INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE

IFPRI

IFPRI Discussion Paper 02256

June 2024

A Network-Driven Data Collection Approach for Agri-Food Value Chains

Kate Ambler

Jeffrey R. Bloem

Alan de Brauw

Sylvan Herskowitz

Julia Wagner

Markets, Trade, and Institutions Unit

INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE

The International Food Policy Research Institute (IFPRI), a CGIAR Research Center established in 1975, provides research-based policy solutions to sustainably reduce poverty and end hunger and malnutrition. IFPRI's strategic research aims to foster a climate-resilient and sustainable food supply; promote healthy diets and nutrition for all; build inclusive and efficient markets, trade systems, and food industries; transform agricultural and rural economies; and strengthen institutions and governance. Gender is integrated in all the Institute's work. Partnerships, communications, capacity strengthening, and data and knowledge management are essential components to translate IFPRI's research from action to impact. The Institute's regional and country programs play a critical role in responding to demand for food policy research and in delivering holistic support for country-led development. IFPRI collaborates with partners around the world.

AUTHORS

Kate Ambler (k.ambler@cgiar.org) is a Senior Research Fellow in the Markets, Trade, and Institutions Unit at the International Food Policy Research Institute (IFPRI), Washington, DC.

Jeffrey R. Bloem (j.r.bloem@cgiar.org) is a Research Fellow in the Markets, Trade, and Institutions Unit at the International Food Policy Research Institute (IFPRI), Washington, DC.

Alan de Brauw (a.debrauw@cgiar.org) is a Senior Research Fellow in the Markets, Trade, and Institutions Unit at the International Food Policy Research Institute (IFPRI), Washington, DC.

Sylvan Herskowitz (sherskowitz@worldbank.org) is an Economist in the World Bank's Development Impact Department (DIME), Washington, DC.

Julia Wagner (j.wagner@cgiar.org) is a Research Analyst in the Markets, Trade, and Institutions Unit at the International Food Policy Research Institute (IFPRI), Washington, DC.

Notices

ABSTRACT

A key challenge in systematically collecting data on intermediary agri-food value chain actors is that value chains take the form of a network, with actors linked by a series of transactions. Moreover, we have limited *ex ante* knowledge about the structure or scale of these networks, which complicates the construction of valid sampling frames and limits traditional random sampling approaches to collect data. To address these challenges, we adapt the respondent-driven sampling approach to collect data on intermediary agri-food value chain actors within their transaction-linked network and implement this approach in the arabica coffee and soybean value chains in Uganda and the rice and potato value chains in Bangladesh. We observe meaningful heterogeneity in the structure and scale of agri-food value chains across commodities and countries. Focusing on traders, we show that the respondent-driven sampling approach generates a larger sample of traders who differ in observable characteristics (i.e., value added, enterprise scale, and financial access) compared to a sub-sample of traders generated in a way that mimics traditional random sampling approaches used to study traders. We conclude by discussing how this respondent-driven sampling approach, applied within transaction-linked networks, can provide a useful data collection method for studying intermediary agri-food value chain actors.

Keywords: Agri-food value chains, respondent-driven sampling, data collection.

ACKNOWLEDGMENTS

The authors are listed in alphabetical order. We are grateful to Seth Garz, Maria May, Hope Michelson, Bart Minten, Marc Bellemare, and Rich Sexton for useful comments. We are thankful for generous support from the Bill and Melinda Gates Foundation in funding this research. This work was also undertaken as part of the CGIAR Research Initiative on Rethinking Food Markets, and so we thank all funders who supported the CGIAR Trust Fund. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Food Policy Research Institute, the CGIAR, the World Bank, their affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent. All errors are our own.

1 Introduction

Existing research in agricultural and development economics focuses almost exclusively on producers and consumers. Classic dual sector models of agricultural development and structural transformation abstract away from the intermediary roles of agribusinesses and assume that primary producers directly supply consumers through complete and competitive market interactions (Barrett, Reardon, Swinnen and Zilberman, 2022). Additionally, when empirical research does consider intermediate agri-food value chain actors, it often focuses on a specially targeted subset of actors within a given value chain, limiting our knowledge of how these actors interact with each other and conduct their businesses (Bellemare, Bloem and Lim, 2022). The relative neglect of the intermediary segments of agri-food value chains is striking given the importance of intermediary actors in adding value to and transforming agricultural produce into healthy, reliable, and safe food for consumers.[1]

An existing qualitative literature describes value chains for specific commodities (Ayele, Duncan, Larbi and Khanh, 2012; Stür, Khanh and Duncan, 2013; Kilelu, Klerkx and Leeuwis, 2013; Beshah, Kitaw and Dejene, 2013; Vicol, Neilson, Hartatri and Cooper, 2018; Soullier and Moustier, 2021) and tends to describe complex relationships within specific value chains. Importantly, this literature does not generalize by segment, where we consider a segment to be a group of actors who complete specific tasks (i.e., collecting, processing, aggregating, or selling to consumers). Moreover, different types of actors within those complex descriptions of agri-food value chains often have overlapping tasks. As a result, while this qualitative literature is useful in describing the complex structure of agri-food value chains, it does not

[1] Recent research suggests that intermediary actors—those operating between the farmgate and retailers—account for a substantial share of value-added produced by the entire agricultural sector of an economy. Reardon (2015) suggests the value-added attributable to intermediary agri-food value chain actors might be as large as 40 percent and Barrett *et al.* (2022) note that these figures vary across countries and over time and typically range from 66 to 75 percent. Yi, Meemken, Mazariegos-Anastassiou, Liu, Kim, Gómez, Canning and Barrett (2021) use data from 61 countries between the years of 2005 and 2015 and find, on average, value-added attributable to intermediary agri-food value chain actors is 73 percent.

enable us to learn about the relative importance of specific types of actors (i.e., in terms of the volume of product they move or the value they add to the product as it moves through the value chain). Nor is it possible to generalize either across agri-food value chains for different commodities within a country or across different countries, due to the complex nature of many agri-food value chains.

When researchers design surveys and data collection plans to quantitatively study agri-food value chains, they generally adopt a simplified structure, with a number of mutually exclusive intermediary steps between producers and consumers (UNCTAD, 2009; FAO, 2014, 2015). Although this simplification can, in principal, allow us to compare actors across value chains, and even across countries, it leads to at least two fundamental problems. First, some actors are particularly mobile (especially traders) and their business is seasonal, so they may be hard to find. Second, the total number of actors within a given segment defined by the simplified structure of a particular agri-food value chain is unknown. As a result, developing a sampling frame for specific types of value chain actors is immediately a challenging task. Given this challenge, many existing efforts to study the intermediary segments of agri-food value chains do not attempt to provide representative samples of intermediary actors, within their transaction-linked networks, with sufficient sample sizes or the ability to disaggregate by actor or commodity (Barrett *et al.*, 2022; Ambler, de Brauw, Herskowitz and Pulido, 2023). These conditions combine to lead to a widespread lack of internationally comparable data with sufficient sample sizes on agri-food value chain actors, limiting research on the role of intermediary actors in agricultural development and structural transformation (Reardon and Timmer, 2007; Reardon, 2015; Yi *et al.*, 2021; Ambler *et al.*, 2023; Barrett *et al.*, 2022; Bellemare *et al.*, 2022).

In this paper, we develop and implement a novel sampling approach to collect data on these intermediary actors within agri-food value chains. Specifically, we adapt a sampling method called "respondent-driven sampling," an approach used in sociology to study hard-to-find populations (Heckathorn, 1997; Heckathorn and Cameron, 2017).[2] We begin with the notion that value chains take the form of a network, with actors linked by transactions but due to our limited *ex ante* knowledge about the structure or scale of these networks we refrain from making strong assumptions about the structure of these value chains. Instead, we follow transaction links as referrals to develop the sample. The referrals allow us to identify with whom specific enterprises transact, and include well-connected and less-well-connected enterprises within value chains in our sample. We use the information on linkages within the data set to generate sampling weights that adjust and account for convenience sampling bias, allowing us to generate descriptive statistics about specific types of intermediary actors within value chains that are more likely to be representative of each type of actor. We apply this method to four distinct agri-food value chains in two countries: the arabica coffee and soybean value chains in Uganda and the rice and potato value chains in Bangladesh.

Our effort builds on several different methods used to study aspects of agri-food value chains in low and middle income countries. Many studies in this literature conduct case studies that analyze a mix of both qualitative and quantitative data involving a small number of intermediary actors often from just one segment within the value chain (Gow, Streeter and Swinnen, 2000; Minten, Randrianarison and Swinnen, 2009; Dries, Germenji, Noev and Swinnen, 2009; Minten, Vandeplas and Swinnen, 2012; Minten, Assefa and Hirvonen, 2017). While these studies make important contributions in their own right, they do not collect or use data with sufficient sample sizes to disaggregate by actor type and commodity.

A second method randomly samples actors within multiple segments of agri-food value chains. Reardon, Liverpool-Tasie and Minten (2021) review the literature implementing surveys using this approach, what the authors call "stacked surveys," identifying 37 unique surveys across 41 studies. These stacked surveys rely on *ex ante* knowledge about both the

[2] Respondent-driven sampling is similar, yet importantly distinct from "snowball sampling" methods; see, Heckathorn (2011) for a detailed discussion of the differences.

structure of a given value chain and where to find intermediary value chain actors to create a sampling frame, and thus the reliability of data from these surveys ultimately rests on the accuracy of this existing knowledge. The *ex ante* knowledge used to develop a sampling frame is often limited. For example, Fafchamps, Hill and Minten (2008) aimed to collect data on representative samples of growers, traders, processors, and wholesalers operating within several agri-food value chains in India. The authors randomly sample growers from villages and traders from markets, but encountered difficulties in constructing a reliable sampling frame when sampling processors and wholesalers. These difficulties ultimately prevent them from conducting their survey with processors and wholesalers.

Finally, several impact evaluations examine outcomes among intermediate agri-food value chain actors when exposed to a given policy or institutional intervention. However, these studies tend to focus on actors within one segment of the agri-food value chain such as traders (Bergquist and Dinerstein, 2020; Jensen, 2007; Aker, 2010; Goyal, 2010; Casaburi and Reed, 2022; Wiseman, 2023), market retailers (Banerjee, Fischer, Karlan, Lowe and Roth, 2023), cooperatives (Casaburi and Macchiavello, 2019), or processors (Macchiavello and Morjaria, 2021). A notable exception is Macchiavello and Miquel-Florensa (2019) who use panel data on coffee farmers and detailed transaction-level data on all coffee sales to investigate how a quality upgrading scheme, implemented by a large multi-national buyer, influences farmer and intermediary actor welfare along the coffee value chain in Colombia. Even this exceptional data, however, does not allow for disaggregation between various types of intermediary actors.

We make three main contributions to the literature. First, we develop and implement a method of adapting respondent-driven sampling to agri-food value chains. This sampling and data collection approach is well-suited for settings in which a well-defined sampling frame either is difficult to construct or does not exist, and where respondents are linked together through a social or economic network. In particular, respondent-driven sampling allows respondents themselves to inform the path of the interviews within their transaction-linked networks and provides researchers with information that can be used to generate sampling weights adjusting and accounting for convenience sampling bias. A fundamental contribution within our application of the respondent-driven sampling approach to agri-food value chains is our construction of a novel sampling weight—the segment-adjusted multiplicity weight—which draws on insights from both respondent-driven sampling (Heckathorn, 1997; Heckathorn and Cameron, 2017) and "multiplicity sampling" methods (Rothbart, Fine and Sudman, 1982; Sirken, 1970). This re-weighting approach is akin to the methods of Hsieh, Hsu, Ko, Kovárık and Logan (2024), who focus on correcting non-representative network data in a regression context. Fundamentally, our application of the respondent-driven sampling approach to the arabica coffee and soybean value chains in Uganda and the rice and potato value chains in Bangladesh offers a useful proof-of-concept and motivates its use in future research.

