(+|D)	0.13	0.67	0.44	0.35	0.30	0.38
P(¬D|-)	0.73	0.39	0.57	0.63	0.66	0.60
correctly classified	0.67	0.54	0.51	0.54	0.56	0.56
% of HHs poor	0.266	0.609	0.427	0.371	0.343	0.403

Robustness test/alternative specification: HH predicted to be poor if Pr(poor) ≥ 0.5.

Table 3.A3: Classification Accuracy (alternative poverty rate)

	Ethiopia	Malawi	Nigeria	Tanzania	Uganda	avg
PMTs						
P(+|D)	0.86	0.55	0.65	0.68	0.70	0.69
P(¬D|-)	0.54	0.87	0.82	0.82	0.75	0.76
correctly classified	0.79	0.79	0.76	0.77	0.73	0.77
PSCs						
P(+|D)	0.83	0.60	0.66	0.67	0.73	0.70
P(¬D|-)	0.46	0.88	0.81	0.79	0.77	0.74
correctly classified	0.75	0.79	0.77	0.76	0.74	0.76
Random assign.						
P(+|D)	0.77	0.23	0.34	0.36	0.45	0.43
P(¬D|-)	0.23	0.77	0.66	0.64	0.55	0.57
correctly classified	0.64	0.64	0.55	0.54	0.51	0.58
% of HHs poor	0.766	0.231	0.339	0.360	0.449	0.429

Robustness test/alternative specification: poverty rate calculated from survey consumption data using PPP $1.90 poverty line.

