

SEAFOOD MISLABELING, FISH EFFICIENCY, AND CHILD TIME USE: THREE ESSAYS
IN AQUACULTURE AND AGRICULTURAL ECONOMICS

By

Eric Abaidoo

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Agricultural, Food, and Resource Economics – Doctor of Philosophy

2023

ABSTRACT

This dissertation consists of three essays, exploring (1) the potential disruptive effects of seafood mislabeling (2) how rural non-farm employment (RNFE) conditions the relationship between agricultural diversification and aquaculture efficiency, and (3) the impact of parental education on child time use.

The first chapter, titled “Fish demand in the U.S. Great Lakes region in the face of seafood mislabeling” investigates whether consumer WTP for local seafood is impacted by information about seafood fraud. The globalization of seafood trade has triggered heightened vulnerability for fraud within the seafood supply network. Consumer perceptions of these vulnerabilities are not limited to imported seafood products, as spillover effects are likely to influence purchasing behavior for domestically produced seafood as well. Applying a discrete choice methodology, I show that consumers derive positive utility from consuming locally sourced relative to imported seafood. Upon further disaggregation, however, I find that for one consumer segment, which I term the price-sensitive group, localness does not command a significant positive premium. Most importantly, I demonstrate that information regarding international seafood fraud largely did not alter local seafood demand. That said, I find some evidence of a negative spillover effect of the information treatment on US-labeled seafood in one consumer subgroup.

The second chapter, titled “Does rural non-farm employment (RNFE) resolve (or exacerbate) the agricultural diversification-farm efficiency tradeoff?” studies how RNFE conditions the relationship between agricultural diversification and fish efficiency? Competition for scarce productive resources typically implies a compromise between agricultural diversification and efficiency. Yet the potential for non-farm income to resolve this tradeoff remains understudied. Cash from non-farm sources may support productivity-enhancing input purchase, thereby

improving efficiency. On the other hand, by diversifying both on and off-farm, households may be spreading their labor resources too thin, thus lowering fish efficiency. Using micro-level data on fish farming households in Southern Bangladesh, I show that at higher levels of the non-farm income share, diversification into crops results in significant allocative inefficiencies. Results are weaker for the technical efficiency measure.

The third chapter, titled “Parental educational attainment and child labor outcomes: Evidence from Malawi” revisits a hot-button topic—child labor use in agricultural production. Prior studies present anecdotal evidence thus far with evidence on the causal interpretation of this relationship rarely explored. I draw on insights from the demography literature, wherein findings suggest that the direct influence of grandparents or lack thereof on grandchildren’s socioeconomic outcomes hinges crucially on familial living arrangements. Hence, conditional on a range of parental characteristics and multigenerational co-residence, I use as a set of instruments grandparents’ educational attainment to exploit plausibly exogenous variation in parents’ schooling. Using a nationally representative Malawian household panel data set, I generally find evidence of a negative parental educational attainment impact on child labor outcomes. The effect of maternal education on household farm work, however, is not significant. My 2SLS results are also shown to be robust to varying degrees of violation of the exclusion restriction. With respect to potential mechanisms, the results suggest that engagement in non-farm employment pursuits among educated parents may mediate these effects.

ACKNOWLEDGEMENTS

Glory be to God Almighty for bringing me over the finish line. I sought His face at the beginning of this journey and He stood with me every step of the way. For that, I am forever grateful.

I would also like to extend my sincere gratitude to my major professor, Dr. Ben Belton, for his mentorship, and support over the years. His commitment to seeing me succeed in this program has been pivotal to the timely completion of this dissertation. I am also grateful to the other members of my dissertation committee, Drs. Thomas Reardon, Songqing Jin, and Trey Malone, whose insights and constructive feedback have immensely benefitted this dissertation.

Thanks are also due to my fellow graduate student colleagues (past and present) in the AFRE and Economics programs, for their advice, and friendship during this journey. In the same breath, I say a massive thank you to my church, Grace International Outreach Church (GIOC), for being my community in Christ and helping me grow my faith as I worked towards this degree.

Most importantly, I would like to acknowledge the support and sacrifices of my beloved family, Mrs. Dorcas Tetteh (my partner), Joel Abaidoo (our son), Mrs. Rebecca Sackey (my mom), Mrs. Becky Hubbell (my U.S. mom), Mr. Ebenezer Abaidoo (my dad), Mr. John Hubbell (my U.S. dad), and siblings for their constant supply of inspiration, energy, and calm throughout this journey.

Special thanks to the Bailey Scholars Program (BSP) for the community and providing me the platform for self-discovery. And to all my wellwishers—you have all contributed your part in ways that you could never imagine. Thank you!

TABLE OF CONTENTS

CHAPTER 1: FISH DEMAND IN THE U.S. GREAT LAKES REGION IN THE FACE OF SEAFOOD MISLABELING	1
1.1 Introduction	1
1.2 Background	4
1.3 Mapping Fraud in Seafood Supply Chains	7
1.4 Consumer Preferences under Fraud Uncertainty	10
1.5 Methods	14
1.6 Empirical Strategy	15
1.7 Data and Descriptives	18
1.8 Econometric Results	19
1.9 Conclusions	26
BIBLIOGRAPHY	28
APPENDIX A: TABLES AND FIGURES	33
APPENDIX B: DEFINITIONS AND EXCERPT	52
 CHAPTER 2: DOES RURAL NON-FARM EMPLOYMENT RELIEVE (OR EXACERBATE) THE AGRICULTURAL DIVERSIFICATION-FARM EFFICIENCY TRADEOFF: THE CASE OF AQUACULTURE IN BANGLADESH	 54
2.1 Introduction	54
2.2 Data and Descriptives	58
2.3 Empirical Strategy	61
2.4 Regression Results	66
2.5 Conclusions	72
BIBLIOGRAPHY	75
APPENDIX A: TABLES AND FIGURES	77
APPENDIX B: SUPPLEMENTARY TABLES	89
 CHAPTER 3: PARENTAL EDUCATIONAL ATTAINMENT AND CHILD LABOR: EVIDENCE FROM MALAWI	 95
3.1 Introduction	95
3.2 Related Literature	99
3.3 Data	101
3.4 Empirical Strategy	104
3.5 Addressing Endogeneity	106
3.6 Results	108
3.7 Imperfect Instruments Sensitivity Analysis	113
3.8 Potential Mechanisms	114
3.9 Conclusions	115
BIBLIOGRAPHY	118
APPENDIX A: TABLES AND FIGURES	121
APPENDIX B: THEORETICAL MODEL	143

CHAPTER 1: FISH DEMAND IN THE U.S. GREAT LAKES REGION IN THE FACE OF SEAFOOD MISLABELING

1.1 Introduction

Growing demand and increasingly sophisticated globalized agri-food networks have coincided with a precipitous increase in international seafood trade over the past few decades. Indeed, seafood is currently one of the world's most widely traded food commodities (Asche, et al., 2022; Gephart, et al., 2019; Kroetz, et al., 2020). Between 1986 and 2018, global seafood export volume almost doubled, while seafood exports climbed from \$37 billion to \$164 billion in value (FAO, 2020). This growth has fostered wider access to seafood originating far from the point of purchase which, in turn, has fueled the recent demand for traceability and origin-labeling (FAO, 2020). Figure 1.1 presents evidence of global seafood export volume growth since the mid-1980s. During this time, wild-caught fishery production has remained relatively stable (Abaidoo et al., 2021). While global seafood markets are predicted to double in size by the year 2050, wild-caught production is expected to contribute little to this additional growth (Waite, et al., 2014; Belton, Reardon, & Zilberman, 2020).

This trend is particularly important for U.S. seafood markets where more than 70 percent of domestic seafood consumption originates outside the United States (NOAA, 2021). Other accounts suggest a more conservative estimate due to reexports. According to Gephart et al. (2019), foreign imports account for 62 to 65% of domestic seafood consumption. Population growth and changing consumer tastes and preferences will likely drive this percentage up further. To meet this demand, U.S. retailers and restaurants rely on imported aquatic products from countries such as Norway, China, and Canada (Abaidoo et al., 2021). As traditional seafood production regions become increasingly strained, supply networks will grow longer and more

complex, creating additional vulnerabilities to fraud. Indeed, food fraud concerns have emerged in recent food policy debates (Meerza & Gustafson, 2020; Spink & Moyer, 2011; Spink, Ortega, Chen, & Wu, 2017). In particular, seafood fraud has dominated global news headlines in recent years (Warner et al., 2013).

Seafood constitutes one of the most susceptible categories to food fraud (see Figure 1.2) (Johnson, 2014; Kroetz, et al., 2020; Meerza & Gustafson, 2020). Prior studies indicate that seafood markets are the most susceptible to adulteration in the United States, followed by dairy and meat (Bitzios et al., 2017; Schug, 2016). In one such investigation, 44% of visited retail outlets sold mislabeled fish (O'Neill et al., 2015; Warner et al., 2013). This nationwide query further highlighted the considerable variability in seafood mislabeling rates by retail outlet type, with sushi spots emerging as the most targeted (74%), followed by restaurants (38%), then grocery stores (18%) (Warner et al., 2013).

To date, the effect of fraud on consumer preferences and the ensuing demand for affected food products remains understudied (Theolier et al., 2021). In particular, despite the prevalence of seafood fraud, few studies have explored its effects on consumer preferences. A notable exception is McCallum et al. (2022), who use an artefactual field experiment with European consumers to estimate their willingness to avoid the risk and/or uncertainty of purchasing inauthentic fish; the authors find that consumers are indeed willing to pay a premium to avoid food fraud.

Instead, prior research has largely focused on estimating consumer willingness-to-pay (WTP) for select food safety attributes for a variety of imported food products (Hayes et al., 1995; Ortega et al., 2014, 2015). To illustrate, in the food safety domain, Ortega et al. (2014) evaluated U.S. consumer preferences for enhanced food safety claims, finding that U.S. consumers have a

higher WTP for the food safety attributes of U.S.-farmed seafood products relative to those primarily sourced from Asia.

This article studies how U.S. consumers weigh tradeoffs between locally sourced and imported seafood given the potential for fraud. We contribute to the literature in three ways. First, we map out a seafood supply chain and highlight possible areas of food fraud vulnerability along the chain. Second, we model consumer demand for select traceability, production method, and processing attributes of a diversity of seafood species in the Great Lakes region. Finally, we examine whether and how different consumer market segments respond to information about the prevalence of seafood fraud across domestic and imported seafood supply chains. This research question is motivated in large part by concerns from local seafood producers regarding negative spillover effects due to fraud perceptions. That is, producers might lose out on product premiums if consumers perceive that some degree of fraud is inevitable.

This study also adds to a growing literature on information effects on U.S. consumer preferences and demand for seafood (Marette et al., 2008a, 2008b; Uchida et al., 2017; Weir, Uchida, & Vadivelo, 2021). While studying the market potential for genetically modified (GM) salmon, Weir et al. (2021) conclude that *ex ante* negative biases do matter, in that providing both negative and positive information about GM fresh salmon had similar effects on WTP as presenting negative information only. By contrast, Uchida et al. (2017) find evidence of no spillover effect of unfavorable information about the mercury content of swordfish on consumer bids for wild and farmed salmon, paralleling one of the main findings of our study.

The remainder of this article is structured as follows. In the next section, we provide a brief background on food fraud as it pertains to domestic and international food systems. We then present a conceptual framework of seafood supply chains, highlighting potential areas of

vulnerability to a range of fraudulent activities. We then describe our main hypotheses. After presenting our empirical strategy, we report summary statistics on key variables in our data set as well as the estimation results. We follow this up with a detailed discussion of our results and conclude.

1.2 Background

Variously described as the adulteration, substitution, dilution, stealing, tampering, diversion, and misrepresentation or mislabeling of food products for economic gain, highly publicized food fraud incidents have received substantial media attention in the past two decades (Spink & Moyer, 2011; Spink et al., 2017). Scandals ranging from horsemeat in European beef markets (Premanadh, 2013) to melamine in infant formula in China are some examples of globally recognized food fraud events in recent history (Ingelfinger, 2008; Chan et al., 2008). In extreme cases, failure to detect these adulterants promptly can result in devastating health outcomes for consumers. For example, the high-profile melamine scandal of 2008 affected about 300,000 children, with almost 50,000 hospitalizations, leading to 6 deaths (Ingelfinger, 2008; Chan et al., 2008; FAO, 2008; Meerza & Gustafson, 2020; Spink et al., 2017; Yang, et al., 2022).

Wherever conditions for opportunistic behavior exist, and mechanisms to remedy such shortcomings are non-existent or ineffective, food fraud will be sure to feature. Indeed, almost every food product has had some history of fraud. Alum and chalk in bread flour, exhausted tea leaves in tea bags, and inferior spirits in branded spirit bottles have all made notable appearances in agri-food supply networks, jeopardizing human health and causing significant economic losses to consumers (Shears, 2010; Meerza & Gustafson, 2020). As a consequence, food fraud remains an enduring concern with far-reaching repercussions due to globalized food supply chains. This

phenomenon is so prevalent that in some cases consumers have begun to develop a strong taste for adulterated food products and beverages (Shears, 2010).