Second, consistent with existing qualitative descriptions of agri-food value chains, we document meaningful heterogeneity in the structure and scale across each of the four agri-food value chains we study. Much of the existing quantitative literature, however, characterize agri-food value chains as a simplified set of links where producers sell to traders, who sell to processors, who sell to wholesalers, who ultimately sell to retailers and then end consumers.[3] Our data—which explicitly relies on transaction links to generate the sample—reveals that agri-food value chains can be much more complex than this simple characterization. In some cases, it is difficult to observe an empirical basis for the simple characterization found in the quantitative literature, with intermediary actors simultaneously using multiple sources for buying and selling a given commodity. Additionally, we explore heterogeneity within the set of traders in our data and find that buying patterns reveal consistent variation along observ-

[3] We do note that the literature does suggest some of these roles can be combined to simplify the agri-food value chain, and that longer chains is a sign of economic development (Barrett *et al.*, 2022).

able characteristics. Moreover, we estimate the average length of each agri-food value chain, defined by the number of transactions needed for a commodity to move from the primary producer to a given end point. We again document meaningful heterogeneity in the length of agri-food value chains—both within and across commodities. These findings highlight the *ex ante* challenge in characterizing the structure and scale of agri-food value chains.

Finally, focusing on traders, we document differences in the data generated by the respondent-driven sampling approach compared with a data generating process that mimics the traditional "stacked sample" approach of randomly sampling traders from local retail markets. We observe meaningful and statistically significant differences between the full sample of traders and the sub-sample that either buy or sell a given commodity at local retail markets. Specifically, we show that the respondent-driven sampling approach generates a much larger sample of traders relative to a sub-sample of market-based traders. In particular, with the respondent-driven sampling approach, we find 1,400 coffee traders, 507 soybean traders, 1,066 rice traders, and 1,117 potato traders. By contrast, with the sub-sample of market-based traders, we find 242 coffee traders, 87 soybean traders, 375 rice traders, and 299 potato traders. Additionally, the sub-sample of market-based traders differs from the full sample in observable characteristics such as estimated value added, enterprise scale, and a measure of financial access. These findings suggest that the data generated from traditional random sampling approaches commonly used in the literature to collect data among intermediary agri-food value chain actors (Reardon *et al.*, 2021) might lead to an incomplete understanding about the characteristics, activities, and financial needs of traders. Our data illustrate that if we had randomly sampled traders from local retail markets we could have generated data indicating higher rates of employment and larger sales volumes among coffee and soybean traders in Uganda, larger sales volumes among rice traders in Bangladesh, and lower sales volumes among potato traders in Bangladesh.

The remainder of this paper is organized as follows. In the next section we introduce the adapted respondent-driven sampling approach and define the sampling weights we develop to account for convenience sampling bias. Section three discusses characteristics of our sample and the geography of our study locations. Section four reports our main findings relating to the structure of agri-food value chains, estimated value-chain length, and compares data on traders generated with the respondent-driven sampling approach and a data generating process that mimics the random sampling approach traditionally used to study traders. Finally, section five concludes with a discussion.

2 Our Sampling Approach

A fundamental challenge to research on intermediary actors is that it is difficult to study them with publicly available data. By design, large, publicly available data sources lack information on the types of enterprises that work in agri-food value chains. For example, the World Bank's LSMS data sets only include information on household-based enterprises, which tend to be small and informal, and information on commercial enterprises is limited. Moreover, they lack data on transactions that facilitate any agricultural trading conducted by these households. Meanwhile, the World Bank Enterprise Surveys use a sampling frame based on registered and formal enterprises with five or more employees, and thus lack information on unregistered, informal enterprises; these data are therefore skewed toward larger enterprises with formal accounting structures in urban areas. As the majority of food in low and middle income countries is thought to pass through micro, small, and medium enterprises, neither of these surveys are well-suited for the study of the intermediary segments of agri-food value chains (HLPE, 2017; Liverpool-Tasie, Wineman, Young, Tambo, Vargas, Reardon, Adjognon, Porciello, Gathoni, Bizikova *et al.*, 2020).[4]

[4]Some researchers have used aggregated national accounts data in the form of input-output or supply-and-use tables combined with partial or general equilibrium models to study agri-food value chains (Breisinger, Thomas and Thurlow, 2009; Canning, Weersink and Kelly, 2016; Yi *et al.*, 2021), but neither allow for disaggregation by actor or commodity, nor can these methods examine how intermediary actors transact

The goal of our sampling approach is to enable the estimation of representative statistics about intermediary actors within agri-food value chains. Intermediary agri-food value chain actors are often highly mobile and relatively informal, limiting our *ex ante* knowledge about the structure and scale of these networks. As a result, standard random sampling methods are difficult to implement, as one does not have a stationary population from which to sample. Moreover, we want to learn about the structure of the value chains we study. We use the fact that agri-food value chains take the form of a network, with actors at various stages of the value chain connecting with each other through buying and selling transactions. So even if we could develop sampling frames for each type of value chain actor, we would miss the linkages between different types of actors. To use the network structure in sampling and overcome these challenges, we draw on methodological insights developed by sociologists that allow for the calculation of key population parameters within a network-based sampling approach (Heckathorn, 1997; Heckathorn and Cameron, 2017). The resulting respondent-driven sampling technique allows respondents to inform the path of the interview process within their transaction-linked networks and researchers to generate sampling weights enabling credible estimation of key population parameters.

Our data collection approach follows an iterative process designed to capture information about intermediary actors within both multiple segments of an agri-food value chain and their transaction-linked network. This process refrains from making strong assumptions about the structure of the agri-food value chain. That is, a simple characterization of an agri-food value chain defines segments of value-addition within the agri-food value chain: (i) farmers, (ii) traders, (iii) processors, (iv) wholesalers, and (v) retailers. Although this simple characterization of an agri-food value chain can approximate the structure of a real-life agri-food value chain, in many cases these segments may be combined within one firm (i.e., a firm that both processes and sells wholesale), or there may be additional linkages to other firms

with one another.

(i.e., transporters or storage facilities), and still finally there may be linkages between actors within the same segment of the value chain. Each of these realistic departures change the structure and scale of an agri-food value chain and complicate traditional random sampling approaches to data collection that require a valid sampling frame to generate reliable data. Our goal, therefore, is to design a data collection process flexible enough to incorporate heterogeneity in the structure and scale of agri-food value chains across commodities and countries, but is also sufficiently standardized as to allow for the collection of comparable information across commodities and countries.

The iterative data collection process starts with selecting a small number of farmers in randomly selected villages within each of the districts selected as our primary study sites, as we will discuss in Section 3. These farmers represent the "seeds" for the respondent-driven sampling approach. We interview farmers with a short survey, ask them to whom they sold the given commodity (i.e., arabica coffee or soybeans in Uganda and rice or potatoes in Bangladesh) in the last 30 days, and record the name and phone number of these selling links. Next, we contact the intermediary actors identified by the farmers and interview them with our full intermediary actor survey. We again record the name and phone number of these selling links, and interview the identified actors if they have not yet been interviewed. This iterative process continues until we either achieve "saturation" whereby referred intermediary value chain actors are already surveyed and included in our data or we are referred to intermediary value chain actors that are outside the geographic scope of our survey.

2.1 Sampling Weights

The respondent-driven sampling approach allows the respondents themselves to direct the path of the interviews within their transaction-linked networks and enables us, the researchers, to calculate sampling weights that adjust and account for convenience sampling bias. We define these sampling weights by first describing the traditional "multiplicity weight" and then adding a refinement to better fit our adaptation of respondent-driven sampling to agri-food value chains. The refinement leads us to define a novel sampling weight, which we call the "segment-adjusted multiplicity weight."

Drawing from the literature on respondent-driven sampling (Heckathorn, 1997; Heckathorn and Cameron, 2017), we generate sampling weights to facilitate the estimation of key population parameters. A fundamental feature of a sampling weight within the context of our respondent-driven sampling approach is to correct for the probability that each actor is selected into our sample. All else equal, we want to assign a larger weight to those actors who are less likely to be selected into our sample, and lower weights when actors are highly likely to be selected into the sample. For example, in this context, we can imagine that we might not capture all of the traders who only buy and sell small quantities or with few people, but larger traders who work with many people in the value chain would have almost certainly been captured in the sample. These sampling weights allow us to make statements about the population of intermediary actors within a given commodity's value chain that buy and sell the commodity produced in our study regions.

To construct our sampling weights, we start by following "multiplicity sampling" methods developed by Sirken (1970) and Rothbart *et al.* (1982). Multiplicity sampling differs from conventional survey methods in that respondents are selected into the sample via referrals from other respondents within the sample rather than through a specific probabilistic sampling design. The multiplicity weight adjusts for the probability that any given intermediary agri-food value chain actor is included in the sample, and is defined as follows:

$$MW_i = \frac{1}{B_i} \tag{1}$$

Where for each intermediary agri-food value chain actor, i, with some number of buying links, B_i, the multiplicity weight, MW_i, is the reciprocal of the number of buying links.

For example, consider a trader who reports selling coffee to three processors (selling links) and purchasing coffee from two farmers (buying links). This trader will have a multiplicity weight equal to 1/2. The intuition for this approach is that intermediary actors with more buying links have a greater likelihood for inclusion in our sample because they have more links pointing toward them as we proceed through our iterative data collection process and are thus more likely to be referred into our sample. The calculation of the multiplicity weight follows a similar motivation to a sampling weight in a randomly sampled household survey where the sampling weight is the reciprocal of the probability that a household is selected from the sampling frame and observations with a relatively high probability of being selected into the sample are given a relatively low sampling weight.

An important departure of our application of respondent-driven sampling from its standard applications is that agri-food value chains include several distinct segments (i.e., traders, processors, and wholesalers), with each segment engaging in complementary roles that together transform a raw agricultural commodity into food marketed to consumers. As we move along the value chain, intermediary actors within each segment may interact more or less directly with primary producers (i.e., the farmers that represent the "seeds" of our respondent-driven sampling approach). Given that it is through referrals—starting with farmers and progressing through multiple value chain segments—that our sample is constructed, we need to adjust for these features in the construction of our preferred sampling weight. Thus, we adjust the multiplicity weight by the share of actors within each segment of the value chain that purchase directly from farmers. The intuition for this adjustment is that actors purchasing directly from farmers are more likely to be referred into our sample, given that farmers represent the "seeds" for our respondent-driven sampling approach. Intermediary actors that do not purchase directly from farmers are referred into our sample via another intermediary actor. This segment-adjusted multiplicity weight is defined as follows:

$$SAMW_{si} = \frac{1}{B_i} \times \frac{1}{F_s} \tag{2}$$

Where the inverse of the number of buying links (i.e., the standard multiplicity weight as defined in equation (1) above) is multiplied by the inverse of the share of actors within a value chain segment, s, that report purchasing directly from farmers, F_s. This equation implies that, all else equal, the larger the share of actors within a given stage that purchases directly from farmers, the smaller the segment-adjusted multiplicity weight.

3 Data and Sample Characteristics

We implement our adaptation of the respondent-driven sampling approach in four different agri-food value chains in two countries: arabica coffee and soybeans in Uganda and rice and potatoes in Bangladesh. In Uganda, we chose to work within the arabica coffee and soybean value chains to represent one relatively established, organized value chain (i.e., coffee) and one less well organized but rapidly growing value chain (i.e., soybeans). In Bangladesh, we chose to work within the potato value chain to represent a rapidly growing chain, and the rice value chain as it is historically important for food security with some government involvement in the value chain organization. Together, these agri-food value chains provide a heterogeneous set of case studies in our implementation of the respondent-driven sampling approach.

In Uganda data collection took place in April and May of 2023, in Mbale district on the east side of the country near Mount Elgon, Kasese district on the west side of the country near the Rwenzori mountains, and Lira district in the north. Both Mbale and Kasese districts represent prominent geographical areas for the production, trading, and processing of arabica coffee. Lira district is one of the most prominent production and trading areas for soy in Uganda. In Bangladesh data collection took place between June and August 2023, in Bogra and Rangpur districts, both located in the northern region. Bogra and Rangpur districts represent a large share of both the local and hybrid varieties of potato production in Bangladesh. Rice is produced throughout Bangladesh and, therefore, we can survey intermediary actors in rice production in both Bogra and Rangpur districts. Importantly, both districts are well known as rice surplus districts (Ahmed and Bakhtiar, 2020).

In Tables 1 and 2 we present a series of summary statistics describing the sample composition and helping to contextualize the scale and demographic characteristics of intermediary actors in both Uganda and Bangladesh, respectively. We also provide information on the selling links reported by each respondent, to summarize how the iterative data collection process worked in practice. We disaggregate each statistic by commodity and by the primary value chain segment self-identified by the respondent (i.e. trader, processor, or wholesaler). There are many ways to define these roles, and indeed the boundaries between a "trader" and a "wholesaler" can sometimes be opaque. Thus, how respondents themselves characterize their primary role of value addition within a given value chain carries meaning and represents our method of defining the roles of intermediary actors within our data.