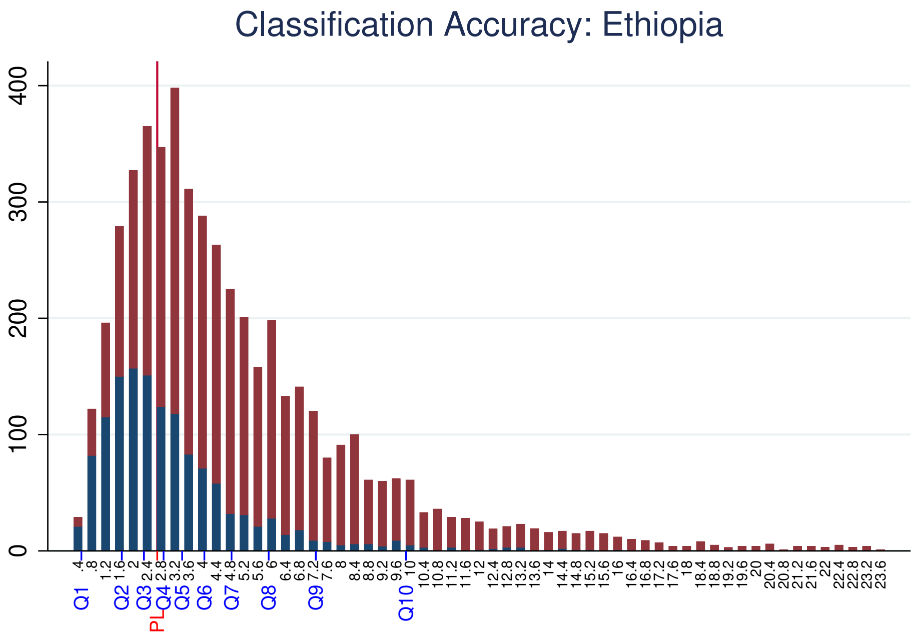


Figure 3.A3: PMT Classification Accuracy by Consumption: Ethiopia

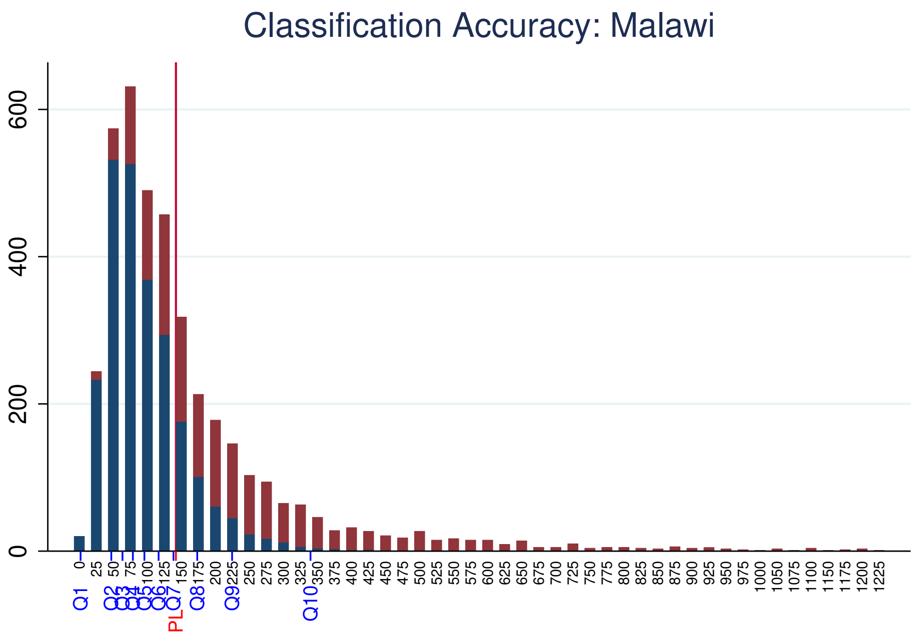


Figure 3.A4: PMT Classification Accuracy by Consumption: Malawi

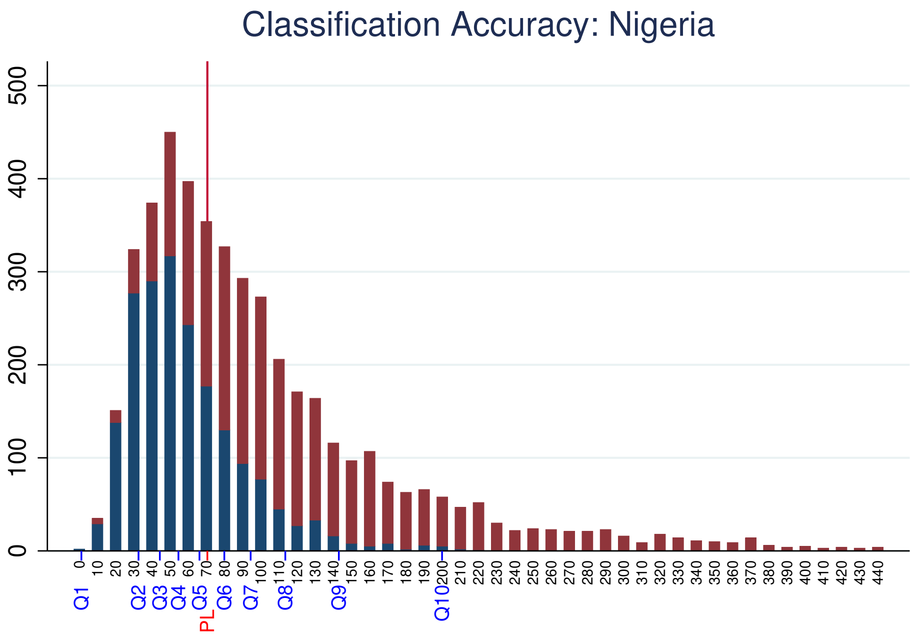


Figure 3.A5: PMT Classification Accuracy by Consumption: Nigeria

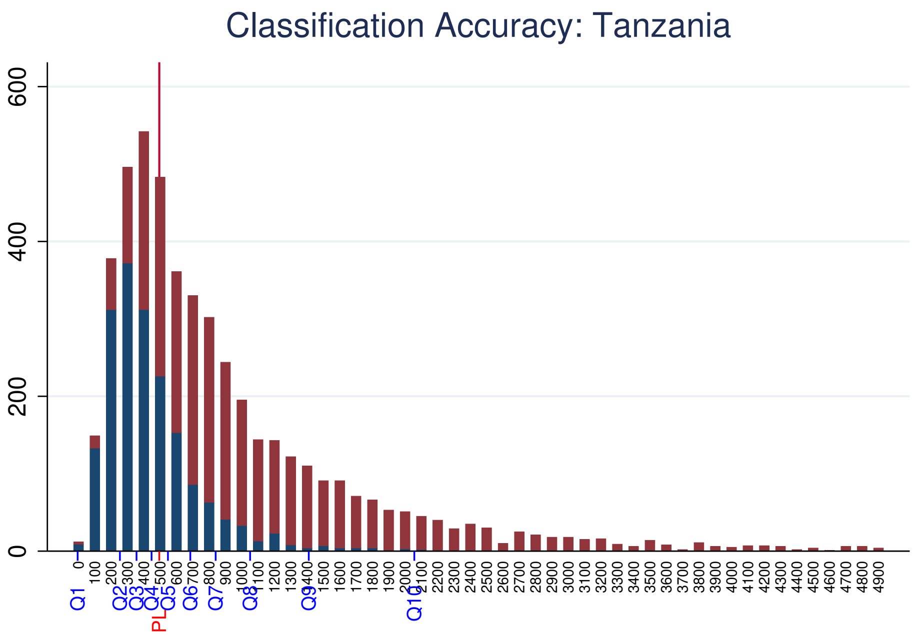


Figure 3.A6: PMT Classification Accuracy by Consumption: Tanzania

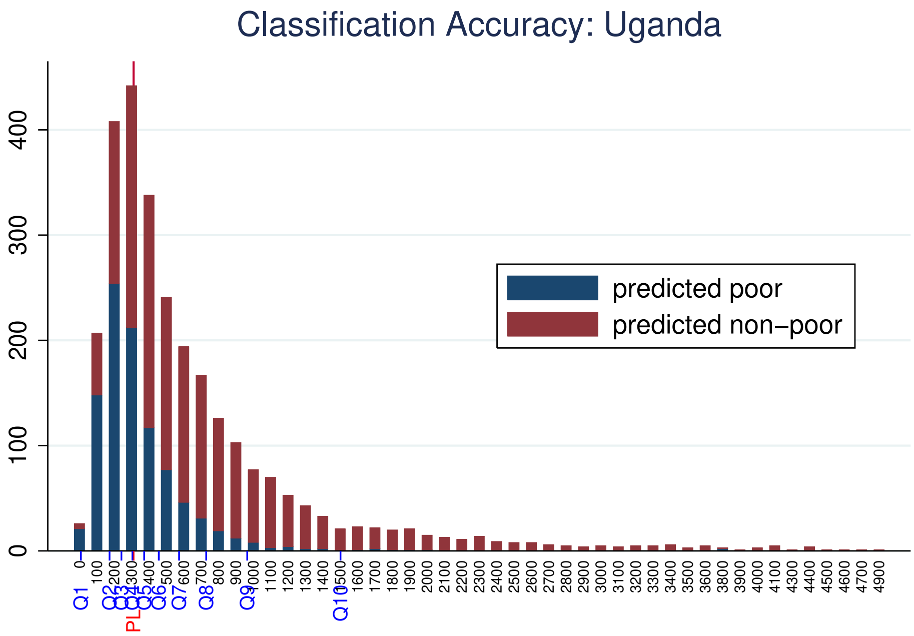


Figure 3.A7: PMT Classification Accuracy by Consumption: Uganda

Bibliography

Adu-Sarkodie, Y. and Steiner, M. (2000). Syndromic management of urethral discharge in Ghanaian pharmacies. *Sexually transmitted infections*, 76:439–442.

Alatas, V., Banerjee, A., Hanna, R., Olken, B. A., and Tobias, J. (2012). Targeting the Poor: Evidence from a Field Experiment in Indonesia Author. *American Economic Review*, 102(4):1206–1240.

Amin, S., Das, J., and Goldstein, M. (2008). *Are you being served?*

Banerjee, A. V., Chattopadhyay, R., Duflo, E., Keniston, D., and Singh, N. (2014). Improving Police Performance in Rajasthan, India: Experimental Evidence on Incentives, Managerial Autonomy and Training.

Beegle, K., De Weerdt, J., and Dercon, S. (2011). Migration and Economic Mobility in Tanzania: Evidence from a Tracking Survey. *Review of Economics and Statistics*, 93(3):1010–1033.