Food fraud events can occur at any point throughout the system, but some food types are more vulnerable and thus easier to manipulate than others. For example, the adulteration of high-quality extra virgin olive oil (EVOO) with inexpensive, low-quality seed oil is a widespread practice given that such manipulations are impossible to detect without the aid of accurate science and high-precision technologies. Also, quality differentiation along origin, olive type, and chemical composition lines imply that consumers will have a tough time deciphering mislabeled or adulterated EVOO, as it is often the case with credence attributes more generally (Meerza & Gustafson, 2020). Food safety management systems may be effective at detecting harmful additives but may miss or fail to alert consumers to substitutions or dilutions that do not present a human health risk.

In the fisheries and aquaculture sector, food fraud can be extremely challenging to detect (Reilly, 2018; Warner et al., 2013). Asymmetric information between consumers and suppliers has fostered fraudulent activities including species substitution, intentional mislabeling, and undisclosed use of water-adhesive agents to increase fish weight for economic gain (Reilly, 2018). Moreover, the practice of processing seafood offsite and then reexporting to the origin further complicates traceability, fostering mislabeling (Asche et al., 2022). Despite routine monitoring and testing by food safety surveillance agencies such as the U.S. Food and Drug Administration (FDA), imported food products including seafood have become common targets of food fraud (Reilly, 2018; Meerza & Gustafson, 2020). To date, the Seafood Import Monitoring Program (SIMP), established to oversee compliance with general recordkeeping and reporting requirements

for imported seafood, covers only 13 seafood species groups (Warner et al., 2013).¹ This risk-based traceability program was designed to mitigate instances of illegal, unreported, and unregulated fishing (IUU), and seafood fraud. However, given the inexhaustive nature of the program's coverage, seafood types not currently covered by SIMP are fraught with mislabeling (Warner et al., 2013). As a result, consumer demand for food safety and authentication labeling is fast gaining traction as certification entities represent only a partial solution to the asymmetric information problem underlining food fraud activities (Giannakas, 2002; Ortega et al., 2014; Zilberman et al., 2018).

Instead of focusing on fraud more broadly, the prior literature has largely emphasized the food safety of imported food products. For example, Ortega et al. (2015) explored media coverage effects of food safety incidents on U.S. consumer preferences for imported aquaculture products originating in Asia. Findings from this study suggest that U.S. consumer WTP for aquatic food products was impacted by exposure to major food safety news headlines. Specifically, their results indicate that consumer WTP for enhanced food safety claims declined substantially for shrimp originating in China, and Thailand following exposure to food safety media information. By contrast, no notable changes in consumer valuation of the enhanced food safety attribute were observed for domestic seafood products given the information shock.

Admittedly, not all forms of food fraud pose food safety risks or health challenges to consumers. However, such practices can dislodge consumer confidence in food labeling and the safety of certain food industries (Giannakas, 2002; Meerza & Gustafson, 2020). For instance, some pork products have been found to be fraudulent, dislodging the trust of some religiously affiliated

¹ Seafood species groups covered by SIMP include Abalone, Atlantic Cod, Blue Crab (Atlantic), Dolphinfin (Mahi Mahi), Grouper, King Crab (RED), Pacific Cod, Red Snapper, Sea Cucumber, Sharks, Shrimp, Swordfish and Tunas (Albacore, Bigeye, Skipjack, Yellowfin, and Bluefin) (Warner et al., 2013).

customers in the meat sector (Bonne & Verbeke, 2008; Premanandh, 2013). Consumers are then forced to rely on authenticity cues such as price, country of origin, and security package labels via certification to make informed food purchasing decisions (El Benni, et al., 2019; Ortega et al., 2014). Indeed, previous studies have noted a consumer preference for domestic finfish over imported seafood due to concerns about potential mislabeling (Garlock et al., 2020; Marko, et al., 2004).

Fraud events linked to one product can have spillover effects on an entire market. While studying the effect of fraud on the olive oil market, Meerza & Gustafson (2020) found evidence suggestive of a negative spillover effect. Specifically, the authors note that exposure to information about Italian olive oil fraud negatively impacted U.S. consumer demand for both U.S. and Greek EVOO. Of the handful of studies eliciting consumer demand for food products under the risk of fraud, Meerza & Gustafson (2020) constitute one of a few with some application to U.S. consumers. Our paper builds on the idea of spillover effects as described by Meerza & Gustafson (2020) to examine whether the premium for local seafood drops or perhaps increases with knowledge about seafood fraud possibly initiated overseas. That said, we deviate from prior research in the following important ways. First, this study focuses on fraud information-induced consumer demand response in the context of seafood, making it the first to do so to our knowledge. Second, we explore a richer definition of local by considering both Great Lakes produced, and other U.S. states sourced seafood products. Finally, we explore potential heterogeneous effects in consumer response to food fraud information published by the media.

1.3 Mapping Fraud in Seafood Supply Chains

We base our conceptual framework on the potential vulnerability of seafood to fraud at various stages of the supply network. Figure 1.3 maps the supply chain for seafood from source to

the final consumer. We present a simplified version of the U.S. seafood supply chain and reference existing or emerging avenues for a range of common fraudulent activities along the chain. We adapt the comprehensive seafood supply chains depicted in Fox et al. (2018), wherein key stages of the supply networks for finfish, shellfish, and crustaceans are laid out in succession. The length and complexity of seafood supply chains follow directly from the production method (that is, via aquaculture or wild-caught production) and, at an extra level of granularity, the species. This is particularly important for our study as global trade includes both wild-caught and aquaculture, which account for 54 and 46 percent of the global production volume, respectively. For instance, in aquaculture production, eggs and fishmeal procured through either domestic or international sources represent critical inputs in the production process, whereas naturally occurring juvenile fish in the wild already have the requisite conditions for survival.

Seafood supply networks are susceptible to a variety of fraudulent activities (Kroetz et al., 2020), with fraud manifesting in diverse ways at multiple levels of the value chain.^{2 3} For example, species substitution can take many forms and occur at any stop in the seafood supply network. For upstream supply chain actors such as fishers and farmers, post-harvest species substitution as fish are held in storage units awaiting further processing can be tempting if there are substantial gains to be made. For instance, the practice whereby high-value species are substituted for low-value species yet sold at a premium is a common occurrence of seafood fraud (Fox et al., 2018; Reilly, 2018). It is also common for high-value species to be misclassified as low-value for tax evasion purposes (Reilly, 2018). Further down the supply chain, the detection of species substitution may

² Examples of these fraudulent practices include species substitution, mislabeling, short weighting, adulteration, and indiscriminate antibiotic use, among others.

³ In a move to comprehensively capture other lesser-known fraudulent opportunities, Fox et al. (2018) extends the scope of seafood fraud to include modern day slavery and animal welfare infractions. For the purposes of this study, we examine seafood fraud outside of these latter ethical considerations.

be complicated due to processing where the morphological identification of species becomes infeasible after they are transformed into fish sticks, fillets, and other pre-prepared fish meals (Marko et al., 2004; Chen, et al., 2014; Fox et al., 2018; Reilly, 2018).

In the processing and distribution segment of the seafood supply network, a unique form of seafood fraud referred to as “short-weighting” can also occur. This involves the overglazing or overbreeding of seafood products to artificially inflate their true weight for economic gain (Reilly, 2018). More recently, Asche et al. (2022) revealed a major discrepancy between Chinese exports, on one hand, and imports plus domestic seafood production numbers, on the other, suspected to result from certain forms of mislabeling and “short-weighting”. A typical example of this fraudulent practice involves the addition of glaze water to frozen seafood products during processing. While seafood species, such as crab, salmon, trout, and halibut, primarily marketed as fresh will likely be less subject to this form of fraud, other predominantly frozen seafood, such as tilapia and shrimp will be common targets (Love, et al., 2022). Relatedly, seafood adulteration with carbon monoxide to enhance fish flesh appearance during frozen storage is just as prevalent although such practices ought to be declared on fish product labels in compliance with most national food safety protocols (Reilly, 2018). Seafood may also be adulterated with antibiotics either directly or indirectly through fish feed to improve production efficiency and fish quality (Fox et al., 2018).

The majority of food fraud investigations are conducted downstream with several studies uncovering seafood fraudulent behavior among retailers (Warner et al., 2013; Fox et al., 2018; Reilly, 2018). Although not entirely obvious whether these actors are themselves victims of seafood fraud initiated higher up the supply chain, surveillance reports have returned some

alarming results. For example, DNA tests of 1,200 seafood samples across 674 retail outlets within the United States revealed that a third of the tested samples were substituted (Warner et al., 2013).

Beyond species substitution, mislabeling and misleading production claims are equally endemic. While seafood might be initially labeled correctly by name, the product may be mislabeled later as wild-caught when it was in fact farmed. This form of mislabeling can occur at any point along the seafood marketing chain, but most frequently occurs among distributors and final seafood retailers such as restaurants and fishmongers (Jacquet & Pauly, 2008). A review of recently published reports on seafood fraud by Pardo et al. (2016) indicated that 30 percent of DNA-tested seafood product samples were mislabeled, with a majority occurring in the food services sector. While mislabeling may not necessarily lead to outright food safety issues, other considerations on sustainability grounds cannot be ignored as such practices could also exacerbate current challenges with depleting fish stocks (Asche et al., 2022; Kroetz et al., 2020).

1.4 Consumer Preferences under Fraud Uncertainty

In this section, we develop a random utility model of heterogeneous consumers with preferences for locally sourced seafood in the face of fraud (McFadden, 1973). Suppose a consumer n derives utility U_{njt} from purchasing seafood alternative j in choice situation t such that:

$$U_{njt} = X'_{jt}\beta_{nj} + \epsilon_{njt} \quad (1)$$

where X'_{jt} denotes the product attributes of alternative j , β_{nj} are coefficients to be estimated, and ϵ_{njt} is the unobserved component of utility that is *independent* and *identically* Gumbel *distributed*.

We then allow the observable component to have the following structure:

$$\begin{aligned}
X'_{jt}\beta_{nj} = & ASC^j + \beta_{Pr}Price_{jt} + \sum_{k \in K \setminus \{GL, US\}} (\bar{\beta}_k + \sigma_k Z_{nk}) x_{jt} + (\bar{\beta}_{GL} + \sigma_{GL} Z_{n,GL}) GL_{jt} \\
& + (\bar{\beta}_{US} + \sigma_{US} Z_{n,US}) US_{jt}
\end{aligned} \tag{2}$$

where ASC^j is an alternative specific constant for alternative j , $Price_{jt}$ is the price, K is a set of experimentally-designed non-price attributes, Z_{nk} is a standard normal random variable, x_{jt} is a $(|K| - 2) \times 1$ vector⁴ of observable product attributes of alternative j , GL_{jt} and US_{jt} denote the seafood alternative's place of origin (that is, the Great Lakes and other US states, respectively)⁵, $\bar{\beta}_k$ is the mean of attribute k 's parameter estimate while σ_k is the standard deviation of the distribution around this mean. Under the scenario with no seafood mislabeling, let us define the mean marginal willingness-to-pay between the *local* and *imported* attribute levels as follows:

$$mWTP_{GL,IM} = -\frac{\bar{\beta}_{GL}}{\beta_{Pr}} \tag{3}$$

$$mWTP_{US,IM} = -\frac{\bar{\beta}_{US}}{\beta_{Pr}} \tag{4}$$

where $mWTP_{GL,IM} > 0$ indicates that the *GL* attribute level attracts a higher premium or a lower discount relative to the *imported* attribute level, on average.

Now, suppose there is a mislabeling information shock such that there exists a non-zero probability $\pi \in (0,1)$ that seafood alternative j is mislabeled. Then consumer n 's expected utility becomes:⁶

⁴ Where $|K|$ denotes the cardinality of set K .

⁵ Where the *GL* and *US* attribute levels are expressed relative to the omitted category, *IM* representing imports.

⁶ Notice that implicit in the representation of the expected utility function is the assumption that the consumer is risk-neutral. The additive structure of the observable component of the utility function renders this assumption.

$$EU_{njt} = (1 - \pi)ASC^j + \pi ASC^l + \beta_{Pr}Price_{jt} + \dots + \sum_{g \in \{GL, US\}} \{[(1 - \pi)\bar{\beta}_g + \pi\bar{\beta}'_g] + [(1 - \pi)\sigma_g + \pi\sigma'_g]Z_{n,g}\}g_{jt} + (1 - \pi)\epsilon_{njt} + \pi\epsilon'_{njt} \quad (5)$$

where $j \neq l$ and ASC^l denotes the alternative specific constant under mislabeling. Analogously, $\tilde{\beta}_g \equiv (1 - \pi)\bar{\beta}_g + \pi\bar{\beta}'_g$ for all $g \in \{GL, US\}$ is defined as the estimated parameter denoting the place of origin attribute level g under mislabeling uncertainty.⁷ Taken together, we obtain the following expression for the mean marginal willingness-to-pay between the *GL* and *imported* attribute levels, for example, in the seafood fraud information setting:

$$mWTP_{GL,IM}^{Treat} = -\frac{\tilde{\beta}_{GL}}{\beta_{Pr}} \quad (6)$$

We hypothesize that consumers' willingness-to-pay for locally produced versus imported seafood partly depends on beliefs of food fraud risk in the respective supply chains. Hence, we expect consumers' preferences for the *local* relative to the *imported* place of origin attribute levels to vary under differing informational settings. We use the theoretical model above to examine two opposing hypotheses regarding the effect of seafood fraud information on consumer preferences for locally produced as opposed to imported food fish. We term these competing hypotheses (1) *the spillover effect*, and (2) *the signaling effect*.