Our sample predominantly consists of traders (i.e., 72.5 percent of the sample in Uganda and 70.6 percent of the sample in Bangladesh). Our sampling approach yielded an insufficient number of soybean processors in Uganda and potato processors in Bangladesh, as the majority of processing for these commodities takes place outside of the geographic scope of our data collection. Thus, for soybeans and potatoes, we only disaggregate by trader and wholesaler. Other roles within the value chain include storage facility owners and aggregators, but these roles are found so infrequently in our data that we leave them out of this analysis.

Panel A in Tables 1 and 2 summarize the scale of the enterprises between each segment of each value chain. In Uganda, our data collection occurred during a relative low-point in the seasonal cycle, with between 80 and 96 percent of intermediary actors reporting that the amount they bought and sold in the last 30 days is less than the usual amount they bought and sold per month over the last year. In Bangladesh, our data collection occurred closer to the high-point in the seasonal cycle, with only between 37 and 56 percent of intermediary actors reporting that the amount they bought and sold in the last 30 days is less than the usual amount they bought and sold per month over the last year. In Panel B of Tables 1 and 2, we report typical costs associated with operating intermediate agri-food value chain enterprises. Notably, in both Uganda and Bangladesh, transportation costs (inclusive of payments for renting vehicles) represents the largest cost for intermediary actors.

Demographic characteristics for respondents appear in Panel C of Tables 1 and 2. We discuss four categories of descriptive statistics. In both Uganda and Bangladesh, the average intermediary actor in our data is between the age of 34 and 40. Respondents are also overwhelmingly male; in Uganda, between 80 and 90 percent within each category are male, while virtually all the actors interviewed in Bangladesh are male. Second, although education attainment varies across intermediary actors, the majority of intermediary actors in both Uganda and Bangladesh have only attained a secondary level of education. Third, in Uganda, although between 43 and 54 percent of coffee traders, coffee wholesalers, and soybean wholesalers have electricity at their home, roughly 80 percent of coffee processors and soybean wholesalers have electricity at their home. By contrast, in Bangladesh, nearly all intermediary actors have electricity in their home.

We finally document information about selling links provided by each respondent in our data. These details are crucial for the implementation of the respondent-driven sampling approach to work as designed. Reassuringly, most respondents reported selling link information and, on average, respondents reported between 1.5 and 2.5 selling links in Uganda and

3.0 and 3.5 in Bangladesh. In both Uganda and Bangladesh, the overwhelming majority of reported links came with an active phone number and roughly half of all recorded links were referred by more than one person (i.e., duplicate links). These descriptive statistics about selling links highlight that although our data collection did not achieve full "saturation" we do find a meaningful share of duplicate links. This finding motivates the use of sampling weights to correct for bias that might arise within this application of respondent-driven sampling.

4 Main Findings

We present three main findings. First, we discuss data characterizing the structure of each of the four agri-food value chains in our data. This analysis includes illustrations of buying and selling patterns reported by intermediary actors and also an exploration of heterogeneity within the sample of traders in our data. Second, leveraging information on transaction links, we document heterogeneity in value chain length (i.e., defined as the average number of transactions we observe between the primary producer and a given value chain segment in our data). Finally, focusing on traders, we estimate differences in observable characteristics between our full data set generated with respondent-driven sampling and a data generating process that mimics the traditional "stacked sample" approach of randomly sampling traders from local retail markets.

4.1 Agri-Food Value Chain Structure

Analysis with quantitative data often characterize agri-food value chains as a simple chain linking actors operating in distinct segments between producers and consumers—see, e.g., UNCTAD (2009) and FAO (2014, 2015). Our data makes clear that this characterization oversimplifies the buying and selling patterns in agri-food value chains, which are reported by actor type in Figures 1 through 4 for all four value chains in Uganda and Bangladesh. In each figure, Panel (a) uses an indicator variable for whether the respondent reported buying from given sources in the previous 30 days, and Panel (b) uses an indicator variable for whether the respondent reported selling to given buyers in the previous 30 days. These figures demonstrate complex buying and selling patterns, highlighting important heterogeneity in the structure of agri-food value chains across commodities and countries.

In the arabica coffee value chain in Uganda, Panel (a) in Figure 1 shows that traders, processors, and wholesalers all most frequently purchase from farmers. However, we do observe that traders are a close second most common buying source among processors and wholesalers, demonstrating that traders sometimes play an intermediary role between farmers and key segments of value addition within the chain. Moreover, roughly 30 percent of traders report buying coffee from other traders. Finally, less than 20 percent of all traders, processors, and wholesalers within the arabica coffee value chain buy from local retail markets.

In panel (b) of Figure 1, we observe that local retail markets represent a relatively infrequent selling source, with less than 10 percent of coffee traders and less than five percent of coffee processors selling to local retail markets. Instead, coffee traders most frequently sell to other traders, coffee processors most frequently sell to either traders or other processors, and coffee wholesalers most frequently sell to processors. Again, these findings contrast with a simple characterization of an agri-food value chain in which a farmer sells to a trader, who then sells to a processor, who finally sells to a wholesaler.

In the soybean value chain, Panel (a) of Figure 2 illustrates that traders and wholesalers both report farmers as their most common buying source. As with coffee, despite the difference in the level of value chain organization, we observe that traders are the second most common buying source among wholesalers. In contrast to the arabica coffee value chain, however, only 15 percent of soybean traders buy from other soybean traders. As with arabica coffee, less than 10 percent of all soybean traders and wholesalers buy from local retail markets. Panel (b) of Figure 2 shows that traders are the most common selling source for both soybean traders and soybean wholesalers. Wholesalers are a close second most common selling source for soybean traders. And finally, less than 10 percent of soybean traders and wholesalers sell to local retail markets.

Turning to Bangladesh, Panel (a) of Figure 3 shows that buying patterns in the rice value chain are somewhat more suggestive of the simplified characterization of an agri-food value chain. Rice traders most frequently purchase from farmers, rice processors most frequently purchase from traders, and wholesalers most frequently purchase from processors. However, the simple characterization of the value chain ignores the 40 percent of rice processors who buy directly from farmers, and the nearly 50 percent of rice wholesalers who buy from rice traders.

In Panel (b), we observe that rice traders and processors both sell to local retail markets relatively infrequently, with less than 20 percent of traders and less than 30 percent of processors selling to local retail markets. Rice wholesalers, however, most commonly sell to local retail markets. Among rice traders, other traders are the most common selling source with processors a close second most common selling source. Rice processors most commonly sell to wholesalers. These observations largely mirror the buying patterns reported in Panel (a) and again provide a rough empirical basis for a simple characterization of the rice value chain in Bangladesh—albeit one that abstracts away from incorporating the diversity of selling sources among intermediary actors.

Finally, the potato value chain also roughly conforms with the simple characterization of agri-food value chains, as shown in Panel (a) of Figure 4. Traders most frequently purchase from farmers, and wholesalers most frequently purchase from traders. However, over 30 percent of potato traders buy from other traders, again highlighting important heterogeneity in the role played by traders. Moreover, nearly 40 percent of potato wholesalers buy directly from farmers. As with rice, local retail markets appear to represent a relatively infrequent buying source within the potato value chain. Panel (b) of Figure 4 shows that traders are the most common selling source for both potato traders and potato wholesalers. Wholesalers are a close second most common selling source for potato traders. Finally, although less than 20 percent of potato traders sell to local retail markets, roughly 50 percent of potato wholesalers sell to local retail markets.

Taken together, Figures 1 through 4 highlight three findings. First, we can broadly state that the simple characterization of agri-food value chains, often assumed by the quantitative literature, can mask the complexity of transactions taking place within agri-food value chains often documented by the qualitative literature. Although we observe some empirical justification for the simplified characterization within the rice value chain in Bangladesh, that characterization ignores the diversity of buying and selling sources among intermediary actors within that chain. Second, we observe many cases of within-segment buying and selling. For example, within all four value chains we observe a meaningful share of traders buying from and selling to other traders. This highlights important heterogeneity in the role played by traders. Finally, with the exception of rice and potato wholesalers in Bangladesh, local retail markets represent a strikingly infrequent buying or selling source among intermediary actors in our data.

We next explore whether respondents who call themselves "traders" may include multiple distinct types of intermediary actors. We do this by investigating the apparent heterogeneity in our sample of traders by testing for differences along observable characteristics based on whether the trader reports buying from another trader. While this is a relatively coarse dimension of heterogeneity, as shown in Figures 1 through 4, the share of traders who buy from other traders range from between 15 percent (among soybean traders) to over 30 percent (among coffee, rice, and potato traders). Tables 3 and 4 show heterogeneity among coffee, soybean, rice, and potato traders in our data and reports means for a list of observable characteristics about the traders. Each table also reports the differences in these means and indicates whether this difference is statistically significant, and a consistent pattern emerges. Among coffee and soybean traders in Uganda and among rice and potato traders in Bangladesh a simple indicator of whether the trader buys from other traders generally predicts that the trader runs a larger business, receives higher prices, and is more likely to use available financial services.

Among coffee and soybean traders in Uganda, Table 3 shows that traders who buy from other traders appear to operate larger businesses in that they receive higher prices, are more likely to have employees, report purchasing more (soybean traders) or selling more (coffee traders), and pay more rent for both structures and vehicles. Additionally, both coffee and soybean traders who buy from other traders are more likely to use financial services in that they are less likely to buy and sell with cash and are more likely to use a financial account. Moreover, coffee traders are more likely to be able to access loans when needed and use insurance to manage risk.

Similar patterns exist among rice and potato traders in Bangladesh. Table 4 shows that traders who buy from other traders appear to operate larger businesses in that they are more likely to have employees, report buying and selling higher quantities, and pay more rent for structures and vehicles (rice traders). Additionally, rice and potato traders who buy from other traders are more likely to use financial services in that they are less likely to buy (potato traders) and sell with cash, are more likely to use a financial account. Moreover, rice traders are more likely to access loans when needed and use insurance to manage risk while rice traders receive higher prices.

4.2 Agri-Food Value Chain Links and Length

Using the information we collect on transaction links, we now estimate the average length of each of the agri-food value chains in our data. This analysis provides additional insight into the structure and scale of the agri-food value chains by specifically investigating the average number of transactions that take place between a farmer and a given value chain segment identified via our sampling approach.

To determine the typical length and composition of the four agri-food value chains, we first create a database containing all possible chains for each farmer by implementing the following procedure. First, we connect each interviewed farmer with their selling links—the actors to whom they sold the commodity of interest (i.e., arabica coffee or soybeans in Uganda and rice or potatoes in Bangladesh) in the last 30 days. Second, we check whether the intermediary actors identified by the farmers were also interviewed.[5] Third, if these intermediary actors were also interviewed, we identify each of their selling links and repeat the process. We continue until we reach an endpoint. In this exercise, we define an endpoint as any intermediary actor with whom we did not conduct a subsequent interview. We purposely exclude any bi-directional links, such as two intermediaries naming each other as selling links within a chain, to avoid generating endless loops in our database.[6] The resulting database describes how products might flow through connections between intermediary actors within a given agri-food value chain.

It is important to emphasize that our link database does not necessarily represent the actual number of transactions that a single unit of the specified commodity goes through on its way from the primary producer to a given endpoint. In particular, because our data collection process follows sales links rather than a single unit of the commodity, some chains in our database involve a greater number of actors than might be realistically expected. Thus, to obtain more accurate estimates of the average length of each agri-food value chain in our data, we randomly select one chain for each farmer and repeat this process 1,000

[5]This identification is based on respondents' phone numbers, so for this exercise we only consider transaction links that came with an active phone number. This leads us to slightly underestimate the length of value chains in our data because we cannot follow links that we could not contact via phone.

[6]These excluded bi-directional links represent 2 percent of the database in Uganda and 12 percent in Bangladesh. Bi-directional links can take multiple forms. For example, it could be actor A sells to actor B and actor B sells to actor A. Or it could take a more complex form where actor A sells to actor B, actor B sells to actor C, and actor C sells to actor A.

times. This simulation exercise allows us to estimate a distribution of possible pathways through our link database, with more frequent pathways representing more "realistic" paths a commodity might take.