Beguería, S. and Vicente-Serrano, S. M. (2017). SPEI: Calculation of the Standardised Precipitation-Evapotranspiration Index.

Beguería, S., Vicente-Serrano, S. M., Reig, F., and Latorre, B. (2014). Standardized precipitation evapotranspiration index (SPEI) revisited: Parameter fitting, evapotranspiration models, tools, datasets and drought monitoring. *International Journal of Climatology*, 34(10):3001–3023.

Behrens, K., Duranton, G., and Robert-Nicoud, F. (2014). Productive Cities: Sorting, Selection, and Agglomeration. *Journal of Political Economy*, 122(3):507–553.

Bennett, D., Huang, C.-L., and Lauderdale, T.-L. (2014). Health Care Competition and Antibiotic Use in Taiwan.

Bennett, D. and Yin, W. (2014). The Market for High-Quality Medicine: Retail Chain Entry and Drug Quality in India.

Björkman-Nyqvist, M., Svensson, J., and Yanagizawa-Drott, D. (2016). Can Good Products Drive Out Bad? Evidence from Local Markets for Antimalarial Medicine in Uganda.

Black, R. E., Laxminarayan, R., Temmerman, M., and Walker, N., editors (2016). *Reproductive, Maternal, Newborn, and Child Health*, volume 2. World Bank Group, Washington DC, 3 edition.

Bloom, G., Standing, H., Lucas, H., Bhuiya, A., Oladepo, O., and Peters, D. H. (2011). Making health markets work better for poor people: the case of informal providers. *Health policy and planning*, 26 Suppl 1(Who 2007):i45–52.

Borjas, J. G. (1987). Self-Selection and the Earning of Immigrants. *American Economic Review*, 77(4):531–553.

Brown, C., Ravallion, M., and van de Walle, D. (2016). A Poor Means Test? Econometric Targeting in Africa.

Bryan, G., Chowdhury, S., and Mushfiq, A. (2014). Underinvestment in a Profitable Technology: The Case of Seasonal Migration in Bangladesh. *Econometrica*, 82(5):1671–1748.

Bryce, J., Victora, C. G., Habicht, J.-P., Black, R. E., and Scherpbier, R. W. (2005). Programmatic pathways to child survival: results of a multi-country evaluation of Integrated Management of Childhood Illness. *Health policy and planning*, 20 Suppl 1:i5–i17.

Buabeng, K. O., Duwiejua, M., Dodoo, A. N. O., Matowe, L. K., and Enlund, H. (2007). Self-reported use of anti-malarial drugs and health facility management of malaria in Ghana. *Malaria journal*, 6:85.

Buabeng, K. O., Matowe, L. K., Smith, F., Duwiejua, M., and Enlund, H. (2010). Knowledge of medicine outlets' staff and their practices for prevention and management of malaria in Ghana. *Pharmacy world & science : PWS*, 32(4):424–31.

Central Statistical Agency (Ethiopia) and The World Bank (2015). Construction of Consumption Aggregates for the Ethiopia Socioeconomic Survey Wave 2: Supplemental Information. Technical report.

Cnobloch, R. and Subbarao, K. (2015). Targeting Methods to Identify the Poorest in Malawi. In del Ninno, C. and Mills, B., editors, *Safety Nets in Africa: Effective Mechanisms to Reach the Poor and Most Vulnerable*, Africa Development Forum, pages 129–157. The World Bank.

Coady, D., Grosh, M., and Hoddinott, J. (2004). Targeting outcomes redux. *World Bank Research Observer*, 19(1):61–85.

Dalberg Global Development Advisors and the MIT-Zaragoza International Logistics Programme (2008). Private Sector Role in Health Supply Chains: Review of the Role and Potential for Private Sector Engagement in Developing Country Health Supply Chains. Technical Partner Paper 13. Technical report, Results for Development Institute, Washington DC.

Das, J. and Hammer, J. (2007). Money for nothing: The dire straits of medical practice in Delhi, India. *Journal of Development Economics*, 83(1):1–36.

Das, J. and Hammer, J. (2014). Quality of Primary Care in Low-Income Countries: Facts and Economics. *Annual Review of Economics*, 6(1).