First, the *spillover effect* posits that coupling information about the dominance of imports in U.S. seafood markets with knowledge of fraudulent behavior will reduce the premium for locally produced seafood products. Toledo & Villas-Boas (2019) and Meerza & Gustafson (2020) present findings consistent with this assertion. In these studies, the authors argue that

⁷ It is important to note the distinction between uncertainty and risk. Risk will suggest that the probability of seafood mislabeling is known whereas uncertainty suggests otherwise.

contaminating or negative spillover effects could prevail even if food safety or fraud disproportionately affects a specific product or source and not others. For example, while studying consumer egg purchasing responses to recalls during the 2010 Salmonella outbreak, Toledo & Villas-Boas (2019) observed that consumers also reduced egg purchases from unaffected stores due to the outbreak. These findings suggest that unfavorable food fraud news can result in negative spillover effects:

$$H_{spillover}: mWTP_{g,IM}^{Treat} - mWTP_{g,IM} < 0 \quad (7)$$

where $g \in \{GL, US\}$ indicates local seafood varieties and *Treat* denotes the scenario under which consumers are subjected to unfavorable seafood fraud news.

Second, *the signaling effect* hypothesizes that given general information on seafood fraud, the indication of origin might signal to consumers that they can trust food product quality or safety if consumers associate localness to stricter food safety standards or more effective surveillance and monitoring. In other words, consumers worried about food fraud may perceive local products as less likely to be subject to fraud. This perception could be born out of the strong association of lengthy food supply chains with more opportunities for fraud (Theolier et al., 2021). The core assumptions of this hypothesis are consistent with findings in studies such as Umberger et al. (2003) and Loureiro & Umberger (2003, 2005), who note that most consumers who preferred country of origin labels interpreted these labels as providing additional food safety guarantees. The signaling effect implies that unfavorable food fraud news can benefit products with a shorter supply chain:

$$H_{signaling}: mWTP_{g,IM}^{Treat} - mWTP_{g,IM} > 0 \quad (8)$$

where $g \in \{GL, US\}$.

1.5 Methods

We utilize a discrete choice experiment (DCE) to estimate consumer demand for seafood and investigate the effect of seafood fraud information on WTP for local, domestic, and imported seafood products. DCEs have been extensively used to ascertain consumer preferences for food product attributes in similar settings (Tonsor et al., 2009; Olynk et al., 2010; Ortega et al., 2011, 2014). In our application, the product attributes represent a bundle of characteristics including the species, place of origin, production method, and form of processing.

Table 2.1 lists the species, attributes, and attribute levels in our DCE. The first attribute, seafood species, includes two species popular with U.S. consumers (salmon and trout) and one species popular with consumers around the Great Lakes (whitefish). These species were selected following a pilot survey of extension scientists and educators in the region which elicited their opinions about seafood product characteristics. The other attributes include price, place of origin, production method, and the form of processing. These attributes were selected as our pilot survey identified them as the attributes consumers mostly think about when making seafood purchasing decisions. Price levels ranging from \$7.99 to \$13.99 were selected based on retail prices (per 8oz fillets) in major grocery stores in the Great Lakes Region. In all, four price levels were considered, as well as three places of origin (Great Lakes, United States⁸, and imported), three production methods (wild-caught, farmed/aquaculture, and unlabeled), and two forms of processing (fresh and frozen) labels.

Figure 1.4 shows an example of the choice questions presented to respondents. Consumers in each choice task were asked to select among three alternative profiles of fish along with a no-buy option. A full factorial experimental design would require 373,248 ($2^{3 \times 1} \times 3^{3 \times 2} \times 4^{3 \times 1}$)

⁸ United States represents any other state outside the Great Lakes Region.

choice tasks. Using an orthogonal fractional factorial design (labeled design), we reduce this number to 36 choice tasks. The 36 choice tasks were then blocked into three segments of 12 choice questions each to reduce the number of treatment combinations presented to any one participant (Stopher & Hensher, 2000; Louviere, 2004; Hensher et al., 2005; Caussade et al., 2005). Thus, each participant is faced with only 12 choice questions with three product alternatives and a no-purchase option in the final design (D-error of 0.04). The order in which consumers answered the choice questions was also randomized to account for possible order effects. We also presented a cheap talk script at the beginning of the DCE section of the survey to partly mitigate potential hypothetical bias in our WTP estimates (Lusk & Schroeder, 2004).

A randomly selected half of the respondents were provided a news article excerpt describing the results of a recent food fraud investigation, which also explained that a considerable share of domestically consumed seafood originated from international sources. A copy of this excerpt is located in the APPENDIX. No such information was presented to the control units.

1.6 Empirical Strategy

1.6.1 Latent Class Model (LCM)

We estimate a latent class model (LCM) to capture heterogeneity in consumer preferences by sorting the sample into a finite number of groups or classes. While they offer less flexibility than mixed logit models (MXL), latent class models impose fewer distributional assumptions about random parameters to capture unobserved heterogeneity (Hensher et al., 2005). The model accommodates heterogeneity across latent consumer groups while estimating common parameters for respondents within each group. At its core, the choice probability for each class is derived from estimating a multinomial logit model. Thus, conditional on assignment to class s , the probability

that individual n chooses alternative j while faced with a choice among J alternatives in choice situation t is expressed as:

$$P_{nt|s}(j) = \frac{\exp(X'_{njt}\beta_s)}{\sum_{i=1}^J \exp(X'_{nit}\beta_s)} \quad (9)$$

where X denotes a vector of select food product attributes, and β_s is a vector of parameters to be estimated common to all members in class s . Analogously, the prior probabilities for class membership for individual n can be specified as:

$$C_{ns} = \frac{\exp(Z'_n\omega_s)}{\sum_{s=1}^S \exp(Z'_n\omega_s)} \quad (10)$$

where $s \in \{1, 2, 3, \dots, S\}$, with Z_n denoting a set of observable covariates factored into modeling the class membership probabilities. Under the independence of choice tasks assumption given class assignment, the log-likelihood for the entire sample is defined as:

$$\ln L = \sum_{n=1}^N \ln \left[\sum_{s=1}^S C_{ns} \left(\prod_{t=1}^T P_{nt|s} \right) \right] \quad (11)$$

with the vector of parameters (including the latent class membership parameters) estimated via the conventional maximum likelihood estimation methods (Hensher et al., 2015).

In what follows, we estimate the effect of the information treatment on consumer preferences for the *localness* attribute levels by latent classes. In doing so, we estimate the following model:

$$U_{njt,s} = ASC_s^{j*} + \beta_{s,Pr}^* Pr_{jt} + \beta_{s,GL}^* GL_{jt} + \beta_{s,US}^* US_{jt} + \beta_{s,FR}^* FR_{jt} + \beta_{s,WC}^* WC_{jt} + \beta_{s,FAR}^* FAR_{jt} + \alpha_{1,s} GL_{jt} \times Treat + \alpha_{2,s} US_{jt} \times Treat + \varepsilon_{njt,s} \quad (12)$$

where $Treat$ is an indicator variable which takes the value 1 if the respondent was presented the news article excerpt, and 0 otherwise; the estimated coefficients on the interaction terms, $\alpha_{1,s}$ and $\alpha_{2,s}$ denote the difference in the preferences for the respective place of origin attribute levels over treatment status for consumer n in market segment s ; ASC_s^j denotes the alternative specific constants representing salmon, whitefish, and trout, with the constant of the no purchase option set to 0; Pr_{jt} is a continuous price variable representing each of the four price levels considered in the study; GL_{jt} and US_{jt} constitute indicator variables for the experimentally-designed place of origin attribute (the Great Lakes region and the United States, respectively), whose coefficients are interpreted relative to the *imported* attribute level; WC_{jt} and FAR_{jt} are dummy variables which take the value 1 if the seafood product carries the wild-caught and farmed production method labels, respectively and 0 otherwise, with the estimated coefficients expressed in relation to the no label attribute level; FR_{jt} is a dummy variable which takes the value 1 if the seafood product is fresh, and 0 if frozen; $\beta_{s,\cdot}^*$ represents the non-stochastic estimated parameter coefficients and $\varepsilon_{njt,s}$ is the unobserved *independent and identically Gumbel distributed* error term.

1.6.2 Mixed Logit Model (MXL)

We also consider an alternative approach to capturing consumer preference heterogeneity by estimating a mixed logit model (MXL).⁹ We estimate a MXL model separately for the treatment and control group using 500 Halton draws. Parameter coefficients of the experimentally designed attributes are assumed to follow a normal distribution and are estimated in WTP space. That is, we reparameterize the MXL models such that WTP measures are directly estimable. This approach has been generally recommended in the discrete choice experiment literature for producing more

⁹ In the rest of the paper, we use the terms mixed logit (MXL) and random parameters logit (RPL) interchangeably.

reliable WTP estimates (Train & Weeks, 2005; Scarpa et al., 2008; Train, 2009). The price coefficients and the alternative specific constants, however, were assumed to be non-stochastic. We estimate these models using the simulated maximum likelihood estimation technique (Train, 2009).

1.7 Data and Descriptives

Survey data were collected online in early September 2021 in collaboration with market research and data collection company Qualtrics. The survey targeted consumers in the Great Lakes region including associated ceded territories of Tribal Nations, who were over the age of 18, were the primary shoppers for food in their respective households and had purchased seafood in the past year. The survey consisted of socio-demographic questions, as well as queries on household seafood purchasing and consumption behaviors. We restrict our analyses to the 1,272 consumers who completed the entire survey, resulting in an equal assignment of respondents to the treatment and control groups.

Tables 1.2 and 1.3 present descriptive statistics on key socio-demographic variables for the sampled participants. In particular, Table 1.2 provides summary statistics by treatment status and a balance test, while Table 1.3 compares our sample to census and other nationally representative survey data. For comparability, we consider census data from the 2021 American Community Survey (ACS) and the 2017-2018 National Health and Nutrition Examination Survey (NHANES).¹⁰ Table 1.3 shows broad agreement between our sample demographics and the Great Lakes region-specific census data. However, important deviations include an oversampling of individuals with more education and persons aged 35-44. Most of the sampled respondents are middle-aged (35-44 years old) with at least some college education (78%).

¹⁰ The ACS estimates are restricted to the adult population residing in the Great Lakes region, while the NHANES applies to the US population subgroup who indicated seafood consumption in the past 30 days.

Table 1.4 provides descriptive statistics on seafood consumption variables across treatment groups. The average respondent typically consumed seafood at home, 2-3 times a month, and preferred wild-caught seafood. To ascertain consumers' *ex ante* level of seafood fraud concern, we posed a choice question on a 0-100 numeric sliding scale, with 100 indicating the maximum level of concern. On average, respondents were moderately concerned about seafood fraud (55.2).¹¹ Reassuringly, a balance test revealed that there appear to be no statistically significant differences in observable respondent characteristics across the information treatment and control groups.¹² The only exception is the household size variable, which revealed a statistically significant difference at the 5% level.

1.8 Econometric Results

1.8.1 Market Segments

Table 1.5 presents the LCM results. We focus our discussion on the LCM with four distinct classes.¹³ The estimates indicate that classes generally differ from one another by the degree of preference for *localness*. To this end, we label these classes such that consumers fall into one of the following groups: (1) *Locavores*, (2) *COO* (where COO denotes country of origin), (3) *Information-sensitive* (hereafter, referred to as *IS*), and (4) *Price-sensitive* (hereafter, *PS*) groups. For instance, *locavores* and respondents in the *COO* group find the locality attributes relevant to their seafood choices but differ in their relative preference for the *GL* and the *US* attribute levels.

¹¹ One might be concerned about potential anchoring bias in consumers' reported levels of seafood fraud concern due to non-randomization of the slider position between subjects, *ex ante*. Given that the slider was positioned at 50 by default, we conduct a simple t-test of the null that the average level of concern is not different from 50. We reject this null in favor of the alternative at the 1% level ($p - value < 0.0001$), indicating that the consumers' average level of concern is significantly different from 50.

¹² Balance test results are presented in Tables 1.2 and 1.4.

¹³ While the Akaike and Bayesian Information Criteria indicate that extending the number of "latent" classes does improve the model fit, doing so yielded unwieldy results and overcomplicates model interpretation. In particular, the estimated standard errors of some coefficients were substantially large; in part, because of the small number of observations assigned to some classes (Heckman & Singer, 1984).