Another important detail of our iterative data collection process, and therefore our link database, is where the value chains end. Because wholesalers represent the final level of our data collection, we concluded all value chains where a wholesaler was reached. In addition, endpoints may occur because the actor didn't provide any further selling links, or the referred actor was previously referred by a different respondent, refused to participate in the survey, was no longer in business, was unreachable, or was located outside the districts selected as our primary study sites. While in most instances our iterative data collection process ends "naturally" (i.e., because a referred actor was previously referred by a different respondent or the referred actor was a retailer), there are some cases where a referred actor was not interviewed because we could not connect with them or they were out of geographic scope.[7] This suggests that our estimates of value chain length in our data likely represent underestimates. To account for this systematic underestimation, we apply an attrition correction to the following categories of value chains in our link database. For chains that "end" because the link could not be reached or we could not connect with them for other reasons, we multiply the value chain length estimate by the inverse of the following ratio:

$$\text{Attrition Ratio} = \frac{\text{No. interviewed}}{\text{No. interviewed} + \text{No. not reached} + \text{No. not interviewed}} \tag{3}$$

The ratio represents the proportion we expect to reach, so the correction applied is the inverse. In Bangladesh, in addition to the main interviews, we followed a portion of the links located outside the two main districts through a phone survey. Thus, for these chains in

[7] As noted above, our geographic scopes for these surveys are defined as follows: In Uganda, we surveyed arabica coffee value chain actors in Mbale and Kasese districts and we surveyed soybean value chain actors in Lira district. In Bangladesh, we surveyed both rice and potato value chain actors in Bogra and Rangpur districts.

Bangladesh, we multiply the value chain length estimate by the inverse of the attrition ratio for phone surveys.[8]

Table 5 reports summary statistics on the average number of transactions required for a commodity to move through the intermediary actors within our data, and the types of actors involved in these agri-food chains. We find meaningful heterogeneity in the length of agri-food value chains in our data. The average length of the agri-food value chains in our data is roughly 1.8 transactions for coffee, 1.4 transactions for soybeans, 3.2 transactions for rice, and 2.8 transactions for potatoes. For both Bangladesh and Uganda, the mean length of the value chains is shorter when the end point is a trader compared to when it is a wholesaler.[9] The longest chains in our data are observed in the rice segment in Bangladesh, where it takes an average of 3.2 transactions for rice to move from the primary producer to a wholesaler. Our estimations also provide insights into the average composition of these agri-food value chains. Across all four commodities, traders emerge as the most prevalent actors. In Uganda, arabica coffee and soybeans pass through one trader, on average, whereas these commodities are much less likely to pass through processors and wholesalers. In Bangladesh, the same pattern holds. Rice and potatoes pass through two traders, on average, and are less likely to pass through processors and wholesalers. These results demonstrate that there is meaningful heterogeneity in the length of agri-food value chains, both within and across commodities. Notably, the simplification of a value chain where a farmer sells to a trader, who sells to a processor, who sells to a wholesaler does not, in general, appear to be the typical route a commodity travels through the segments of the value chains in our data. This underscores the challenge in characterizing the structure and scale of agri-food value

[8] Based on these proportions, the value chain length estimates in Bangladesh are multiplied by 1.47 (=1/0.68) for the in-person survey in the rice value chain, by 1.39 (=1/0.72) for the in-person survey in the potato value chain, and by 1.27 (=1/0.79) for the phone survey in Bangladesh (for both the rice and potato value chains). In Uganda, the value chain length estimates are multiplied by 1.69 (=1/0.59) in the coffee value chain and by 1.30 (=1/0.77) in the soybean value chain.

[9] The missing estimate of the average length with endpoint processor for the soybean value chain is due to the fact that our sampling approach yielded an insufficient number of soybean processors in Uganda.

chains *ex ante* and motivates the use of our adaptation of the respondent-driven sampling approach.

4.3 Respondent-Driven vs. Random Sampling

We now ask how our data might have looked if, instead of implementing the approach inspired by respondent-driven sampling, we had implemented a more traditional "stacked sampling" approach. As defined by Reardon *et al.* (2021), a stacked sample is a set of surveys conducted within representative samples of actors in each segment of an agri-food value chain. To generate these representative samples, researchers apply *ex ante* knowledge about the structure and scale of a particular agri-food value chain to construct a sampling frame.[10]

Within agri-food value chains, the challenge of constructing a valid sampling frame is perhaps most salient among traders, who are highly mobile, often operate informal enterprises, and might engage in different agri-food value chains depending on the season and commodity-specific market conditions. Of the studies reviewed by Reardon *et al.* (2021), only a subset collect information and report statistics on traders. Nearly all of these studies rely on relatively small samples of traders, with fewer than 500 traders and often with fewer than 100 traders in the data. For example, Fafchamps *et al.* (2008) report on 400 fruit and vegetable traders in India, Minten, Tamru, Engida and Kuma (2016) survey 204 teff traders in Ethiopia, and Minten, Murshid and Reardon (2013) and Reardon, Chen, Minten and Adriano (2012) use the same survey to study 60 rice traders and 60 potato traders, respectively, in Bangladesh. A notable exception is Liverpool-Tasie, Reardon, Sanou, Ogunleye,

[10]As noted in the introduction, Fafchamps *et al.* (2008) document this challenge in their effort to generate representative samples of growers, traders, processors, and wholesalers operating within several agri-food value chains in India. While the authors were able to randomly sample growers from villages and traders from markets, the authors ultimately could not construct "a reliable sampling frame" and struggled "getting selected enterprises to respond to the questionnaire" when sampling processors and wholesalers. These challenges prevented them from conducting their survey with processors and wholesalers.

Ogunbayo and Omonona (2017), who implement an extensive data collection effort by randomly sampling maize traders from over 65 local retail markets across Nigeria and surveyed 1,406 traders. Despite these differences in sample size, each of these studies use the same sampling approach of randomly sampling traders from local retail markets.

The approach of randomly sampling traders from local retail markets is a sensible way to address the challenge of finding traders to survey. Local retail markets represent a defined geographic area where traders can conduct their business and are likely to be serving farmers from the surrounding area. Our data show, however, that although local retail markets are one possible source for traders to buy and sell a given commodity, it is not the only buying or selling source. As shown in Figures 1 through 4, among traders local retail markets are never the most common source for either buying or selling, and a substantial share of traders do not report buying (i.e., 80-90 percent) or selling (i.e., 85-95 percent) at all in local retail markets. Therefore, sampling at local retail markets could miss all of these traders.

We assess the difference in several observable characteristics between our full sample of traders generated with our respondent-driven sampling approach and the sub-sample of market-based traders that mimics the traditional "stacked sample" method of randomly sampling traders from local retail markets. We report these results in Table 6. By definition, a sub-sample can be no larger than the full sample. However, it is notable how much smaller the sub-sample of market-based traders is relative to the full sample of traders. With the respondent-driven sampling approach, we find 1,400 coffee traders, 507 soybean traders, 1,066 rice traders, and 1,117 potato traders. With the sub-sample of traders that either buy or sell at local retail markets, we find 242 coffee traders, 87 soybean traders, 375 rice traders, and 299 potato traders. The implication is that if we had instead implemented the traditional "stacked sample" approach and randomly sampled traders from local retail markets, we could have only captured between 17 and 35 percent of the active traders within each given value chain.[11]

Table 6 also reports the mean and standard deviation of value added (i.e., the difference between average selling price and average buying price), an indicator variable indicating if the respondent reports employing anyone in the last 30 days, the amount sold in the last 30 days (i.e. reported in 1,000 kgs), and an indicator variable indicating if the respondent reports using an account at a financial institution. We observe meaningful differences between the full sample of traders and the subset who never trade at local retail markets. The estimated value added is nearly one-third lower and over two-thirds lower in the market-based sub-sample than in the full sample of traders in the arabica coffee and soybean value chains, respectively. In Bangladesh, though, the difference is smaller for the rice value chain, and among potato traders the value add is slightly higher in the market-based subset. In Uganda, we find that in both value chains the probability of employing someone is 23 and 15 percentage points higher in the market-based sub-sample than in the full sample; probabilities are small and not meaningfully different in the two value chains in Bangladesh. In three of the chains we find that market-based traders report selling more than the overall average, but we find the opposite for potatoes. Finally, although the probability of using an account at a financial institution is slightly higher in the market-based sub-sample than in the full sample of traders in the arabica coffee and rice value chains, it is slightly lower in the soybean and potato value chains.

The statistics reported in Table 6 represent the statistics observed in our data and differences between the full sample and the market-based sub-sample could be due to inherent differences between these groups or due to sampling variation. To account for bias driven

[11] Here we assume we would have gone to all the markets that the sample of traders visit. Had we gone to a larger set of markets, we would have been able to increase the sample size, but would still have missed the majority of traders who do not buy or sell at local retail markets. Additionally, this might also depend on whether traders who buy from other traders are not making these transactions at local markets. Even if this does occur frequently, the traditional "stacked sample" approach would lead to the inclusion of a sub-set of the traders included in our respondent-driven sampling approach.

by sampling variation, we perform a bootstrapping exercise. We first take our data and re-sample the data with replacement. Next, we calculate the mean of each of the four observable characteristics reported in Table 6 in the full re-sampled data and in the market-based sub-sample of the re-sampled data. Finally, we calculate the difference between these two means. We repeat this three-step process 10,000 times and plot these differences in Figures 5 through 8. These figures plot the distribution of the difference in means, along with the associated 90 percent confidence interval (i.e., solid lines) and the mean of the distribution (i.e., dotted line). Cases in which zero does not fall between the 5th and 95th percentile is akin to estimating the statistical significance of the estimated difference in means with a bootstrapped standard error, at the 95 percent level.

Across the four value chains, we find four differences that appear different from zero. In the arabica coffee value chain, we find the difference in means between the full sample and the stacked survey sub-sample are not likely to be zero for the employment and the amount sold variables (Figure 5). In the soybean value chain the difference is not zero for the employment variable (Figure 6). There are fewer clearly significant differences in the rice and potato value chains; we find no clear significant differences in the rice value chain (Figure 7), and only the amount sold variable appears to have a mean different from zero for the potato value chain (Figure 8). That said, given the low overall levels of employment among traders in Bangladesh, for at least that variable the null finding is not surprising. In sum, although this bootstrapping exercise represents a relatively conservative assessment of statistical significance, we observe several instances where the difference in means between the full sample generated with respondent-driven sampling is both meaningfully and statistically different from the market-based sub-sample.

These findings suggest that data generated from the "stacked sampling" approach might lead to biased samples and an incomplete understanding of the activities of traders. In particular, our data show that between 35 and 83 percent of traders would be missing from our data if we had implemented the "stacked sample" approach by randomly sampling traders from local retail markets, depending upon the value chain. These missing data would have led us to incorrectly conclude that traders have higher rates of employment and larger sales volumes among coffee and soybean in Uganda than they actually do based on our respondent-driven sampling approach. In Bangladesh these missing data would have led us to incorrectly conclude that rice traders have larger sales volumes and potato traders have lower sales volumes. This could lead to biased inferences that lead to the implementation of misguided policies and ultimately represents a limitation that contributes to the so-called "missing middle" of agri-food value chains in academic research and policy discussions (Reardon, 2015). In short, if we would have implemented traditional approaches to sampling and collecting data on intermediary agri-food value chain actors a substantial portion of intermediary actors would likely be missing from our data. These missing actors would be left out of our data analysis and omitted from discussions about possible policies and interventions that could better support these actors.

5 Discussion and Conclusion

Despite decades of research on primary producers and end consumers, the activities of intermediary agri-food value chain actors remain understudied (Reardon and Timmer, 2007; Reardon, 2015; Barrett *et al.*, 2022; Bellemare *et al.*, 2022). The lack of research on intermediary agri-food value chain actors is particularly notable because these intermediary actors are thought to account for a substantial share of value added produced by the entire agricultural sector (Reardon, 2015; Yi *et al.*, 2021; Barrett *et al.*, 2022). A key reason for this apparent knowledge gap is the challenge in systematically collecting reliable data on intermediary agri-food value chain actors and, as a result, a limited set of quality data capable of producing credible statistics on intermediary actors with the ability to disaggregate by actor and commodity exist (Ambler *et al.*, 2023; Bellemare *et al.*, 2022).

The core challenge in collecting reliable data on intermediary agri-food value chain actors is that value chains are fundamentally a transaction-based network and, particularly in the context of low and middle income countries, intermediary actors are highly mobile and relatively informal. These features limit the *ex ante* knowledge available to researchers about the structure and scale of these value chain networks and complicate the construction of valid sampling frames required to implement a traditional random sampling approach to collect data.