Das, J., Holla, A., Das, V., Mohanan, M., Tabak, D., and Chan, B. (2012). In urban and rural India, a standardized patient study showed low levels of provider training and huge quality gaps. *Health Affairs*, 31(12):2774–84.

Deaton, A. and Paxson, C. (1998). Economies of Scale, Household Size, and the Demand for Food. *Journal of Political Economy*, 106(5):897.

Deaton, A. and Paxson, C. (2003). Engel's What? A Response to Gan and Vernon. *Journal of Political Economy*, 111(6):1378–1381.

del Ninno, C. and Mills, B., editors (2015). *Safety Nets in Africa: Effective Mechanisms to Reach the Poor and Most Vulnerable*. The World Bank, Washington DC.

Devereux, S., Masset, E., Sabates-Wheeler, R., Samson, M., Rivas, A. M., and te Lintelo, D. (2017). The targeting effectiveness of social transfers. *Journal of Development Effectiveness*, 9(2):162–211.

Diamond, A., Gill, M., Dellepiane, M. R., Skoufias, E., Vinha, K., and Xu, Y. (2016). Estimating Poverty Rates in Target Populations: An Assessment of the Simple Poverty Scorecard and Alternative Approaches.

Duranton, G. and Puga, D. (2004). Micro-Foundations of Urban Agglomeration Economies. *Handbook of Regional and Urban Economics*, 4(04):2063–2117.

Dustmann, C. and Glitz, A. (2011). Migration and Education. In *Handbook of the Economics of Education*, volume 4, pages 327–439. Elsevier B.V.

FAO (2017). GIEWS Country Brief United Republic of Tanzania. Technical report.

Ferreira, F., Chen, S., Dabalen, A., Dikhanov, Y. M., Hamadeh, N., Jolliffe, D. M., Narayan, A., Prydz, E. B., Revenga, A. L., Sangraula, P., Serajuddin, U., and Yoshida, N. (2015a). A Global Count of the Extreme Poor in 2012.

Ferreira, F., Jolliffe, D. M., and Prydz, E. B. (2015b). The international poverty line has just been raised to $1.90 a day, but global poverty is basically unchanged. How is that even possible? *The World Bank Let's Talk Development Blog*.

Fink, G., Weeks, J. R., and Hill, A. G. (2012). Income and health in Accra, Ghana: results from a time use and health study. *The American journal of tropical medicine and hygiene*, 87(4):608–15.

Fitzpatrick, A. (2016). Customer Information, Price, and Quality: Evidence from a Randomized Audit Study.

Friedman, W., Woodman, B., and Chatterji, M. (2015). Can mobile phone messages to drug sellers improve treatment of childhood diarrhoea?–A randomized controlled trial in Ghana. *Health Policy and Planning*, 30(suppl 1):i82–i92.

Gan, L. and Vernon, V. (2003). Testing the Barten Model of Economies of Scale in Household Consumption: Toward Resolving a Paradox of Deaton and Paxson. *Journal of Political Economy*, 111(6):1361–1377.

Geldsetzer, P., Williams, T. C., Kirolos, A., Mitchell, S., Ratcliffe, L. A., Kohli-Lynch, M. K., Bischoff, E. J. L., Cameron, S., and Campbell, H. (2014). The recognition of and care seeking behaviour for childhood illness in developing countries: A systematic review. *PLoS ONE*, 9(4).

Gollin, D., Lagakos, D., and Waugh, M. E. (2014). The Agricultural Productivity Gap. *Quarterly Journal of Economics*, 129(2):939–993.

Gommes, R. and Houssiau, M. (1982). Rainfall variability, types of growing seasons and cereal yields in Tanzania. In *Proc. of the Tech. Conf. on Climate in Africa*, pages 312–324. WMO/OMM Publication no. 596.

Gove, S. (1997). Integrated management of childhood illness by outpatient health workers: technical basis and overview. The WHO Working Group on Guidelines for Integrated Management of the Sick Child. *Bulletin of the World Health Organization*, 75 Suppl 1:7–24.

Grosh, M., del Ninno, C., Tesliuc, E., and Ouerghi, A. (2008). *For protection and promotion: the design and implementation of effective safety nets*.