As the group names infer, *locavores* indicate a stronger preference for seafood produced in the Great Lakes region relative to seafood produced in other parts of the United States. The opposite is true for the *COO* group. Additional characteristics representative of each group are presented in Table 1.6. As the table indicates, approximately less than half (48%) of respondents in the *locavore* group identified as female, while females are overrepresented across the remaining classes. The *locavore* and *IS* groups mostly consist of younger consumers, whereas the *COO* and *PS* groups are over-representative of participants aged 65 years and older. We also observe that at least 58% of the sampled respondents across all classes consume seafood at home. Table 1.6 also shows that the price-sensitive group is the least concerned about seafood fraud, while the remaining classes report average levels of concern close to the full sample mean of 55.2. Market shares for each latent class of consumers are 52% (*locavores*), 25% (*COO*), 11% (*IS*), and 12% (*PS*).

Results for the *locavore* latent class reveal that the coefficients on the *GL* and *US* attribute levels are positive and statistically significant. That is, consumers in this group strongly favor the two *local* place of origin labels. This finding is in line with conclusions drawn in previous work (Davidson et al., 2012; Fonner & Sylvia, 2015; Brayden et al., 2018). The results also show that relative to the *imported* label, the presence of the *GL* attribute label induced a higher utility increment relative to the *US* label. The reverse is true for the *COO* group. Both *IS* and *PS* groups, however, do not appear to prefer either the *GL* or *US* attribute labels. We also find support for disutility for price increments across all four classes, consistent with consumer demand theory.

While *locavores* and consumers in the *COO* group indicated a strong preference for the fresh label, no statistically significant result was reported for the other groups. Perhaps for these consumer segments, seafood bearing a fresh label indicates a lower likelihood of being imported (Campbell et al., 2014; Fonner & Sylvia, 2015). In other results, consumers in the *IS* group

exhibited a positive and statistically significant preference for any of the production method claims (either wild-caught or farmed) relative to no label. However, this was not the case for the *PS* group. The alternative specific constants also suggest stark differences across the respective classes. While *locavores* demonstrate a strong preference for the purchase alternatives (that is, salmon, trout, and whitefish) to the no-buy option, such a preference is absent across the remaining classes. In fact, in some cases, consumers exhibited a strong preference for the no-buy option (for example, the *COO* group). Estimates for the *IS* group deviate slightly from this pattern, with consumers appearing to prefer the salmon alternative to the no purchase option. To some extent, while we do not collect self-reported attribute nonattendance (ANA) data, ANA behavior can be inferred from our LCM estimates. For instance, the *PS* group appears to ignore all the non-price attributes, while the *IS* consumer segment does not attend to any of the places of origin attribute labels.

We report marginal willingness-to-pay (mWTP) estimates for the LCM, which is expressed as the ratios of the corresponding coefficients of the attributes of interest and the price coefficient. Linearity of the different attributes and the price variable in the indirect utility specification yields the following expression for the marginal WTP for attribute k for a given latent class s :

$$WTP_{k,s} = -\frac{\beta_{s,k}}{\beta_{s,Pr}} \quad (13)$$

where the corresponding asymptotic standard errors of these ratios are estimated using the Delta method with 10,000 draws. Table 1.7 reports these marginal WTP estimates for each class with the corresponding 95% confidence intervals.

The mWTP estimates across all attributes for consumers in the *PS* group are not statistically significant, indicating that consumers in this class do not place significant importance on any of the production method, place of origin, or processing form attributes represented in the

survey. By contrast, the *WILD* and *FARM* attribute levels generate a positive WTP for respondents in the *IS* group though we do not find any significant premiums across the remaining attributes for this group. Turning to the *COO* group, we find that the *US* attribute label carries a higher premium of \$4.63 per 8oz fillets of seafood relative to the *GL* label, which carries a mWTP of \$3.62 per 8oz fillets. By contrast, *locavores* have a higher mWTP for the *GL* place of origin label followed by the *US* label (mWTP estimate of \$6.38 vs \$5.71 per 8oz fillets of seafood). Within the *COO* group, consumers exhibited the highest marginal willingness-to-pay for the wild-caught label with a mWTP estimate of \$9.76 per 8oz fillets of seafood. These consumers also indicated positive mWTP values of \$3.28 and \$1.24 per 8oz fillets for the farmed and fresh attribute labels, respectively. Similarly, the presence of labels denoting that seafood was wild-caught, fresh, and farmed are associated with positive and statistically significant mWTP estimates of \$4.68, \$4.92, and \$3.13 per 8oz fillets of seafood, respectively for the *locavore* group.

1.8.2 Information Treatment Effects for LCM

To test for a differential effect of the treatment on consumers' preferences for the locality attribute labels, we estimate equation (10). The estimated coefficients on the interaction terms, $GL \times Treat$ and $US \times Treat$ capture the changes in utility for *localness* given the information shock. Results are presented in Table 1.8. Our results generally indicate that consumer preferences for the *GL* and *US* attributes were not significantly altered by information about fraud for any of the consumer groups except the *IS* group. For this sub-group, the information shock appears to have eroded any positive premium for the *US* attribute label, which is consistent with the *spillover effect* hypothesis.

Further, to get a sense of whether the *signaling* or *spillover effect* holds for the entire sample, we derive full sample differences in mWTP estimates for the *local* relative to the *imported*

attribute labels due to the information shock. Results are reported in Table 1.9. Using the estimated class probabilities as weights, we obtain weighted average changes in mWTP for the *local* attribute labels given unfavorable food fraud news. Our results suggest a decline in the premium for the *GL* in lieu of the *imported* label (due to mislabeling) amounting to 9 cents per 8oz fillets; however, this effect is not statistically significant. More strikingly, we find that the information shock resulted in a \$1.96 per 8oz fillets reduction in the average mWTP for the *US* relative to the *imported* attribute label across the entire sample. This effect is statistically significantly different from zero at the 10 percent level. Taken together, the evidence provided in Table 1.9 offers some support for the *spillover effect* for US-label seafood products, though not GL-labeled products.

1.8.3 MXL Model Results

Next, we obtain and plot the distribution of individual-specific conditional WTP estimates for the *US* and *GL* attribute levels across treatment status from the MXL model. These respondent-specific estimates are essentially means of the conditional distribution of the WTP parameter estimates, where we condition on the choices we observe the respondents make (Hensher et al., 2015). The distribution of these WTP estimates are shown in figures 1.5 and 1.6 for the *GL* and *US* attribute levels, respectively. As can be seen, the treatment appears to have eroded premia across the two levels but more so for the *US* attribute level. Interestingly, we also observe a bunching of the WTP estimates around the median with respect to the *US* attribute level for the treatment group.

We then present the MXL model estimates on all attributes featured in the choice experiment across treatment arms. Results are reported in Table 1.10. First, as consumer demand stipulates, the price coefficients are negative and statistically different from zero at all conventional levels of significance across the treatment and control groups. We also observe that consumers are

willing to pay a premium of \$1.58 and \$1.96 per 8oz fillets for the *GL* and *US* attribute levels, respectively relative to the *imported* label for the control group. These WTP estimates are both significant at the 5 percent level or better. Nonetheless, we observe a plummeting of these WTP values with treatment (that is, the values tend closer to zero). Specifically, the mean marginal willingness-to-pay for the *GL* and *US* attribute levels dropped to \$1.26 and \$1.50, respectively following the information shock, providing some evidence in favor of a negative *spillover effect*.

In other results, the treatment appears to have induced a positive and significant premium for farmed fish of \$0.68 per 8oz fillets. The premium for the farmed attribute level was marginally distinguishable from zero in the control group. That said, we do observe significant preference heterogeneity for this attribute level within this sub-group. In particular, the mean and standard deviation of the WTP estimates suggest that roughly 55% of consumers in the control group have a positive marginal WTP for farmed fish relative to the no label option. Likewise, we observe significant preference heterogeneity for the other non-price experimentally designed attributes for both treatment and control groups. However, we do notice that the standard deviation on the *US* attribute level tends toward zero and is no longer statistically significant after treatment exposure, suggesting that the treatment homogenizes consumers in terms of their WTP for the *US* label.

We also test whether preferences across the treatment and control groups are the same using a likelihood ratio test of equality of WTP and parameter coefficients across the two groups. In doing so, we follow the approach set forth by Layton & Brown (2000) by pooling across the two models (that is, across the treatment and control groups) and conducting the likelihood ratio test. We document from this test that we can reject the null hypothesis that preferences can be restricted to be the same across the treatment and control groups, with a likelihood ratio statistic of 391.94 against a $\chi^2_{16,0.05}$ critical value of 26.30 (at the 5 % level of significance).

1.8.4 Market Shares Estimation

Next, we investigate the impact of the treatment on predicted market share estimates for each of the seafood alternatives with all prices fixed at \$10.99 per 8oz fillets. We follow Lusk & Tonsor (2016) and Van Loo et al. (2020) by estimating a RPL model with the systematic component expressed as follows:

$$\hat{V}_{nj} = \left(\hat{\beta}_j + \sum_{k=1}^3 \hat{\sigma}_{jk} z_{nk} \right) + \hat{\alpha}_j Pr_j \quad (14)$$

where z_{nk} has a standard normal distribution; $\hat{\beta}_j$ is the alternative specific constant for seafood alternative j ; $\hat{\sigma}_{jk}$ denotes the lower triangular Cholesky decomposition for the variance-covariance matrix of the random parameters with the off-diagonals set to equal zero (that is, $\hat{\sigma}_{jk} = 0$ for $j \neq k$) (Lusk & Tonsor, 2016). That is, the seafood alternative specific constants are assumed to be independently distributed. We then substitute equation (12) into a multinomial logit formula to derive the estimated market shares for the seafood alternatives. We approximate the mean market shares using simulations with a set of 5,000 draws for z_{nk} . Results are reported in Table 1.11.

As the table shows, the predicted unconditional market shares for salmon, trout, and whitefish are 34%, 28%, and 29%, respectively in the absence of the treatment. Following the information shock, whitefish become the species with the lowest market share (17%). Interestingly, the share of consumers who prefer the no-purchase option falls from 9% to 6% after the treatment. By contrast, the estimated market shares for salmon, and trout increase to 39% and 38%, respectively. That is, the information shock appears to have negatively impacted the choice share for whitefish, while a favorable effect of the treatment was observed for salmon and trout. Results from the conditional market share estimates reiterate these findings.

1.9 Conclusions

Increasingly globalized food supply chains create added opportunities for fraud, which is likely to influence consumer behavior. Indeed, no food product is immune to potential food fraud risk. As one of the most targeted food fraud categories, seafood that is domestically sourced could as well suffer consequences for fraudulent activities initiated elsewhere. In this paper, we examine consumers' risk mitigating responses as reflected in their valuation of localness when making seafood purchasing decisions in the face of fraud. Using a between-sample approach, we randomize respondents into differing informational settings to investigate whether consumer WTP for local seafood is impacted by a seafood fraud information shock.

Our results indicate that, consumers broadly derive positive utility from consuming locally sourced seafood (that is, seafood produced in the Great Lakes Region or other states within the United States). Upon further disaggregation, however, we find that for more price-sensitive market segments, localness does not command a significant positive premium. Further, we demonstrate that providing information regarding fraud is unlikely to significantly alter preferences toward the local options across most consumer market segments. That said, we find that the information shock resulted in a \$1.96 per 8oz fillets decline in the willingness-to-pay for *US*-labeled seafood.

In other results, we also show that the intervention disproportionately affected market shares for certain seafood species relative to others. Specifically, the predicted market share for whitefish recorded the largest drop with exposure to unfavorable seafood fraud information, with salmon and trout experiencing an uptick in market share following the treatment. An investigation into the mechanisms driving such differences in consumer response across the various seafood species is beyond the scope of this article. However, for producers and marketers of whitefish, a deeper dive into possible explanations for such consumer risk mitigating behavior can be of value.

As a note of caution, the fact that we do not find overwhelmingly compelling evidence in support of either the *spillover* or *signaling effect* for most consumer segments does not suggest that seafood fraud is not of concern. First, we must point out that we consider a specific form of seafood fraud (that is, mislabeling) in this study. To the extent that other fraudulent activities such as indiscriminate antibiotic use, short weighting, adulteration, among others, stir stronger consumer demand response, our results are not generalizable. Second, our results do not necessarily suggest that consumers will not attach a significant positive premium to product integrity assurances in the form of “food fraud-free” certification labels. For different actors along the seafood supply chain, innovations in seafood DNA testing and authentication are fast emerging. Nonetheless, whether such labeling features will be economically worthwhile remains to be seen and calls for further research.