To address these challenges, we adapt the respondent-driven sampling approach, developed by sociologists to study hard-to-find populations, to the context of agri-food value chains. The respondent-driven sampling approach is well-suited for settings in which a well-defined sampling frame either is difficult to construct or does not exist, and where respondents are linked together through a social or economic network. In adapting this approach to agri-food value chains we develop a novel sampling weight, the segment-adjusted sampling weight, which helps account for potential convenience sampling bias.

We apply this sampling approach in four agri-food value chains in two countries: the arabica coffee and soybean value chains in Uganda and the rice and potato value chains in Bangladesh. We document meaningful heterogeneity in the structure and scale of these agri-food value chains. Across each of the four agri-food value chains in our data, we find that intermediary actors report a diversity of both buying and selling sources; an observation that ultimately leads us to reject simple characterizations of agri-food value chains commonly found in the quantitative literature. We also investigate how different key descriptive statistics would be if we had implemented the traditional "stacked sample" approach that aims to generate representative samples by randomly sampling actors across various segments of the value chain. Focusing on traders, we find that the respondent-driven sampling approach generates a larger sample of traders who differ in observable characteristics to a sub-sample of market-based traders generated in a way that mimics traditional random sampling approach used to study traders. Our application of the respondent-driven sampling approach, therefore, provides a useful proof-of-concept as it is well-suited for settings where we do not have a well-defined sampling frame and where relatively hard-to-find respondents are linked together through some form of social or economic network.

Table 1: Sample Composition and Summary Statistics - Uganda

	Coffee Trader N=1,400 (53.2%)	**Coffee Processor N=111 (4.2%)**	**Coffee Wholesaler N=334 (12.7%)**	**Soybean Trader N=507 (19.3%)**	**Soybean Wholesaler N=280 (10.6%)**
Panel A: Enterprise Scale					
Buying Amount (in 1,000 Kg) in Last 30 Days	0.10 (1.00)	0.37 (6.07)	0.20 (8.01)	0.00 (23.39)	2.80 (146.87)
This Amount is Less than Usual Buying Amount Over Last Year	0.80 (0.40)	0.91 (0.28)	0.79 (0.41)	0.96 (0.20)	0.95 (0.21)
Usual Buying Amount (in 1,000 Kg) per Month Over the Last Year	1.00 (5.92)	5.00 (1067.76)	5.00 (195.64)	15.00 (430.15)	60.00 (991.96)
Selling Amount (in 1,000 Kg) in Last 30 Days	0.10 (0.95)	0.50 (9.66)	0.15 (5.28)	0.55 (44.29)	5.00 (106.51)
This Amount is Less than Usual Selling Amount Over Last Year	0.80 (0.40)	0.89 (0.31)	0.77 (0.42)	0.94 (0.24)	0.93 (0.25)
Usual Selling Amount (in 1,000 Kg) per Month Over the Last Year	1.00 (5.19)	4.00 (1080.42)	4.00 (97.79)	16.00 (723.50)	60.00 (1009.44)
Panel B: Enterprise Costs					
Monthly Payments on Structures Rented (in 1,000 UGX)	35.00 (89.69)	150.00 (679.73)	60.00 (259.94)	90.00 (225.42)	300.00 (319.95)
Monthly Payments on Vehicles Rented (in 1,000 UGX)	40.00 (153.70)	200.00 (9291.70)	100.00 (2080.89)	200.00 (1032.12)	250.00 (878.31)
Monthly Payments on Machines Rented (in 1,000 UGX)	30.00 (80.81)	71.70 (4120.02)	70.00 (563.91)	10.00 (0.00)	(.)
Payments on Transportation in Last 30 Days (in 1,000 UGX)	20.00 (98.97)	80.00 (767.93)	20.00 (373.30)	100.00 (923.17)	200.00 (5637.36)
Payments on Processing in Last 30 Days (in 1,000 UGX)	10.00 (76.39)	45.00 (917.08)	0.00 (240.52)	70.00 (86.53)	0.00 (0.00)
Payments on Storage in Last 30 Days (in 1,000 UGX)	20.00 (96.12)	30.00 (270.65)	200.00 (181.18)	40.00 (243.51)	300.00 (307.07)
Panel C: Demographics					
Age	40.54 (11.35)	38.74 (10.38)	42.08 (9.68)	38.49 (8.57)	38.76 (7.81)
Male	0.84 (0.37)	0.90 (0.30)	0.94 (0.24)	0.91 (0.29)	0.80 (0.40)
Highest Ed Level: Primary	0.48 (0.50)	0.23 (0.42)	0.44 (0.50)	0.42 (0.49)	0.30 (0.46)
Highest Ed Level: Secondary	0.31 (0.46)	0.30 (0.46)	0.38 (0.49)	0.35 (0.48)	0.43 (0.50)
Highest Ed Level: Tertiary	0.04 (0.20)	0.37 (0.49)	0.08 (0.27)	0.12 (0.33)	0.22 (0.42)
Has Electricity at Home	0.43 (0.50)	0.79 (0.41)	0.50 (0.50)	0.51 (0.50)	0.77 (0.42)
Has a Phone	0.98 (0.13)	0.99 (0.11)	0.99 (0.10)	0.97 (0.16)	0.99 (0.09)
Panel D: Selling Links					
Reported Any Link	0.78 (0.42)	0.80 (0.40)	0.83 (0.38)	0.63 (0.48)	0.74 (0.44)
Number of Links Reported	1.64 (1.04)	2.51 (1.67)	1.48 (0.92)	1.53 (0.85)	1.47 (0.70)
Percentage of Links with Phone Number	90.96 (19.96)	89.58 (21.01)	90.76 (19.21)	92.66 (18.92)	95.73 (14.87)
Number of Phone Links	1.47 (0.83)	1.98 (1.20)	1.59 (0.95)	1.35 (0.64)	1.22 (0.51)
Number of Duplicate Links	0.85 (0.73)	0.88 (0.98)	0.84 (0.87)	0.80 (0.74)	0.81 (0.71)

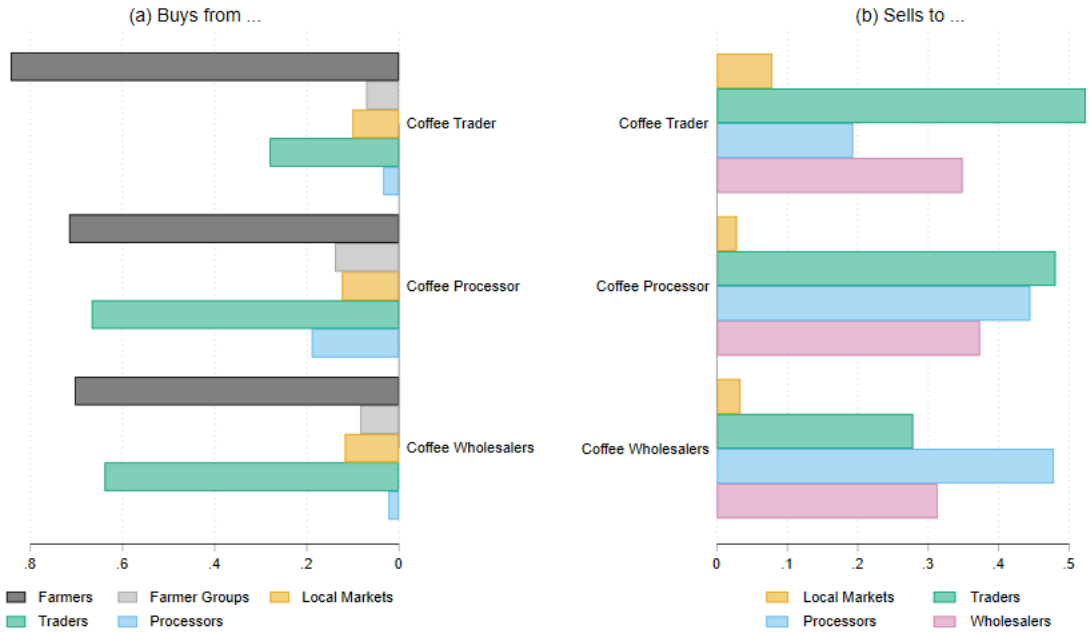
Notes: Statistics calculated by the authors with the segment-adjusted multiplicity weight applied. The parentheses indicate the standard deviation. In Panel A, median values are shown for numerical and non-binary variables, while the mean is displayed for binary variables. In panel B, all figures represent the median. Panels C and D display the mean.

Table 2: Sample Composition and Summary Statistics - Bangladesh

	Rice Trader N=1,066 (34.5%)	**Rice Processor N=456 (14.7%)**	**Rice Wholesaler N=220 (7.1%)**	**Potato Trader N=1,117 (36.1%)**	**Potato Wholesaler N=235 (7.6%)**
Panel A: Enterprise Scale					
Buying Amount (in 1,000 Kg) in Last 30 Days	22.08 (107.93)	23.18 (203.31)	15.00 (74.78)	15.00 (100.17)	18.00 (114.57)
This Amount is Less than Usual Buying Amount Over Last Year	0.50 (0.50)	0.38 (0.49)	0.53 (0.50)	0.51 (0.50)	0.46 (0.50)
Usual Buying Amount (in 1,000 Kg) per Month Over the Last Year	33.12 (138.35)	33.12 (170.12)	16.00 (84.50)	75.00 (233.32)	20.00 (262.10)
Selling Amount (in 1,000 Kg) in Last 30 Days	22.08 (105.35)	19.00 (180.02)	12.50 (70.10)	18.00 (102.58)	18.00 (117.77)
This Amount is Less than Usual Selling Amount Over Last Year	0.52 (0.50)	0.37 (0.48)	0.56 (0.50)	0.53 (0.50)	0.48 (0.50)
Usual Selling Amount (in 1,000 Kg) per Month Over the Last Year	34.50 (141.00)	28.00 (198.54)	15.00 (87.32)	75.00 (228.01)	25.00 (261.30)
Panel B: Enterprise Costs					
Monthly Payments on Structures Rented (in 1,000 BDT)	2.00 (2.33)	5.00 (6.69)	3.00 (5.35)	1.50 (2.25)	2.10 (9.88)
Monthly Payments on Vehicles Rented (in 1,000 BDT)	10.80 (47.92)	20.00 (105.59)	5.00 (36.37)	25.00 (110.68)	8.00 (80.61)
Monthly Payments on Machines Rented (in 1,000 BDT)	0.30 (10.31)	4.38 (7.69)	0.00 (0.68)	2.00 (28.15)	3.75 (43.80)
Payments on Transportation in Last 30 Days (in 1,000 BDT)	9.00 (48.79)	12.00 (65.61)	3.50 (36.91)	6.00 (53.87)	7.00 (48.70)
Payments on Processing in Last 30 Days (in 1,000 BDT)	15.00 (158.15)	8.40 (38.40)	86.10 (30.50)	20.00 (69.43)	50.00 (174.83)
Payments on Storage in Last 30 Days (in 1,000 BDT)	(.)	9.00 (1.34)	6.50 (.)	12.76 (67.40)	14.40 (112.91)
Panel C: Demographics					
Age	43.67 (9.83)	46.08 (11.60)	46.16 (10.18)	43.72 (9.87)	43.59 (10.34)
Male	1.00 (0.03)	1.00 (0.00)	1.00 (0.06)	1.00 (0.02)	0.99 (0.09)
Highest Ed Level: Primary	0.33 (0.47)	0.25 (0.44)	0.30 (0.46)	0.32 (0.47)	0.44 (0.50)
Highest Ed Level: Secondary	0.47 (0.50)	0.51 (0.50)	0.49 (0.50)	0.49 (0.50)	0.41 (0.49)
Highest Ed Level: Tertiary	0.10 (0.30)	0.14 (0.34)	0.16 (0.36)	0.10 (0.30)	0.06 (0.23)
Has Electricity at Home	0.99 (0.11)	1.00 (0.06)	1.00 (0.00)	1.00 (0.04)	0.99 (0.12)
Has a Phone	0.99 (0.09)	0.99 (0.09)	0.99 (0.07)	0.99 (0.11)	1.00 (0.00)
Panel D: Selling Links					
Reported Any Link	0.96 (0.18)	0.95 (0.21)	0.97 (0.16)	0.97 (0.17)	0.95 (0.23)
Number of Links Reported	3.03 (2.02)	3.25 (2.10)	3.16 (1.87)	3.01 (2.20)	3.52 (2.32)
Percentage of Links with Phone Number	97.82 (10.13)	96.82 (12.05)	97.56 (9.99)	98.38 (8.15)	95.05 (14.24)
Number of Phone Links	2.99 (2.01)	3.15 (2.05)	3.11 (1.89)	3.00 (2.21)	3.30 (2.29)
Number of Duplicate Links	1.87 (1.52)	1.94 (1.71)	1.13 (1.41)	1.60 (1.68)	1.22 (1.64)