Heltberg, R., Oviedo, A. M., and Talukdar, F. (2015). What do Household Surveys Really Tell Us about Risk, Shocks, and Risk Management in the Developing World? *Journal of Development Studies*, 51(3):209–225.

Hicks, J. H., Kleemans, M., Li, N. Y., and Miguel, E. (2017). Reevaluating Agricultural Productivity Gaps with Longitudinal Microdata.

Innovations for Poverty Action (2018). Poverty Probability Index.

Kubik, Z. and Maurel, M. (2016). Weather Shocks, Agricultural Production and Migration: Evidence from Tanzania. *Journal of Development Studies*, 52(5):665–680.

Kuznets, S. (1971). *Economic Growth of Nations: Total Output and Production Structure*. Harvard University Press, Cambridge, MA.

La Ferrara, E. and Harari, M. (2018). Conflict, Climate and Cells : A disaggregated analysis. *Review of Economics and Statistics*.

Lagakos, D. and Waugh, M. E. (2013). Selection, Agriculture, and Cross-Country Productivity Differences. *American Economic Review*, 103(2):948–980.

Leonard, K. L. and Masatu, M. C. (2005). The use of direct clinician observation and vignettes for health services quality evaluation in developing countries. *Social Science and Medicine*, 61(9):1944–1951.

Lewis, W. A. (1954). *The Theory of Economic Growth*. Irwin.

Liu, L., Johnson, H. L., Cousens, S., Perin, J., Scott, S., Lawn, J. E., Rudan, I., Campbell, H., Cibulskis, R., Li, M., Mathers, C., and Black, R. E. (2012). Global, regional, and national causes of child mortality: an updated systematic analysis for 2010 with time trends since 2000. *Lancet*, 379(9832):2151–61.

Mackintosh, M., Channon, A., Karan, A., Selvaraj, S., Zhao, H., Cavagnero, E., and Keynes, M. (2016). What is the private sector? Understanding private provision in the health systems of low-income and middle-income. *The Lancet*, 6736(16):1–10.

Madden, J. M., Quick, J. D., Ross-Degnan, D., and Kafle, K. K. (1997). Undercover careseekers: simulated clients in the study of health provider behavior in developing countries. *Social science & medicine*, 45(10):1465–82.

Makinen, M., Sealy, S., Bitran, R. A., Adjei, S., and Muñoz, R. (2011). *Private health sector assessment in Ghana*. World Bank Working Papers. The World Bank, Washington DC.

Makinen, M., Waters, H., Rauch, M., and Almagambetova, N. (2000). Inequalities in health care use and expenditures : Empirical data from eight developing countries and countries in transition. *Bulletin of the World Health Organization*, 78(1):55–65.

Mayhew, S. and Nzambi, K. (2001). Pharmacists' role in managing sexually transmitted infections: policy issues and options for Ghana. *Health policy and planning*, 16(2):152–160.

Mohanan, M., Vera-Hernández, M., Das, V., Giardili, S., Goldhaber-Fiebert, J. D., Rabin, T. L., Raj, S. S., Schwartz, J. I., and Seth, A. (2015). The Know-Do Gap in Quality of Health Care for Childhood Diarrhea and Pneumonia in Rural India. *JAMA pediatrics*, 27708(4):349–357.

Morgan, R., Ensor, T., and Waters, H. (2016). Performance of private sector health care: implications for universal health coverage. *The Lancet*, 6736(16):1–7.

Mueller, V., Gray, C., and Kosec, K. (2014). Heat stress increases long-term human migration in rural Pakistan. *Nature Climate Change*, 4(3):182–185.

National Statistical Office (Malawi) (2014). Integrated Household Panel Survey 2010-2013: Household Socio-Economic Characteristics Report. Technical Report November.

Nsimba, S. E. (2007). Assessing the impact of educational intervention for improving management of malaria and other childhood illnesses in Kibaha District-Tanzania. *East African Journal of Public Health*, 4(1):5–11.

Peters, D. H. and Bloom, G. (2012). Bring order to unregulated health markets. *Nature*, 487:163–165.