BIBLIOGRAPHY

- Abaidoo, E., Melstrom, M., & Malone, T. (2021). The Growth of Imports in US Seafood Markets. *Choices*, 36(4), 1-10.
- Asche, F., Yang, B., Gephart, J. A., Smith, M. D., Anderson, J. L., Camp, E. V., . . . Straume, H.-M. (2022). China's seafood imports-Not for domestic consumption? *Science*, 375(6579), 386-388.
- Belton, B., Reardon, T., & Zilberman, D. (2020). Sustainable commoditization of seafood. *Nature Sustainability*, 3(9), 677-684.
- Bitzios, M., Lisa, J., Krzyzaniak, S.-A., & Mark, X. (2017). Country-of-origin labelling, food traceability drivers and food fraud: Lessons from consumers' preferences and perceptions. *European Journal of Risk Regulation*, 8(3), 541-558.
- Bonne, K., & Verbeke, W. (2008). Religious values informing halal meat production and the control and delivery of halal credence quality. *Agriculture and Human Values*, 25(1), 35-47.
- Brayden, W. C., Noblet, C. L., Evans, K. S., & Rickard, L. (2018). Consumer preferences for seafood attributes of wild-harvested and farm-raised products. *Aquaculture Economics & Management*, 22(3), 362-382.
- Campbell, L. M., Boucquey, N., Stoll, J., Coppola, H., & Smith, M. D. (2014). From vegetable box to seafood cooler: applying the community-supported agriculture model to fisheries. *Society & Natural Resources*, 27(1), 88-106.
- Caussade, S., de Dios Ortúzar,, J., Rizzi, L. I., & Hensher, D. A. (2005). Assessing the influence of design dimensions on stated choice experiment estimates. *Transportation research part B: Methodological*, 39(7), 621-640.
- Chan, E., Griffiths, S., & Chan, C. (2008). Public-health risks of melamine in milk products. *The Lancet*, 372(9648), 1444-1445.
- Chen, S., Zhang, Y., Li, H., Wang, J., Chen, W., Zhou, Y., & Zhou, S. (2014). Differentiation of fish species in Taiwan Strait by PCR-RFLP and lab-on-a-chip system. *Food Control*, 44, 26-34.
- Davidson, K., Pan, M., Hu, W., & Poerwanto, D. (2012). Consumers' willingness to pay for aquaculture fish products vs. wild-caught seafood--A case study in Hawaii. *Aquaculture Economics & Management*, 16(2), 136-154.
- El Benni, N., Stolz, H., Home, R., Kendall, H., Kuznesof, S., Clark, B., . . . Chan, M.-Y. (2019). Product attributes and consumer attitudes affecting the preferences for infant milk formula in China-A latent class approach. *Food Quality and Preference*, 71, 25-33.
- FAO. (2008). *Food safety and quality - Melamine*. Food and Agriculture Organization.

- FAO. (2020). *The State of the World Fisheries and Aquaculture: Sustainability in Action*. Food and Agriculture Organization.
- Fonner, R., & Sylvia, G. (2015). Willingness to pay for multiple seafood labels in a niche market. *Marine Resource Economics*, 30(1), 51-70.
- Fox, M., Mitchell, M., Dean, M., Elliott, C., & Campbell, K. (2018). The seafood supply chain from a fraudulent perspective. *Food Security*, 10(4), 939-963.
- Garlock, T., Nguyen, L., Anderson, J., & Musumba, M. (2020). Market potential for Gulf of Mexico farm-raised finfish. *Aquaculture Economics & Management*, 24(2), 128-142.
- Gephart, J. A., Froehlich, H. E., & Branch, T. A. (2019). To create sustainable seafood industries, the United States needs a better accounting of imports and exports. *Proceedings of the National Academy of Sciences*, 116(19), 9142-9146.
- Giannakas, K. (2002). Information asymmetries and consumption decisions in organic food product markets. *Canadian Journal of Agricultural Economics*, 50(1), 35-50.
- Hayes, D. J., Shogren, J. F., Shin, S. Y., & Kliebenstein, J. B. (1995). Valuing food safety in experimental auction markets. *American Journal of Agricultural Economics*, 77(1), 40-53.
- Heckman, J., & Singer, B. (1984). A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica: Journal of the Econometric Society*, 271-320.
- Hensher, D. A., Rose, J. M., & Greene, W. H. (2005). *Applied choice analysis: a primer*. Cambridge University Press.
- Hensher, D. A., Rose, J. M., & Greene, W. H. (2015). *Applied Choice Analysis* (Vol. 2). Cambridge University Press.
- Ingelfinger, J. R. (2008). Melamine and the global implications of food contamination. *New England Journal of Medicine*, 359(26), 2745-2748.
- Jacquet, J. L., & Pauly, D. (2008). Trade secrets: renaming and mislabeling of seafood. *Marine Policy*, 32(3), 309-318.
- Johnson, R. (2014). *Food fraud and economically motivated adulteration of food and food ingredients*. Washington DC: Congressional Research Service.
- Kroetz, K., Luque, G. M., Gephart, J. A., Jardine, S. L., Lee, P., Chicojay Moore, K., . . . Donlan, C. (2020). Consequences of seafood mislabeling for marine populations and fisheries management. *Proceedings of the National Academy of Sciences*, 117(48), 30318-30323.
- Layton, D. F., & Brown, G. (2000). Heterogeneous preferences regarding global climate change. *Review of Economics and Statistics*, 82(4), 616-624.
- Loureiro, M. L., & Umberger, W. J. (2003). Estimating consumer willingness to pay for country-of-origin labeling. *Journal of Agricultural and Resource Economics*, 287-301.

- Loureiro, M. L., & Umberger, W. J. (2005). Assessing consumer preferences for country-of-origin labeling. *Journal of Agricultural and Applied Economics*, 37(1), 49-63.
- Louviere, J. J. (2004). Random utility theory-based stated preference elicitation methods: applications in health economics with special reference to combining sources of preference data. *Centre for the Study of Choice (CenSoC) working paper*(04-001), 22.
- Love, D. C., Asche, F., Young, R., Nussbaumer, E. M., Anderson, J. L., Botta, R., . . . Gephart, J. A. (2022). An overview of retail sales of seafood in the USA, 2017-2019. *Reviews in Fisheries Science & Aquaculture*, 30(2), 259-270.
- Lusk, J. L., & Schroeder, T. C. (2004). Are choice experiments incentive compatible? A test with quality differentiated beef steaks. *American Journal of Agricultural Economics*, 86(2), 467-482.
- Lusk, J. L., & Tonsor, G. T. (2016). How meat demand elasticities vary with price, income, and product category. *Applied Economic Perspectives and Policy*, 38(4), 673-711.
- Marette, S., Roosen, J., & Blanchemanche, S. (2008). Health information and substitution between fish: Lessons from laboratory and field experiments. *Food Policy*, 33(3), 197-208.
- Marette, S., Roosen, J., Blanchemanche, S., & Verger, P. (2008). The choice of fish species: an experiment measuring the impact of risk and benefit information. *Journal of Agricultural and Resource Economics*, 1-18.
- Marko, P. B., Lee, S. C., Rice, A. M., Gramling, J. M., Fitzhenry, T. M., McAlister, J. S., . . . Moran, A. L. (2004). Mislabeling of a depleted reef fish. *Nature*, 430(6997), 309-310.
- McCallum, C., Cerroni, S., Derbyshire, D., Hutchinson, W. G., & Nayga Jr, R. (2022). Consumers' responses to food fraud risks: an economic experiment. *European Review of Agricultural Economics*, 49(4), 942-969.
- McFadden, D. (1973). *Conditional logit analysis of qualitative choice behavior*. Oakland: Institute of Urban and Regional Development, University of California Oakland.
- Meerza, S. I., & Gustafson, C. R. (2020). Consumers' Response to Food Fraud: Evidence from Experimental Auctions. *Journal of Agricultural and Resource Economics*, 45(2), 219-231.
- NOAA. (2021). *U.S. Aquaculture*. Retrieved from NOAA: <https://www.fisheries.noaa.gov/national/aquaculture/us-aquaculture>
- Olynk, N. J., Tonsor, G. T., & Wolf, C. A. (2010). Verifying credence attributes in livestock production. *Journal of Agricultural and Applied Economics*, 42(3), 439-452.
- O'Neill, L., Holbrook, T., & Russell, C. (2015). *Fishy Business: Economically Motivated Adulteration of Fish in Minnesota Retail Markets*. Minneapolis: Food and Agriculture Organization.

- Ortega, D. L., Wang, H. H., & Olynk Widmar, N. J. (2014). Aquaculture imports from Asia: an analysis of US consumer demand for select food quality attributes. *Agricultural Economics*, 45(5), 625-634.
- Ortega, D. L., Wang, H. H., & Olynk Widmar, N. J. (2015). Effects of media headlines on consumer preferences for food safety, quality and environmental attributes. *Australian Journal of Agricultural and Resource Economics*, 59(3), 433-445.
- Ortega, D. L., Wang, H. H., Wu, L., & Olynk, N. J. (2011). Modeling heterogeneity in consumer preferences for select food safety attributes in China. *Food Policy*, 36(2), 318-324.
- Pardo, M. Á., Jiménez, E., & Pérez-Villarreal, B. (2016). Misdescription incidents in seafood sector. *Food Control*, 62, 277-283.
- Premanadh, J. (2013). Horse meat scandal-A wake-up call for regulatory authorities. *Food Control*, 34(2), 568-569.
- Reilly, A. (2018). Overview of food fraud in the fisheries sector. *FAO Fisheries and Aquaculture Circular*(C1165), 1-21.
- Scarpa, R., Thiene, M., & Train, K. (2008). Utility in willingness to pay space: a tool to address confounding random scale effects in destination choice to the Alps. *American Journal of Agricultural Economics*, 90(4), 994-1010.
- Schug, D. (2016). Preventing food fraud. *Food Engineering*, 88(1), 109.
- Shears, P. (2010). Food fraud-a current issue but an old problem. *British Food Journal*.
- Spink, J., & Moyer, D. C. (2011). Defining the public health threat of food fraud. *Journal of Food Science*, 76(9), 157-163.
- Spink, J., Ortega, D. L., Chen, C., & Wu, F. (2017). Food fraud prevention shifts the food risk focus to vulnerability. *Trends in Food Science & Technology*, 62, 215-220.
- Stopher, P. R., & Hensher, D. A. (2000). Are more profiles better than fewer?: searching for parsimony and relevance in stated choice experiments. *Transportation Research Record*, 1719(1), 165-174.
- Theolier, J., Barrere, V., Charlebois, S., & Godefroy, S. B. (2021). Risk analysis approach applied to consumers' behaviour toward fraud in food products. *Trends in Food Science & Technology*, 107, 480-490.
- Toledo, C., & Villas-Boas, S. B. (2019). Safe or not? Consumer responses to recalls with traceability. *Applied Economic Perspectives and Policy*, 41(3), 519-541.
- Tonsor, G. T., Olynk, N., & Wolf, C. (2009). Consumer preferences for animal welfare attributes: The case of gestation crates. *Journal of Agricultural and Applied Economics*, 41(3), 713-730.
- Train, K. E. (2009). *Discrete choice methods with simulation*. Cambridge University Press.

- Train, K., & Weeks, M. (2005). Discrete choice models in preference space and willingness-to-pay space. In *Applications of simulation methods in environmental and resource economics* (pp. 1-16). Springer.
- Uchida, H., Roheim, C. A., & Johnston, R. J. (2017). Balancing the health risks and benefits of seafood: how does available guidance affect consumer choice? *American Journal of Agricultural Economics*, 99(4), 1056-1077.
- Umberger, W. J., Feuz, D. M., Calkins, C. R., & Sitz, B. M. (2003). Country-of-origin labeling of beef products: US consumers' perceptions. *Journal of Food Distribution Research*, 34, 103-116.
- Van Loo, E. J., Caputo, V., & Lusk, J. L. (2020). Consumer preferences for farm-raised meat, lab-grown meat, and plant-based meat alternatives: Does information or brand matter? *Food Policy*, 95, 101931.
- Waite, R., Beveridge, M., Brummett, R., Castine, S., Chaiyawannakarn, N., Kaushik, S., . . . Phillips, M. (2014). *Improving productivity and environmental performance of aquaculture*. WorldFish.
- Warner, K., Timme, W., Lowell, B., & Hirschfield, M. (2013). *Oceana study reveals seafood fraud nationwide*. Washington, DC: Oceana.
- Weir, M. J., Uchida, H., & Vadivelo, M. (2021). Quantifying the effect of market information on demand for genetically modified salmon. *Aquaculture Economics & Management*, 25(1), 1-26.
- Yang, Z., Zhou, Q., Wu, W., Zhang, D., Mo, L., Liu, J., & Yang, X. (2022). Food fraud vulnerability assessment in the edible vegetable oil supply chain: A perspective of Chinese enterprises. *Food Control*, 109005.
- Zilberman, D., Kaplan, S., & Gordon, B. (2018). The political economy of labeling. *Food Policy*, 78, 6-13.