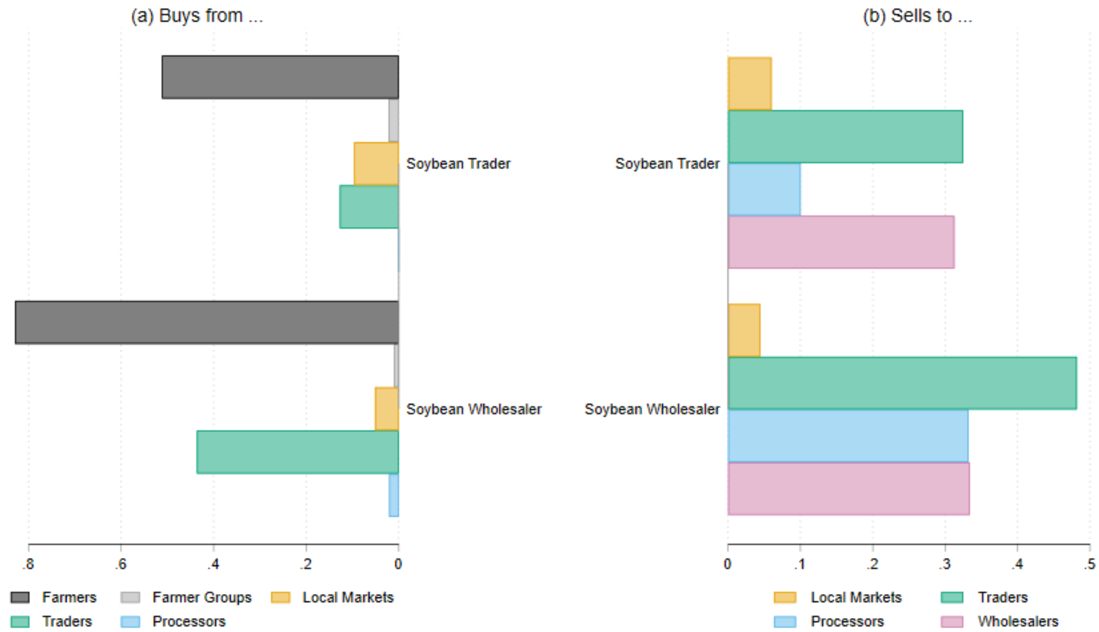
Notes: Statistics calculated by the authors with the segment-adjusted multiplicity weight applied. The parentheses indicate the standard deviation. In Panel A, median values are shown for numerical and non-binary variables, while the mean is displayed for binary variables. In panel B, all figures represent the median. Panels C and D display the mean.

Figure 1: Buying and Selling Patterns—Coffee in Uganda



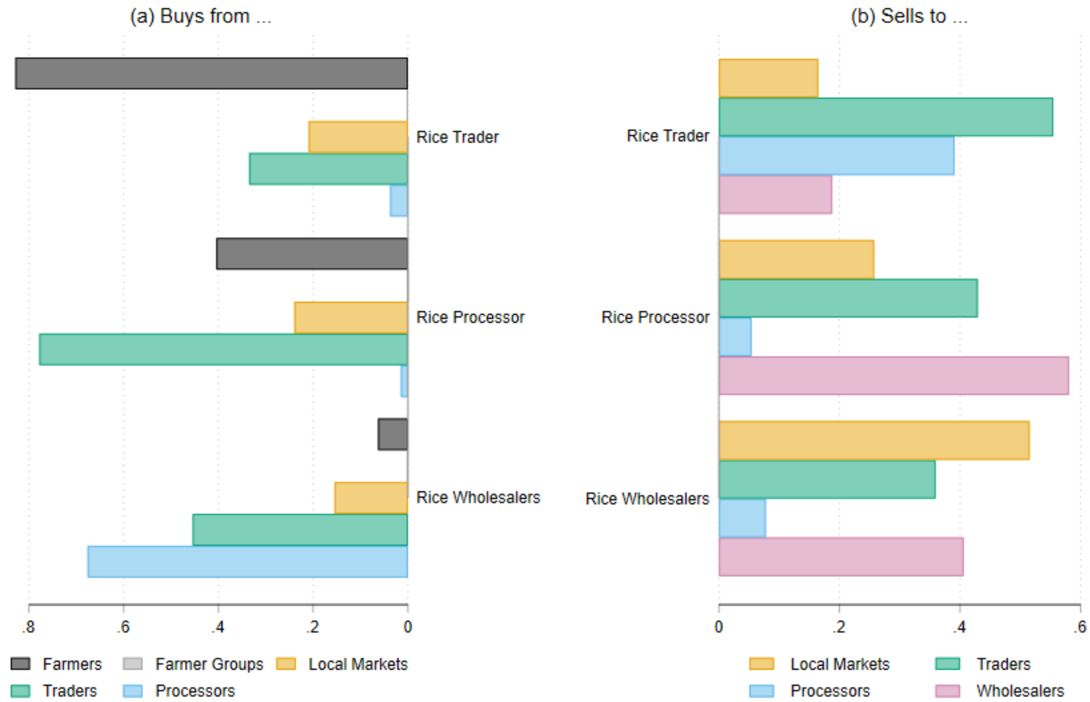
Notes: Statistics calculated by the authors with the segment-adjusted multiplicity weight applied.

Figure 2: Buying and Selling Patterns—Soybeans in Uganda



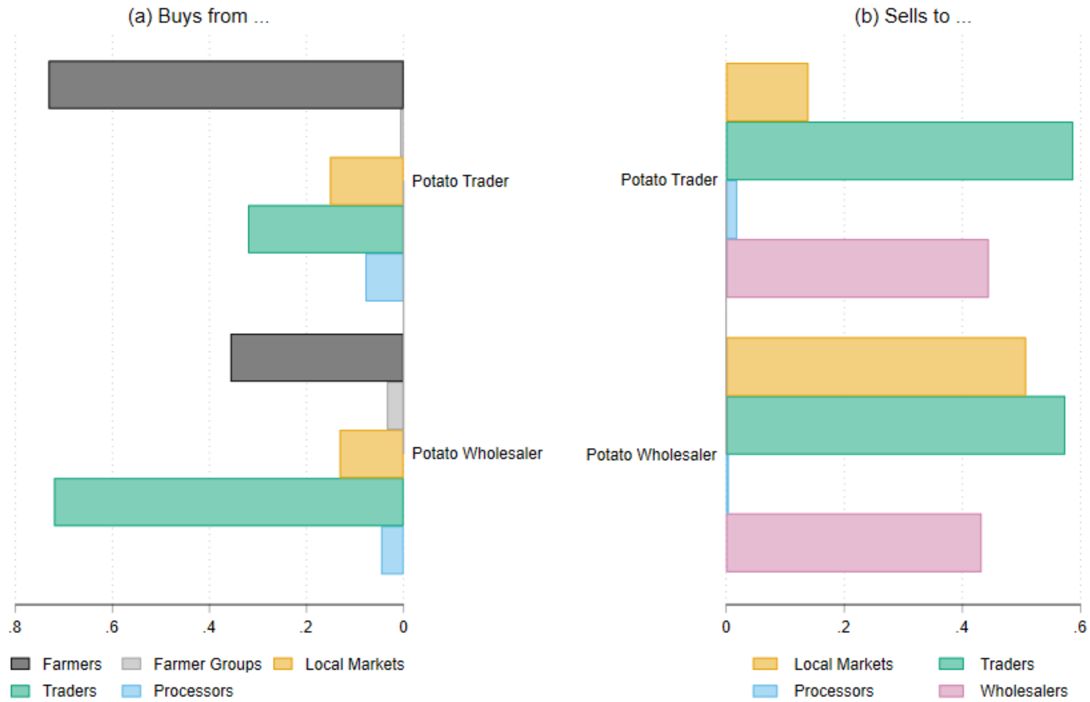
Notes: Statistics calculated by the authors with the segment-adjusted multiplicity weight applied.

Figure 3: Buying and Selling Patterns—Rice in Bangladesh



Notes: Statistics calculated by the authors with the segment-adjusted multiplicity weight applied.

Figure 4: Buying and Selling Patterns—Potatoes in Bangladesh



Notes: Statistics calculated by the authors with the segment-adjusted multiplicity weight applied.

Table 3: Heterogeneity of Coffee and Soybean Traders in Uganda

	Coffee			Soybean		
	(1)	(2)	(1)-(2)	(3)	(4)	(3)-(4)
Do you buy from other traders?	No (N=879)	Yes (N=521)	Pairwise t-test	N (N=422)	Yes (N=85)	Pairwise t-test
	Mean/(SE)	Mean/(SE)	Mean difference	Mean/(SE)	Mean/(SE)	Mean difference
Value added	901.120	1350.939	-449.819**	26.349	83.593	-57.244**
	(71.449)	(213.674)		(7.250)	(24.699)	
Has employees	0.283	0.581	-0.298***	0.420	0.729	-0.308***
	(0.015)	(0.022)		(0.024)	(0.049)	
Number of full-time employees	0.331	0.714	-0.382***	1.619	2.581	-0.961
	(0.045)	(0.085)		(0.164)	(0.329)	
Number of part-time employees	0.852	1.793	-0.941***	1.959	2.177	-0.219
	(0.084)	(0.169)		(0.635)	(0.330)	
Number of seasonal employees	1.427	3.360	-1.933***	1.508	2.686	-1.178
	(0.121)	(0.290)		(0.227)	(0.410)	
Amount purchased	0.479	4.149	-3.670	6.175	31.549	-25.373***
	(0.096)	(3.685)		(1.535)	(11.357)	
Amount sold	0.480	1.232	-0.753***	16.713	28.387	-11.674
	(0.099)	(0.129)		(3.825)	(11.322)	
Rent paid for structures	21.120	43.059	-21.939***	144.579	310.164	-165.585*
	(1.858)	(4.672)		(26.282)	(42.841)	
Rent paid for vehicles	20.974	55.857	-34.883***	124.503	476.831	-352.328***
	(2.823)	(7.131)		(26.956)	(143.448)	
Percent purchase with cash	90.746	88.578	2.168*	94.602	87.653	6.949***
	(0.647)	(0.792)		(0.537)	(2.410)	
Percent sell with cash	87.792	84.005	3.787**	90.851	83.713	7.138**
	(0.788)	(1.142)		(0.894)	(2.926)	
Use a financial account	0.282	0.376	-0.094***	0.604	0.671	-0.068
	(0.015)	(0.021)		(0.024)	(0.051)	
Can access loans when needed	0.201	0.259	-0.057**	0.198	0.251	-0.053
	(0.014)	(0.019)		(0.019)	(0.047)	
Does not manage risk	0.447	0.366	0.081**	0.442	0.418	0.024
	(0.017)	(0.021)		(0.024)	(0.054)	
Use insurance	0.007	0.019	-0.012*	0.009	0.000	0.009
	(0.003)	(0.006)		(0.004)	(0.000)	

Notes: This table shows means of observable characteristics of coffee and soybean traders disaggregated by whether they buy from other traders. Statistics calculated by the authors with the segment-adjusted multiplicity weight applied.