Pop, L. B. (2015). Options for Improving the Targeting of Safety Nets in Ghana. In del Ninno, C. and Mills, B., editors, *Safety Nets in Africa: Effective Mechanisms to Reach the Poor and Most Vulnerable*, Africa Development Forum, pages 67–106. The World Bank.

Raifman, J. R. G., Lanthorn, H. E., Rokicki, S., and Fink, G. (2014). The impact of text message reminders on adherence to antimalarial treatment in northern ghana: a randomized trial. *PloS one*, 9(10):e109032.

Ravallion, M., Chen, S., and Sangraula, P. (2009). Dollar a day revisited. *World Bank Economic Review*, 23(2):163–184.

Schreiner, M. (2015a). Simple Poverty Scorecard Malawi. Technical report.

Schreiner, M. (2015b). Simple Poverty Scorecard Nigeria. Technical report.

Schreiner, M. (2015c). Simple Poverty Scorecard Uganda. Technical report.

Schreiner, M. (2016a). Simple Poverty Scorecard Ethiopia. Technical report.

Schreiner, M. (2016b). Simple Poverty Scorecard Tanzania. Technical report.

Smith, F. (2009). The quality of private pharmacy services in low and middle-income countries: a systematic review. *Pharmacy world & science : PWS*, 31(3):351–61.

Stoeffler, Q., Mills, B., and del Ninno, C. (2016). Reaching the Poor: Cash Transfer Program Targeting in Cameroon. *World Development*, 83:244–263.

Stoeffler, Q., Nguetse-Tegoum, P., and Mills, B. (2015). Generating a System for Targeting Unconditional Cash Transfers in Cameroon. In del Ninno, C. and Mills, B., editors, *Safety Nets in Africa: Effective Mechanisms to Reach the Poor and Most Vulnerable*, Africa Development Forum, pages 39–66. The World Bank.

Sudhinaraset, M., Ingram, M., Lofthouse, H., and Montagu, D. (2013). What is the role of informal healthcare providers in developing countries? A systematic review. *PloS one*, 8(2):e54978.

Tawfik, Y., Nsungwa-Sabitii, J., Greer, G., Owor, J., Kesande, R., and Prysor-Jones, S. (2006). Negotiating improved case management of childhood illness with formal and informal private practitioners in Uganda. *Tropical medicine & international health : TM & IH*, 11(6):967–73.

The World Bank (2018). PovcalNet: an online analysis tool for global poverty monitoring.

Tulloch, J. (1999). Integrated approach to child health in developing countries. *The Lancet*, 354:SII16–SII20.

Vicente-Serrano, S. M., Beguería, S., and López-Moreno, J. I. (2010). A multiscalar drought index sensitive to global warming: The standardized precipitation evapotranspiration index. *Journal of Climate*, 23(7):1696–1718.

Wafula, F. N. and Goodman, C. A. (2010). Are interventions for improving the quality of services provided by specialized drug shops effective in sub-Saharan Africa? A systematic review of the literature. *International journal for quality in health care : journal of the International Society for Quality in Health Care / ISQua*, 22(4):316–23.

Wafula, F. N., Miriti, E. M., and Goodman, C. a. (2012). Examining characteristics, knowledge and regulatory practices of specialized drug shops in Sub-Saharan Africa: a systematic review of the literature. *BMC health services research*, 12(1):223.

Waters, H., Hatt, L., and Peters, D. H. (2003). Working with the private sector for child health. *Health Policy and Planning*, 18(2):127–137.

Watson, M., Norris, P., and Granas, A. (2006). A systematic review of the use of simulated patients and pharmacy practice research. *International Journal of Pharmacy Practice*, 14(2):83–93.

WHO (2014). *Integrated Management of Childhood Illnesses (IMCI) Chart Booklet.* Geneva, Switzerland.

Yohan, I. B. B., Oteng'i, S. B. B., and Lukorito, C. B. (2006). Assessment of the Growing Season over the Unimodal Rainfall Regime Region in Tanzania. *Tanzania Journal of Agricultural Sciences*, 7(1):16–26.

Young, A. (2013). Inequality, the Urban-Rural Gap, and Migration. *Quarterly Journal of Economics*, pages 1727–1785.