APPENDIX A: TABLES AND FIGURES

Table 1.1

<i>Attributes and Attribute Levels included in the Discrete Choice Experiment</i>	
Product attribute	Levels
Price	\$7.99/8oz fillets
	\$9.99/8oz fillets
	\$11.99/8oz fillets
	\$13.99/8oz fillets
Origin	Great Lakes Region
	United States (but outside the Great Lakes)
	Imported
Processing form	Fresh
	Frozen
Production Method	Wild-caught
	Farmed
	Unlabeled

Table 1.2

<i>Sample Demographics</i>				
Variable	All	Treatment	Control	Diff: p-value
Female (%)	53	53	53	0.96
Age (%)				0.07
18 – 24 years old	7	6	8	
25 – 34 years old	21	22	20	
35 – 44 years old	27	30	24	
45 – 54 years old	12	12	12	
55 – 64 years old	13	13	13	
65+ years old	20	17	23	
Marital Status (%)				0.65
Married	57	58	56	
Divorced	9	8	10	
Separated	2	2	1	
Single, Never Married	28	28	28	
Widowed	5	5	5	
Educational level (%)				0.33
Less than High School	2	2	1	
High School/GED	21	23	19	
Some College	21	20	21	
2-Year College Deg. (Assoc.)	9	9	10	
4-Year College Deg. (BA, BS)	25	25	24	
Master's Degree	19	18	20	
Professional Deg. (Ph.D., J.D., M.D., etc.)	4	3	5	
Number of HH members (%)				0.04
1	19	17	22	
2	32	29	34	
3	19	24	15	
4	20	20	20	
5+	10	10	10	
Annual pre-tax HH income in \$ (%)				0.42
Less than 20,000	12	12	13	
20,000 – 39,999	20	18	22	
40,000 – 59,999	17	19	15	
60,000 – 79,999	15	15	15	

Table 1.2 (cont'd)

80,000 – 99,999	8	8	8	
100,000 – 119,999	7	8	6	
120,000 – 139,999	6	7	5	
140,000 – 159,999	6	7	6	
160,000 or greater	9	8	11	
Region of residence (%)				0.32
Midwest	67	67	67	
Northeast	31	31	31	
South	1	2	1	
West	1	1	1	
Observations	1,272	636	636	

Table 1.3*Overall sample demographics and representability*

Variable	Sample	ACS* (Great Lakes)	NHANES [#]
Female (%)	53	51	52
Age (%)			
18 – 24 years old	7	12	12
25 – 34 years old	21	17	16
35 – 44 years old	27	16	15
45 – 54 years old	12	16	17
55 – 64 years old	13	17	19
65+ years old	20	22	20
Marital Status (%)			
Married	57	48	56
Divorced	9	11	10
Separated	2	1	3
Single, Never Married	28	34	25
Widowed	5	6	6
Educational level (%)			
Less than High School	2	9	9
High School/GED	21	29	26
Some College/2-year college deg.	30	29	31
4-Year College Deg. (BA, BS) and beyond	48	40	34
Number of HH members (%)			
1	19	28	13
2	32	34	35
3	19	16	19
4	20	13	15
5+	10	10	18
Annual pre-tax HH income in \$ (%)			
100,000 or less	72	75	81
> 100,000	28	25	19

Notes: *The ACS estimates are derived from the 2021 American Community Survey 1-Year estimates for survey respondents in the Great Lakes region (Illinois, Indiana, Michigan, Minnesota, Ohio, New York, Pennsylvania, and Wisconsin). [#]The survey weight-adjusted National Health and Nutrition Examination Survey (NHANES) estimates are obtained from the 2017 to 2018 demographics file on US seafood consumers.

Table 1.4*Seafood purchasing behavior*

Variable	All	Treatment	Control	Diff: p-value
Seafood purchase frequency (%)				0.14
Every day	2.44	3.14	1.73	
Two to three times a week	9.21	8.33	10.08	
Weekly	22.27	21.86	22.68	
Two to three times a month	32.10	30.82	33.39	
About once a month	20.06	20.75	19.37	
Less than once a month	12.98	14.15	11.81	
Never	0.94	0.94	0.94	
Seafood consumption location (%)				0.95
At home	67.53	67.61	67.45	
Away from home (e.g., restaurants, bars, etc.)	32.47	32.39	32.55	
Preferred seafood source (%)				0.11
Wild caught	36.34	33.60	39.08	
Farmed	8.87	9.67	8.07	
Indifferent	41.81	43.90	39.72	
Not sure	12.98	12.84	13.13	
Level of seafood fraud concern (%)	55.2 (27.9)	55.6 (27.1)	54.8 (28.7)	0.59
Observations	1,272	636	636	

Notes: Standard deviations are reported in parentheses. P-values from the null hypotheses testing of no difference between treatment and control subgroups are also reported. Seafood level of concern is measured on a 0 – 100 scale where 100 depicts maximum level of concern.

Table 1.5*Latent Class Model Parameter Estimates*

4 Latent Classes, Fixed Parameters				
Variable	Locavores	COO	Information-sensitive	Price-sensitive
	Class 1	Class 2	Class 3	Class 4
<i>ASC</i>				
Salmon	2.749*** (0.128)	-0.078 (0.176)	4.201*** (0.558)	0.475 (0.830)
Trout	2.763*** (0.128)	-0.940** (0.183)	0.071 (0.608)	0.013 (0.866)
Whitefish	2.743*** (0.127)	-0.384** (0.183)	0.030 (0.615)	0.895 (0.856)
<i>PRICE</i>	-0.052*** (0.007)	-0.136*** (0.015)	-0.112** (0.046)	-0.426*** (0.079)
<i>GL</i>	0.334*** (0.035)	0.490*** (0.069)	0.268 (0.218)	0.007 (0.292)
<i>US</i>	0.299*** (0.036)	0.626*** (0.069)	0.324 (0.209)	-0.022 (0.265)
<i>FRESH</i>	0.258*** (0.028)	0.444*** (0.058)	-0.048 (0.196)	0.165 (0.228)
<i>WILD</i>	0.246*** (0.037)	1.322*** (0.079)	1.017*** (0.275)	0.037 (0.277)
<i>FARM</i>	0.164*** (0.036)	0.169** (0.080)	0.869*** (0.268)	-0.154 (0.319)
Class prob.	0.524	0.251	0.108	0.117
Log Likelihood	-16252			
AIC	32581			
AIC (Sample adjusted)	2.135			
Observations	1,272			

Notes: Standard errors are reported in parentheses. ***p < 0.01, **p < 0.05, *p < 0.1.

Table 1.6*Descriptive Statistics by Latent Classes*

Variable	Latent Classes			
	Locavores	COO	Information-sensitive	Price-sensitive
Female (%)	48.0	59.6	58.7	53.7
Age (%)				
18 – 24 years old	7.4	6.3	10.5	4.1
25 – 34 years old	25.5	16.3	25.6	8.2
35 – 44 years old	34.5	18.4	21.8	16.3
45 – 54 years old	11.3	11.9	13.5	12.2
55 – 64 years old	10.7	16.9	12.8	15.0
65+ years old	10.6	30.3	15.8	44.2
Marital Status (%)				
Divorced	8.0	10.0	7.5	10.9
Married	57.9	58.1	50.4	53.7
Separated	1.6	1.9	1.5	1.4
Single, Never Married	28.9	24.7	38.4	22.5
Widowed	3.6	5.3	2.3	11.6
Educational level (%)				
Less than High School	1.9	0.9	3.8	1.4
High School/GED	19.1	22.8	19.6	27.9
Some College	21.1	19.1	21.8	21.1
2-Year College Deg. (Assoc.)	8.2	8.1	12.0	15.7
4-Year College Deg. (BA, BS)	24.4	27.8	24.8	20.4
Master's Degree	21.3	16.9	15.8	11.6
Professional Deg. (Ph.D., J.D., M.D., etc.)	4.0	4.4	2.3	2.0
Annual pre-tax HH income in \$ (%)				
Less than 20,000	13.1	10.3	15.0	11.6
20,000 – 39,999	18.8	19.7	21.1	25.9
40,000 – 59,999	15.3	17.8	23.3	17.7
60,000 – 79,999	13.1	15.3	16.5	17.0
80,000 – 99,999	6.1	10.6	4.5	10.2
100,000 – 119,999	8.6	6.9	2.3	11.6
120,000 – 139,999	7.0	6.6	3.0	4.8
140,000 – 159,999	8.8	4.1	2.3	0.7

Table 1.6 (cont'd)

160,000 or greater	9.2	8.8	12.0	3.4
Seafood consumption location (%)				
At home	65.8	75.3	58.7	66.7
Away from home	34.2	24.7	41.3	33.3
	Mean			
	(Std. Dev.)			
Level of seafood fraud concern	57.2	56.7	51.4	46.2
	(27.0)	(27.0)	(31.1)	(29.4)
Observations	672	320	133	147

Table 1.7*Marginal WTP estimates with 95% confidence intervals*

Variable	mWTP estimate (\$/ 8oz fillets of seafood)			
	[95% C.I.]			
	Locavores	COO	Information-sensitive	Price-sensitive
<i>GL</i>	6.38 [4.34, 8.41]	3.62 [2.35, 4.89]	2.37 [-1.71, 6.45]	0.01 [-1.32, 1.35]
<i>US</i>	5.71 [3.79, 7.62]	4.63 [3.25, 6.00]	2.86 [-1.27, 6.99]	-0.05 [-1.27, 1.16]
<i>FRESH</i>	4.92 [3.27, 6.57]	3.28 [2.09, 4.47]	-0.42 [-3.85, 3.00]	0.39 [-0.65, 1.43]
<i>WILD</i>	4.68 [2.91, 6.45]	9.76 [7.26, 12.27]	9.04 [0.53, 17.56]	0.10 [-1.18, 1.37]
<i>FARM</i>	3.13 [1.57, 4.70]	1.24 [0.05, 2.43]	2.22 [0.45, 4.00]	-0.35 [-1.82, 1.13]

Notes: Asymptotic standard errors for the 95% confidence intervals calculated using the Delta Method with 10,000 draws.

Table 1.8*Latent Class Model Parameter Estimates with interactions*

4 Latent Classes, Fixed Parameters				
	Locavores	COO	Information-sensitive	Price-sensitive
Variable	Class 1	Class 2	Class 3	Class 4
<i>ASC</i>				
Salmon	2.746*** (0.127)	-0.103 (0.176)	4.097*** (0.546)	0.537 (0.807)
Trout	2.754*** (0.127)	-0.949*** (0.183)	-0.226 (0.609)	0.050 (0.829)
Whitefish	2.743*** (0.126)	-0.387** (0.181)	-0.208 (0.589)	0.903 (0.799)
<i>PRICE</i>	-0.053*** (0.007)	-0.136*** (0.015)	-0.115** (0.046)	-0.425*** (0.077)
<i>GL</i>	0.328*** (0.050)	0.528*** (0.087)	0.251 (0.308)	0.132 (0.318)
<i>GL × Treat</i>	0.013 (0.069)	-0.069 (0.112)	0.047 (0.397)	-0.489 (0.499)
<i>US</i>	0.326*** (0.050)	0.688*** (0.088)	0.959*** (0.321)	0.173 (0.287)
<i>US × Treat</i>	-0.056 (0.069)	-0.113 (0.112)	-1.076*** (0.399)	-0.649 (0.503)
<i>FRESH</i>	0.258*** (0.028)	0.454*** (0.058)	-0.095 (0.191)	0.105 (0.227)
<i>WILD</i>	0.245*** (0.037)	1.319*** (0.078)	1.158*** (0.281)	0.043 (0.274)
<i>FARM</i>	0.166*** (0.036)	0.149* (0.082)	0.951*** (0.268)	-0.049 (0.351)
Class prob.	0.526	0.249	0.108	0.117
Log Likelihood	-16246			
AIC	32585			
AIC (Sample adjusted)	2.135			
Observations	1,272			

Notes: Standard errors are reported in parentheses. ***p < 0.01, **p < 0.05, *p < 0.1.

Table 1.9*Marginal WTP estimates with interaction terms and 95% confidence intervals*

Variable	mWTP estimate (\$/ 8oz fillets of seafood)			
	[95% C.I.]			
	Locavores	COO	Information-sensitive	Price-sensitive
<i>GL</i>	6.25 [3.86, 8.63]	3.89 [2.38, 5.41]	2.17 [-3.22, 7.56]	0.31 [-1.15, 1.77]
<i>GL × Treat</i>	0.24 [-2.33, 2.81]	-0.51 [-2.13, 1.11]	0.41 [-6.33, 7.15]	-1.15 [-3.49, 1.19]
<i>US</i>	6.20 [3.83, 8.58]	5.07 [3.43, 6.71]	8.31 [0.14, 16.47]	0.41 [-0.91, 1.73]
<i>US × Treat</i>	-1.07 [-3.67, 1.54]	-0.84 [-2.43, 0.75]	-9.32 [-19.24, 0.60]	-1.53 [-3.92, 0.86]
<i>FRESH</i>	4.91 [3.27, 6.55]	3.35 [2.16, 4.54]	-0.82 [-4.12, 2.47]	0.25 [-0.79, 1.28]
<i>WILD</i>	4.67 [2.91, 6.42]	9.73 [7.28, 12.17]	10.04 [1.22, 18.85]	0.10 [-1.16, 1.36]
<i>FARM</i>	3.15 [1.60, 4.71]	1.10 [-0.09, 2.30]	8.24 [0.87, 15.61]	-0.12 [-1.74, 1.50]

Notes: Asymptotic standard errors for the 95% confidence intervals calculated using the Delta Method with 10,000 draws.