Table 4: Heterogeneity of Rice and Potato Traders in Bangladesh

Do you buy from other traders? Mean/(SE)	Rice (1) No (N=729) N Mean/(SE)	(2) Yes (N=337) Mean/(SE) Mean difference	(1)-(2) Pairwise t-test N	Potato (3) No (N=681) Mean/(SE)	(4) Yes (N=436) N	(3)-(4) Pairwise t-test Mean difference
Value added	2.023 (1.504)	1.752 (0.265)	0.272	0.419 (0.045)	1.249 (0.075)	-0.830***
Has employees	0.025 (0.006)	0.063 (0.013)	-0.039***	0.027 (0.006)	0.083 (0.013)	-0.056***
Number of full-time employees	0.030 (0.010)	0.065 (0.018)	-0.035*	0.058 (0.025)	0.097 (0.023)	-0.039
Number of part-time employees	1.272 (0.137)	2.570 (0.243)	-1.298***	3.780 (0.339)	3.077 (0.380)	0.703
Number of seasonal employees	7.617 (0.931)	11.650 (1.347)	-4.033**	15.887 (0.924)	22.655 (1.804)	-6.768***
Amount purchased	58.061 (5.440)	185.908 (15.005)	-127.847***	48.012 (5.209)	134.950 (17.200)	-86.938***
Amount sold	59.453 (5.869)	178.583 (13.936)	-119.130***	60.292 (7.079)	135.425 (17.220)	-75.133***
Rent paid for structures	1.109 (0.121)	2.124 (0.165)	-1.015***	0.805 (0.085)	1.312 (0.217)	-0.507**
Rent paid for vehicles	12.060 (1.548)	37.468 (4.088)	-25.409***	32.898 (8.813)	22.125 (3.494)	10.773
Percent purchase with cash	71.506 (0.946)	69.300 (1.238)	2.206	71.044 (1.025)	62.499 (1.278)	8.545***
Percent sell with cash	62.016 (0.998)	44.861 (1.551)	17.154***	50.404 (1.283)	31.833 (1.407)	18.571***
Use a financial account	0.679 (0.017)	0.865 (0.019)	-0.186***	0.744 (0.017)	0.869 (0.016)	-0.125***
Can access loans when needed	0.394 (0.018)	0.479 (0.027)	-0.085***	0.462 (0.019)	0.428 (0.024)	0.035
Does not manage risk	0.759 (0.016)	0.628 (0.026)	0.131***	0.730 (0.017)	0.631 (0.023)	0.099***
Use insurance	0.025 (0.006)	0.083 (0.015)	-0.059***	0.034 (0.007)	0.038 (0.009)	-0.004

Notes: This table shows means of observable characteristics of rice and potato traders disaggregated by whether they buy from other traders. Statistics calculated by the authors with the segment-adjusted multiplicity weight applied.

Table 5: Value Chain Length

	(1) Coffee	(2) Soybean	(3) Rice	(4) Potato
Average length	1.83	1.37	3.21	2.79
	(0.04)	(0.03)	(0.09)	(0.09)
Length w. endpoint trader	1.85	1.21	2.77	2.72
	(0.05)	(0.03)	(0.13)	(0.12)
Length w. endpoint processor	2.28	n/a	3.53	2.98
	(0.21)		(0.23)	(0.85)
Length w. endpoint wholesaler	1.65	1.66	4.35	3.12
	(0.07)	(0.05)	(0.34)	(0.36)
No. of traders	1.04	0.94	1.85	2.02
	(0.02)	(0.02)	(0.05)	(0.07)
No. of processors	0.07	0.00	0.60	0.00
	(0.01)	(0.00)	(0.04)	(0.00)
No. of wholesalers	0.25	0.34	0.27	0.32
	(0.01)	(0.02)	(0.03)	(0.03)

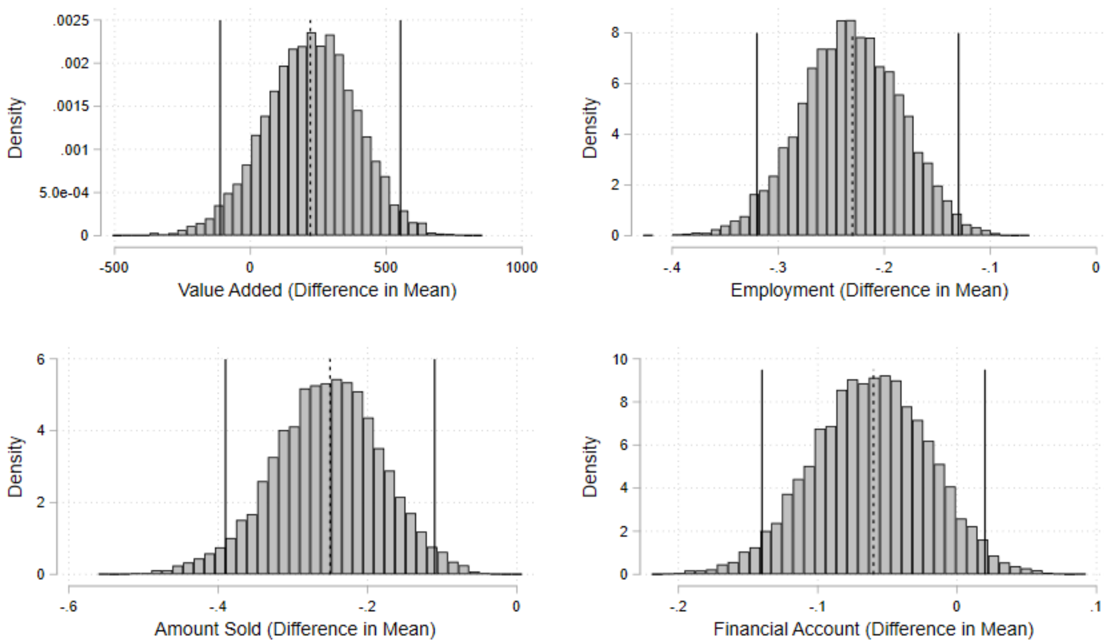
Notes: This table reports means (standard deviations in parenthesis). Statistics calculated by the authors with the segment-adjusted multiplicity weight applied. The value chain length indicates the average number of transactions we observe between the primary producer and a given endpoint in our data. We first create a database containing all possible chains for each seed farmer. To derive the mean length and the value chain composition, we randomly sample one chain for each farmer and repeat this process 1,000 times. To account for the fact that our iterative data collection process doesn't end "naturally" in all instances, we apply an attrition correction to the value chain length estimates.

Table 6: Full Trader Sample vs. Market-Based Trader Sub-Sample

	(1) Value Added	(2) Employment	(3) Amount Sold	(4) Financial Account
Panel A: Coffee in Uganda				
Full sample	1001.19	0.35	0.43	0.30
(n=1,400)	(2967.82)	(0.48)	(0.93)	(0.46)
Market-based sub-sample	781.18	0.58	0.68	0.36
(n=242)	(1751.22)	(0.49)	(1.10)	(0.48)
Panel B: Soybeans in Uganda				
Full sample	30.13	0.44	11.07	0.61
(n=507)	(155.89)	(0.50)	(44.28)	(0.49)
Market-based sub-sample	7.86	0.59	25.68	0.58
(n=87)	(186.92)	(0.50)	(66.45)	(0.49)
Panel C: Rice in Bangladesh				
Full sample	1.94	0.04	84.06	0.73
(n=1,066)	(34.09)	(0.19)	(152.03)	(0.44)
Market-based sub-sample	1.52	0.02	98.48	0.75
(n=375)	(4.18)	(0.16)	(175.76)	(0.43)
Panel D: Potatoes in Bangladesh				
Full sample	0.62	0.04	58.0	0.77
(n=1,117)	(1.32)	(0.20)	(102.5)	(0.42)
Market-based sub-sample	0.72	0.04	46.08	0.72
(n=299)	(1.54)	(0.21)	(87.58)	(0.45)

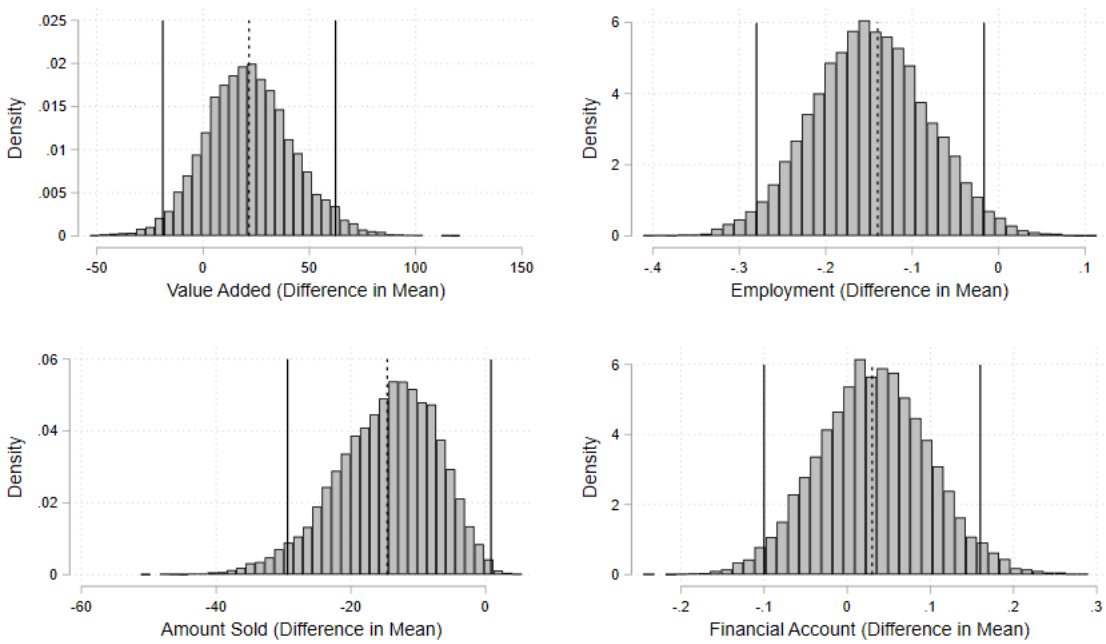
Notes: This table reports means (standard deviations in parenthesis) in our full trader sample (i.e., generated with our respondent-driven sampling approach) and in the sub-sample of market-based traders (i.e., mimicking the traditional "stacked sample" approach). Statistics calculated by the authors with the segment-adjusted multiplicity weight applied.

Figure 5: Respondent-Driven Sampling vs. Random Sampling, Coffee in Uganda



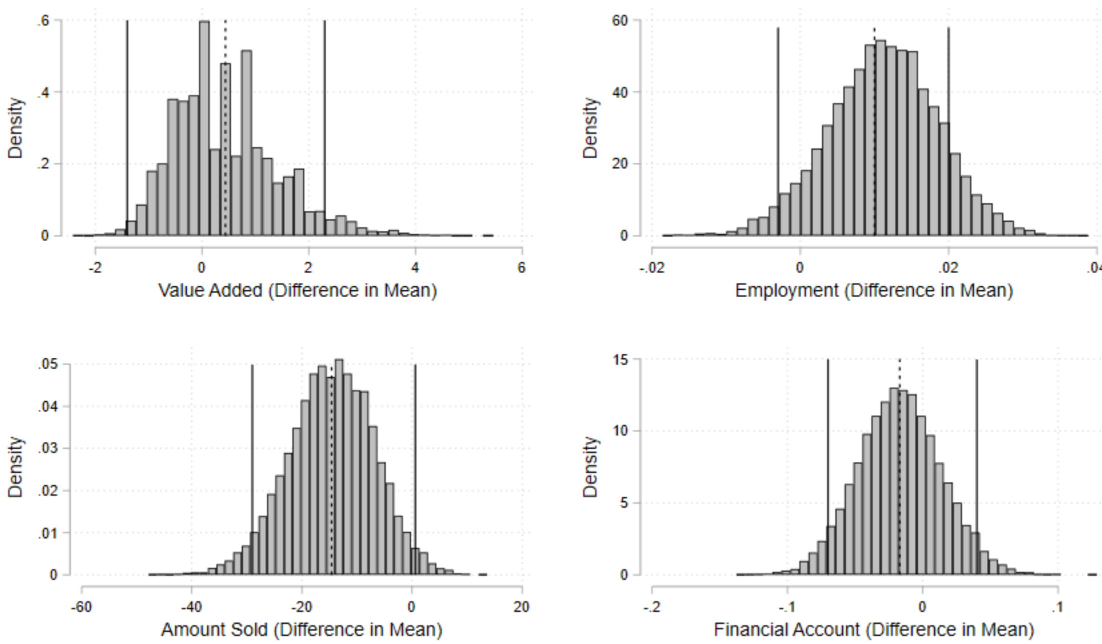
Notes: Each panel in this figure plots the distribution of the difference in the mean of a given variable in our full trader sample (i.e., generated with our respondent-driven sampling approach) and in the sub-sample of market-based traders (i.e., mimicking in the traditional "stacked sample" approach). The data are re-sampled with replacement 10,000 times. The solid lines indicate the 95 percent confidence interval and the dotted line represents the mean of the distribution. Statistics calculated by the authors with the segment-adjusted multiplicity weight applied.