Table 1.10*Random Parameters Logit Model Estimates*

Variable	Treatment	Control
<i>PRICE</i>	-0.37*** (0.04)	-0.37*** (0.03)
<i>ASC</i>		
Salmon	2.80*** (0.11)	3.06*** (0.12)
Trout	2.21*** (0.10)	2.48*** (0.12)
Whitefish	2.29*** (0.11)	2.62** (0.12)
Estimates in WTP space		
<i>GL</i>		
Mean	1.26*** (0.20)	1.58*** (0.20)
S.D.	1.00*** (0.35)	1.16*** (0.20)
<i>US</i>		
Mean	1.50*** (0.21)	1.96*** (0.20)
S.D.	0.03 (0.42)	0.84*** (0.27)
<i>FRESH</i>		
Mean	2.30*** (0.19)	2.04*** (0.21)
S.D.	3.73*** (0.36)	3.64*** (0.24)
<i>WILD</i>		
Mean	2.06*** (0.27)	3.66*** (0.25)
S.D.	5.83*** (0.28)	5.45*** (0.29)
<i>FARM</i>		
Mean	0.68*** (0.25)	-0.46* (0.27)
S.D.	1.57*** (0.33)	3.93*** (0.27)
<i>Std. Dev. of error component</i>	8.13*** (0.44)	9.51*** (0.54)
Log Likelihood	-8,513	-8,373
AIC (Sample adjusted)	2.24	2.20
Number of parameters	16	16
Observations	7,632	7,632

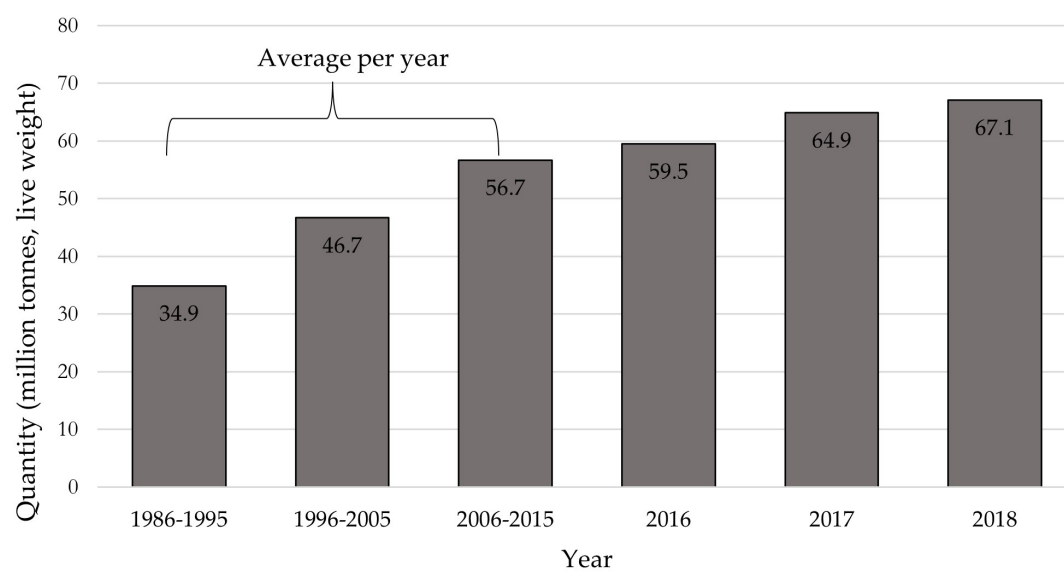
Notes: Standard errors are reported in parentheses. ***p < 0.01, **p < 0.05, *p < 0.1.

Table 1.11*Estimated Market Shares by Treatment*

Species	Unconditional		Conditional	
	Treatment	Control	Treatment	Control
<i>Salmon</i>	39%	34%	42%	37%
<i>Trout</i>	38%	28%	38%	31%
<i>Whitefish</i>	17%	29%	20%	32%
<i>None</i>	6%	9%		

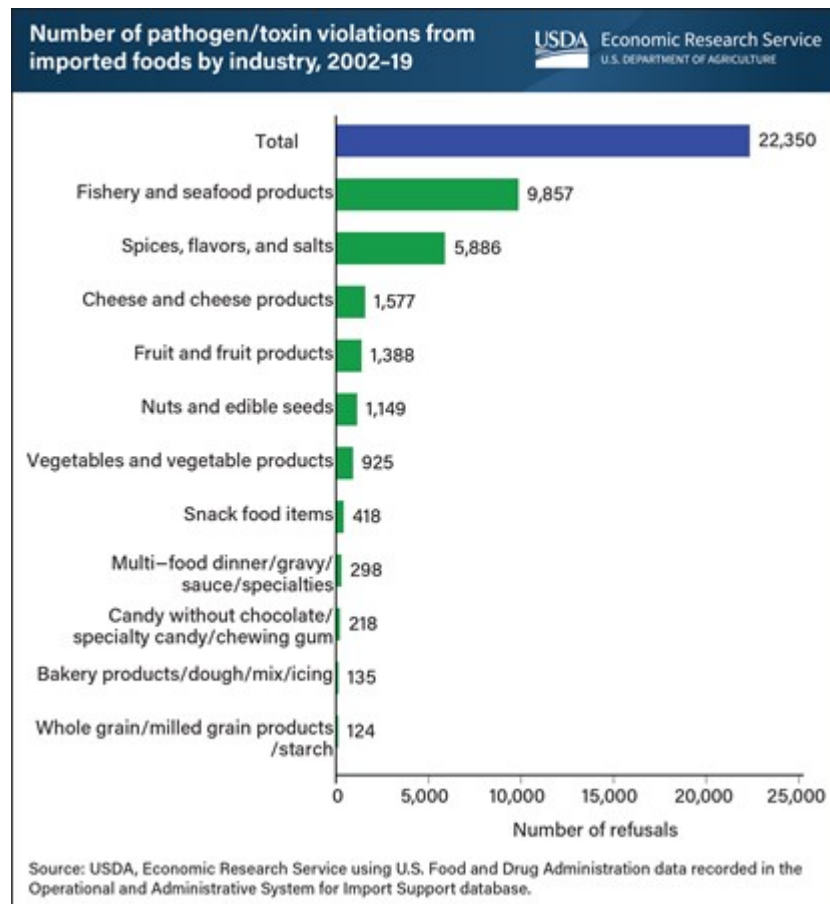
Notes: All prices are fixed at \$10.99 per 8oz fillets with mean shares approximated over 5,000 draws.

Figure 1.1: *Global seafood export volume, 1986 – 2018*



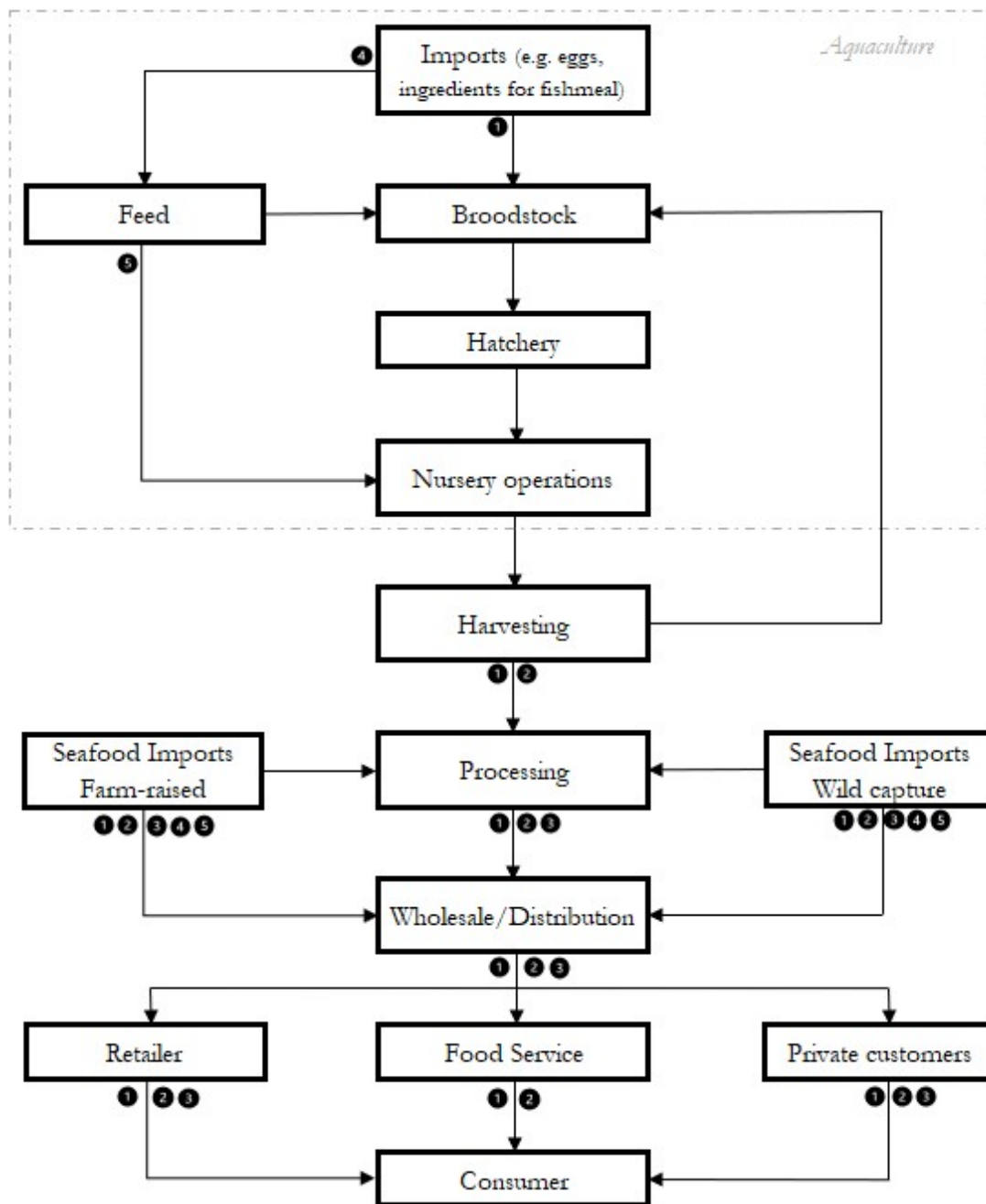
Source: FAO.

Figure 1.2: Bar graph of total number of pathogen/toxin violations from imported foods by industry, 2002 - 2019



Source: USDA - Economic Research Service.

Figure 1.3: Seafood supply chain with fish fraud vulnerability assessment

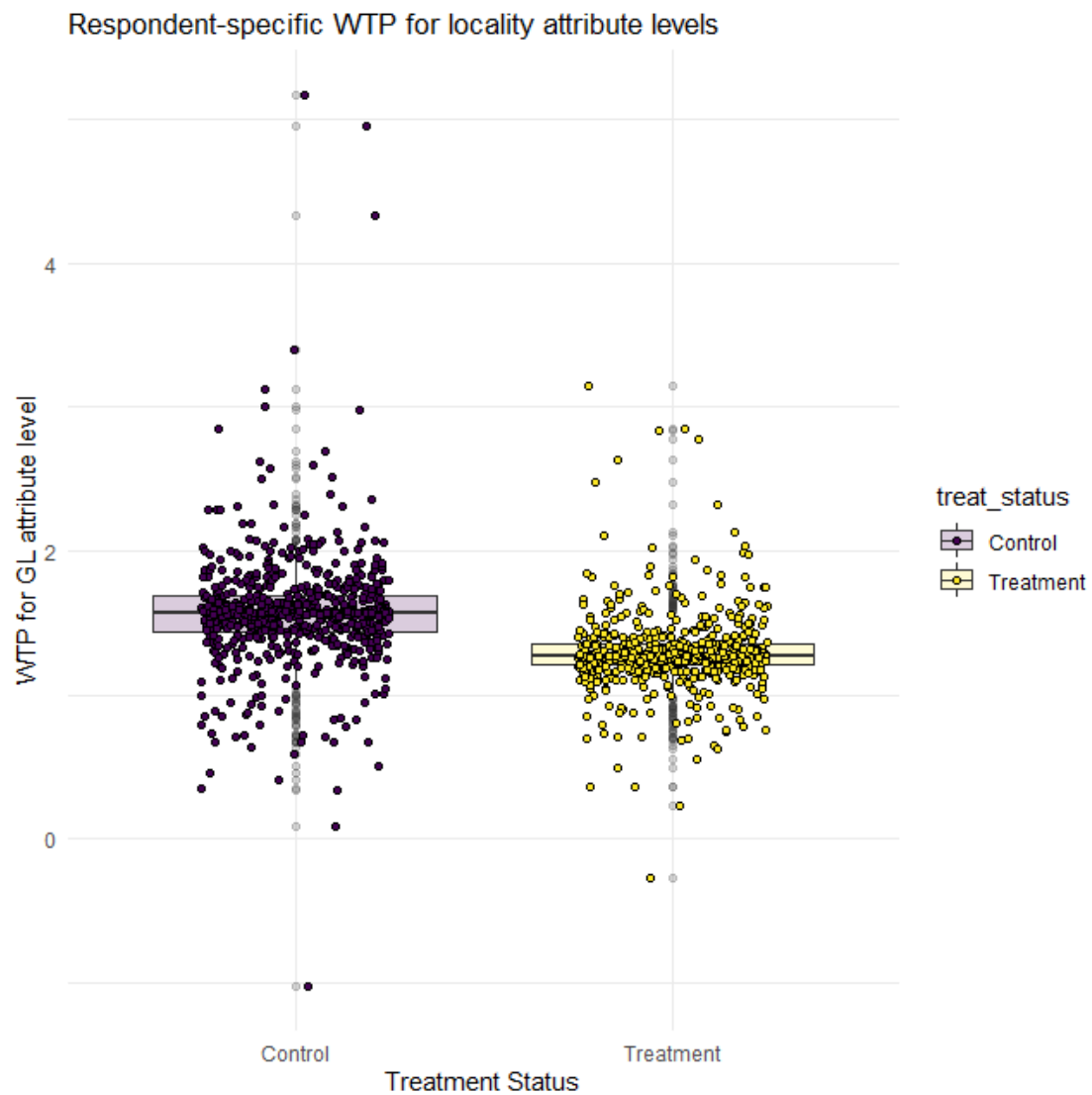


Notes: 1 = species substitution, 2 = mislabeling, 3 = short-weighting, 4 = adulteration, 5 = indiscriminate antibiotic use.