Figure 6: Re-sampling Difference in Means, Soybeans in Uganda



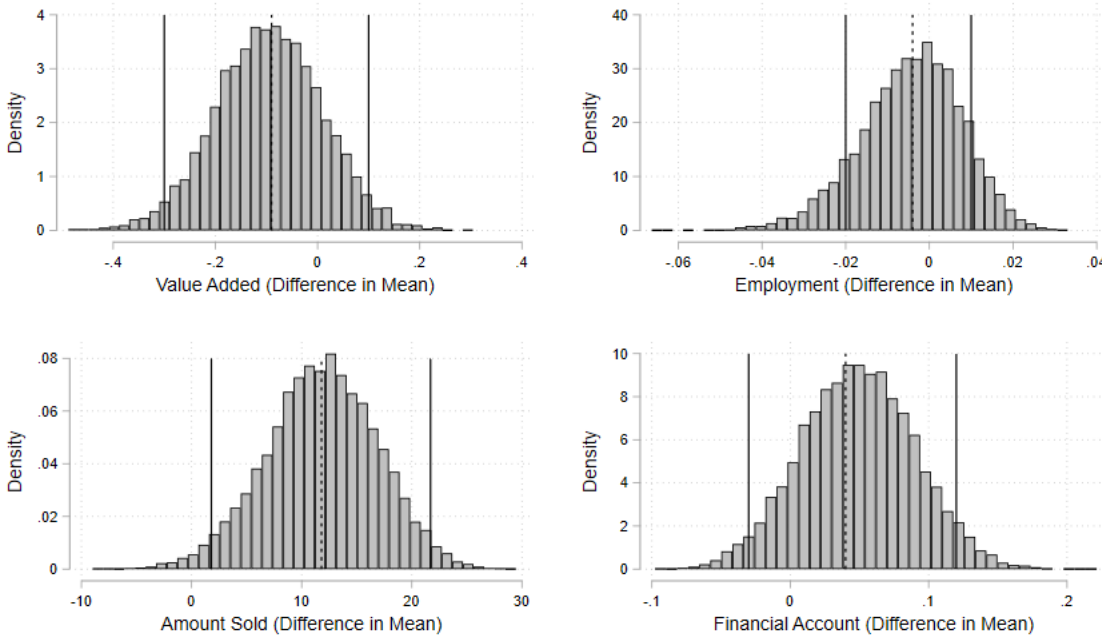
Notes: Each panel in this figure plots the distribution of the difference in the mean of a given variable in our full trader sample (i.e., mimicking with our respondent-driven sampling approach) and in the sub-sample of market-based traders (i.e., generated in the traditional "stacked sample" approach). The data are re-sampled with replacement 10,000 times. The solid lines indicate the 95 percent confidence interval and the dotted line represents the mean of the distribution. Statistics calculated by the authors with the segment-adjusted multiplicity weight applied.

Figure 7: Re-sampling Difference in Means, Rice in Bangladesh



Notes: Each panel in this figure plots the distribution of the difference in the mean of a given variable in our full trader sample (i.e., mimicking with our respondent-driven sampling approach) and in the sub-sample of market-based traders (i.e., generated in the traditional "stacked sample" approach). The data are re-sampled with replacement 10,000 times. The solid lines indicate the 95 percent confidence interval and the dotted line represents the mean of the distribution. Statistics calculated by the authors with the segment-adjusted multiplicity weight applied.

Figure 8: Re-sampling Difference in Means, Potatoes in Bangladesh



Notes: Each panel in this figure plots the distribution of the difference in the mean of a given variable in our full trader sample (i.e., mimicking with our respondent-driven sampling approach) and in the sub-sample of market-based traders (i.e., generated in the traditional "stacked sample" approach). The data are re-sampled with replacement 10,000 times. The solid lines indicate the 95 percent confidence interval and the dotted line represents the mean of the distribution. Statistics calculated by the authors with the segment-adjusted multiplicity weight applied.

References

Ahmed, A. and Bakhtiar, M. M. (2020) *Boro rice procurement in Bangladesh: Implications for policy*, Intl Food Policy Res Inst.

Aker, J. C. (2010) Information from markets near and far: Mobile phones and agricultural markets in niger, *American Economic Journal: Applied Economics*, **2**, 46–59.

Ambler, K., de Brauw, A., Herskowitz, S. and Pulido, C. (2023) Financial needs of the agricultural midstream, *Food Policy*, **121**.

Ayele, S., Duncan, A., Larbi, A. and Khanh, T. T. (2012) Enhancing innovation in livestock value chains through networks: Lessons from fodder innovation case studies in developing countries, *Science and public policy*, **39**, 333–346.

Banerjee, A., Fischer, G., Karlan, D., Lowe, M. and Roth, B. N. (2023) Do microenterprises maximize profits? a vegetable market experiment in india, *Unpublished working paper*.

Barrett, C. B., Reardon, T., Swinnen, J. and Zilberman, D. (2022) Agri-food value chain revolutions in low-and middle-income countries, *Journal of Economic Literature*, **60**, 1316–1377.

Bellemare, M. F., Bloem, J. R. and Lim, S. (2022) Producers, consumers, and value chains in low-and middle-income countries, in *Handbook of Agricultural Economics*, Elsevier, vol. 6, pp. 4933–4996.

Bergquist, L. F. and Dinerstein, M. (2020) Competition and entry in agricultural markets: Experimental evidence from kenya, *American Economic Review*, **110**, 3705–3747.

Beshah, B., Kitaw, D. and Dejene, T. (2013) Quality and value chain analyses of ethiopian coffee, *Journal of Agriculture and Social Research (JASR)*, **13**, 35–41.

Breisinger, C., Thomas, M. and Thurlow, J. (2009) *Social accounting matrices and multiplier analysis: An introduction with exercises*, vol. 5, Intl Food Policy Res Inst.

Canning, P., Weersink, A. and Kelly, J. (2016) Farm share of the food dollar: an io approach for the united states and canada, *Agricultural Economics*, **47**, 505–512.

Casaburi, L. and Macchiavello, R. (2019) Demand and supply of infrequent payments as a commitment device: evidence from kenya, *American Economic Review*, **109**, 523–555.

Casaburi, L. and Reed, T. (2022) Using individual-level randomized treatment to learn about market structure, *American Economic Journal: Applied Economics*, **14**, 58–90.

Dries, L., Germenji, E., Noev, N. and Swinnen, J. F. (2009) Farmers, vertical coordination, and the restructuring of dairy supply chains in central and eastern europe, *World development*, **37**, 1742–1758.

Fafchamps, M., Hill, R. V. and Minten, B. (2008) Quality control in nonstaple food markets: evidence from india, *Agricultural Economics*, **38**, 251–266.

FAO (2014) Developing sustainable food value chains—guiding principles, *Food and Agricultural Organization of the United Nations*.

FAO (2015) Food security and international trade: Unpacking disputed narratives, *Food and Agriculture Organization of the United Nations*.

Gow, H. R., Streeter, D. H. and Swinnen, J. F. (2000) How private contract enforcement mechanisms can succeed where public institutions, *Agricultural economics*, **23**, 253–265.

Goyal, A. (2010) Information, direct access to farmers, and rural market performance in central india, *American Economic Journal: Applied Economics*, **2**, 22–45.

Heckathorn, D. D. (1997) Respondent-driven sampling: a new approach to the study of hidden populations, *Social problems*, **44**, 174–199.

Heckathorn, D. D. (2011) Comment: Snowball versus respondent-driven sampling, *Sociological methodology*, **41**, 355–366.

Heckathorn, D. D. and Cameron, C. J. (2017) Network sampling: From snowball and multiplicity to respondent-driven sampling, *Annual review of sociology*, **43**, 101–119.

HLPE (2017) *Nutrition and food systems*, A report by the High Level Panel of Experts on Food Security and Nutrition of the Committee on World Food Security, Rome.

Hsieh, C.-S., Hsu, Y.-C., Ko, S. I., Kovárık, J. and Logan, T. D. (2024) Non-representative sampled networks: Estimation of network structural properties by weighting.

Jensen, R. (2007) The digital provide: Information (technology), market performance, and welfare in the south indian fisheries sector, *The quarterly journal of economics*, **122**, 879–924.

Kilelu, C. W., Klerkx, L. and Leeuwis, C. (2013) Unravelling the role of innovation platforms in supporting co-evolution of innovation: Contributions and tensions in a smallholder dairy development programme, *Agricultural systems*, **118**, 65–77.

Liverpool-Tasie, L. S. O., Wineman, A., Young, S., Tambo, J., Vargas, C., Reardon, T., Adjognon, G. S., Porciello, J., Gathoni, N., Bizikova, L. *et al.* (2020) A scoping review of market links between value chain actors and small-scale producers in developing regions, *Nature Sustainability*, **3**, 799–808.

Liverpool-Tasie, S., Reardon, T., Sanou, A., Ogunleye, W., Ogunbayo, I. and Omonona, B. T. (2017) The transformation of value chains in africa: Evidence from the first large survey of maize traders in nigeria, Tech. rep.

Macchiavello, R. and Miquel-Florensa, J. (2019) Buyer-driven upgrading in gvcs: The sustainable quality program in colombia.

Macchiavello, R. and Morjaria, A. (2021) Competition and relational contracts in the rwanda coffee chain, *The Quarterly Journal of Economics*, **136**, 1089–1143.

Minten, B., Assefa, T. and Hirvonen, K. (2017) Can agricultural traders be trusted? evidence from coffee in ethiopia, *World Development*, **90**, 77–88.

Minten, B., Murshid, K. and Reardon, T. (2013) Food quality changes and implications: evidence from the rice value chain of bangladesh, *World Development*, **42**, 100–113.

Minten, B., Randrianarison, L. and Swinnen, J. F. (2009) Global retail chains and poor farmers: Evidence from madagascar, *World development*, **37**, 1728–1741.

Minten, B., Tamru, S., Engida, E. and Kuma, T. (2016) Feeding africa's cities: The case of the supply chain of teff to addis ababa, *Economic Development and Cultural Change*, **64**, 265–297.

Minten, B., Vandeplas, A. and Swinnen, J. (2012) Regulations, brokers, and interlinkages: The institutional organization of wholesale markets in india, *Journal of Development Studies*, **48**, 864–886.

Reardon, T. (2015) The hidden middle: the quiet revolution in the midstream of agrifood value chains in developing countries, *Oxford Review of Economic Policy*, **31**, 45–63.

Reardon, T., Chen, K. Z., Minten, B. and Adriano, L. (2012) The quiet revolution in staple food value chains in asia: Enter the dragon, the elephant, and the tiger, *Asian Development Bank and International Food Policy Research Institute, Philippines*.

Reardon, T., Liverpool-Tasie, L. S. O. and Minten, B. (2021) Quiet revolution by smes in the midstream of value chains in developing regions: wholesale markets, wholesalers, logistics, and processing, *Food Security*, **13**, 1577–1594.

Reardon, T. and Timmer, C. P. (2007) Transformation of markets for agricultural output in developing countries since 1950: How has thinking changed?, *Handbook of agricultural economics*, **3**, 2807–2855.

Rothbart, G. S., Fine, M. and Sudman, S. (1982) On finding and interviewing the needles in the haystack: The use of multiplicity sampling, *Public Opinion Quarterly*, **46**, 408–421.

Sirken, M. G. (1970) Household surveys with multiplicity, *Journal of the American statistical Association*, **65**, 257–266.

Soullier, G. and Moustier, P. (2021) The modernization of the rice value chain in senegal: A move towards the asian quiet revolution?, *Development Policy Review*, **39**, O81–O101.

Stür, W., Khanh, T. T. and Duncan, A. (2013) Transformation of smallholder beef cattle production in vietnam, *International journal of agricultural sustainability*, **11**, 363–381.

UNCTAD (2009) World investment report: Transnational corporations, agricultural production and development, *United Nations Conference on Trade and Development*.

Vicol, M., Neilson, J., Hartatri, D. F. S. and Cooper, P. (2018) Upgrading for whom? relationship coffee, value chain interventions and rural development in indonesia, *World Development*, **110**, 26–37.

Wiseman, E. (2023) Border trade and information frictions: Evidence from informal traders in kenya.

Yi, J., Meemken, E.-M., Mazariegos-Anastassiou, V., Liu, J., Kim, E., Gómez, M. I., Canning, P. and Barrett, C. B. (2021) Post-farmgate food value chains make up most of consumer food expenditures globally, *Nature Food*, **2**, 417–425.

ALL IFPRI DISCUSSION PAPERS

All discussion papers are available **here**

They can be downloaded free of charge

INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE
www.ifpri.org

IFPRI HEADQUARTERS
1201 Eye Street, NW
Washington, DC 20005 USA
Tel.: +1-202-862-5600
Fax: +1-202-862-5606
Email: ifpri@cgiar.org