Figure 1.4: *Choice experiment question sample*

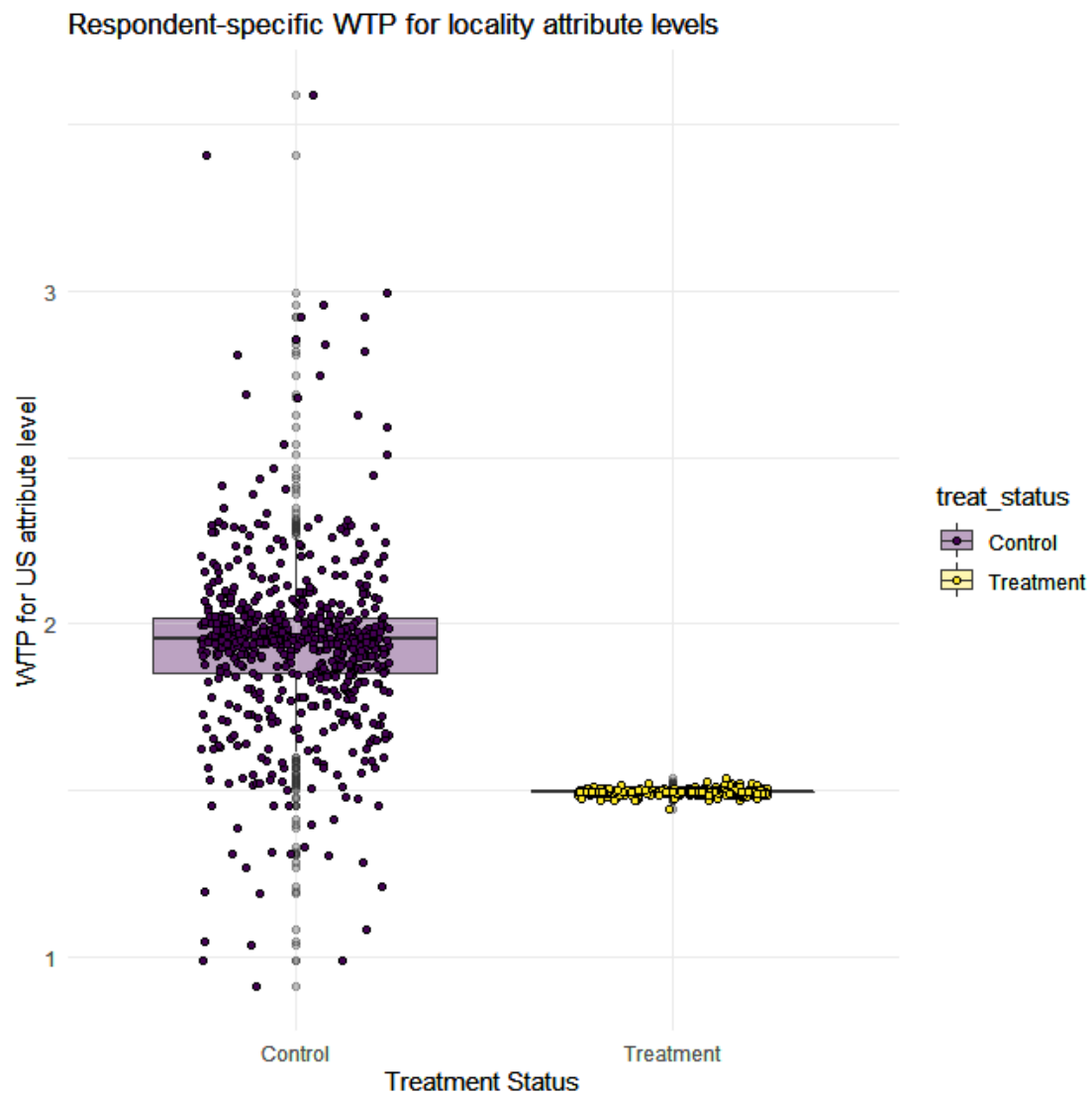
Finfish attribute	Option A	Option B	Option C	Option D
Species	Salmon	Trout	Whitefish	
Price (\$/ 8oz fillets)	\$11.99	\$9.99	\$13.99	I would purchase
Place of origin	United States	Imported	United States	none of these.
Processing form	Frozen	Frozen	Fresh	
Production method	No claim	No claim	Farmed	
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Figure 1.5: *Boxplot of respondent-specific WTP for GL attribute level*



Notes: The black lines in the middle of the colored rectangles denote the medians; the colored boxes show the interquartile range (IQR), and the whiskers are $1.5 \times \text{IQR}$. WTP estimates outside the IQR are shown in grey circles.

Figure 1.6: *Boxplot of respondent-specific WTP for US attribute level*



Notes: The black lines in the middle of the colored rectangles denote the medians; the colored boxes show the interquartile range, and the whiskers are $1.5 \times \text{IQR}$. WTP estimates outside the IQR are shown in grey circles.

APPENDIX B: DEFINITIONS AND EXCERPT

Definitions

Origin refers to where fish was farmed or caught:

Great Lakes Region refers to the region spanning the following states: Illinois, Indiana, Michigan, Minnesota, New York, Ohio, Pennsylvania, and Wisconsin.

The United States refers to any other state within the United States outside the Great Lakes Region.

Imported refers to any country outside the borders of the United States.

Processing form refers to the form in which fish was bought by the final consumer or restaurant:

Fresh means fish has never been frozen since harvest or catch.

Frozen means fish has undergone frozen storage since harvest.

Production method refers to the method of fish production:

Wild-caught means fish was captured in their natural habitat.

Farmed means fish was raised by a fish farmer in a controlled setting (i.e., aquaculture).

Unlabeled means no claims about the fish production method made.

Excerpt on Fish Fraud from a News Article

“There’s something, well, fishy going on with certain favorite fish dishes, according to a new study from the conservation group Oceana.

DNA tests showed that about 21% of the fish [that] researchers sampled was not what it was called on the label or menu.

With so many species and with 80% of the fish Americans eat coming from international sources, labeling is complicated.”

NB: This excerpt was derived from a published news article on the Cable News Network (CNN) website on March 7th, 2019.

URL to the full CNN article attached:

<https://www.cnn.com/2019/03/07/health/fish-mislabeled-investigation-oceana/index.html>

CHAPTER 2: DOES RURAL NON-FARM EMPLOYMENT RELIEVE (OR EXACERBATE) THE AGRICULTURAL DIVERSIFICATION-FARM EFFICIENCY TRADEOFF: THE CASE OF AQUACULTURE IN BANGLADESH

2.1 Introduction

This paper examines fish efficiency and agricultural diversification, and how each and their relation are conditioned by rural non-farm employment (RNFE). This is justified by a gap in the literature, which we show by reviewing the march of the crop diversification, RNFE, and efficiency literatures.

Adam Smith theorized that there are gains from specialization. Conversely, there could be efficiency losses associated with diversification such as in agriculture. This reduction in crop production efficiency, both technical and allocative, from crop diversification may happen for the following reasons. We define technical efficiency as a farm's ability to obtain maximal output from a given input bundle. Allocative efficiency occurs when a household allots resources in a manner that maximizes farm profits given input and output prices.

First, efficiency of the production of crop i could be undermined because of competition for labor and other inputs from adding crop j . For example, in Bangladesh, rice efficiency is undermined by adding jute, which competes with rice for water, land, and labor (Rahman, et al., 2017).

Second, multi-cropping could exert pressure on household labor and management time. In Papua New Guinea, Coelli & Fleming (2004) argue that overlapping labor and management needs among multiple cash crops create diseconomies of diversification.

By contrast, some studies find a positive or null effect of crop diversification on a target crop's efficiency. This may happen for several reasons. First, due to physical complementarities between crops. For example, nitrogen-fixing crops can enhance grain efficiency as found in the farming systems literature in the 1970s/1980s. In Bangladesh, Emran et al. (2022) find that sequentially cropped systems (i.e., rice with mungbean, lathyrus, or groundnut) improve rice productivity.

Second, crop diversification could stimulate cash and knowledge spillovers, which can help the target crop. For example, income from cash cropping can purchase capital and labor (Von Braun & Kennedy, 1994). Crop diversification could also spur productivity spillovers via cross-crop knowledge and skill transfer (Von Braun & Kennedy, 1994).

Third, seasonality can permit serial specialization within diversification, so crops do not compete. See Schreinemachers et al. (2016) for the case of off-season vegetable production in Bangladesh, with low impacts on on-season rice productivity.

Fourth, off-farm development of businesses like cash crop-based input dealers, selling fungible inputs like fertilizer and services like logistics could benefit food crops (Kennedy & Cogill, 1987).

Diversification into crop production among fish farmers is increasingly gaining in importance as a way around malfunctioning food crop markets and a reliable means of rural income diversification. Relatedly, synergies within diversified agricultural systems make a compelling case for integrated aquaculture-agriculture (IAA) technology adoption. Recent studies suggest that IAA adoption is a viable approach to sustainable agricultural intensification with considerable potential for improving agricultural productivity and food security in Bangladesh (Islam, 2021).

However, a factor that can affect each of the above (i.e., crop diversification and efficiency) and their relation is RNFE. The literature shows that RNFE is important to rural household livelihoods (Lanjouw & Shariff, 2004; Deichmann, et al., 2009). In Bangladesh, we show for fish farmers that non-farm activities account for less than half of rural household income, on average. Similarly, Deichmann et al. (2009) report that the non-farm income share of rural household income exceeds 50% in Bangladesh.

RNFE conditions crop diversification and a target crop's technical and allocative efficiency. Indeed, there are two vectors of effect.

First, an emerging literature shows that RNFE affects crop *i*'s efficiency. On one hand, RNFE provides cash for crop input purchase. For example, Begum et al. (2013) show in Bangladesh a positive correlation between RNFE and shrimp efficiency. The authors explain by positing that RNFE buys inputs. In China, Rozelle et al. (1999) find that remittances improve access to physical capital, raising maize yields. Chavas et al. (2005) also find that RNFE improves allocative efficiency in The Gambia, indicative of capital market imperfections in the study area.

Second, RNFE affects agricultural diversification. Others find that RNFE fuels diversification like into fish production. For example, in Myanmar, Faxon (2020) reports that migrant-sending households built fishponds using remittances, spurring diversification from paddy into aquaculture. However, RNFE can also compete for resources and time with agricultural diversification, such as into fruit trees (Huang, et al., 2009).

The literature has left additional gaps, which we attempt to fill. First, how RNFE conditions the crop diversification-farm efficiency tradeoff has not been studied for either type of efficiency. Second, although Begum et al. (2013) study the relation between RNFE and technical efficiency in aquaculture, RNFE impact on allocative efficiency was not addressed. Hence, how RNFE

affects optimal input (such as labor) choices given input prices in fish systems remains unclear. Besides, shrimp farming is less representative of Bangladeshi aquaculture. It accounts for only 4% of aquaculture production in the country. Hence, inferring Bangladeshi aquaculture productivity from shrimp technical efficiency estimates could mislead as shrimp farming uses significantly less inputs. Third, while Faxon (2020) shows that RNFE affects agricultural diversification, the evidence she presents are qualitative. Hence, an empirical analysis using a larger sample can provide additional validation to these qualitative results.

We attempt to fill the aforementioned gaps in the literature using panel data on fish farming households in Southern Bangladesh. We address the follow questions: (1) does non-farm income diversification resolve or exacerbate the crop diversification-efficiency tradeoff? We hypothesize that RNFE relieves the crop diversification-farm efficiency tradeoff. Cash from non-farm activities can buy labor to offset competition for household labor across multiple agricultural activities. On the other hand, RNFE may exacerbate this tradeoff if it diverts family labor away from the farm altogether, further constraining household labor; (2) does RNFE affect fish efficiency (both technical and allocative)? We postulate that RNFE increases fish efficiency by providing cash to buy skilled labor and other inputs. By contrast, RNFE could draw labor away from the farm, especially if off-farm work is year-round; (3) does RNFE affect crop diversification for fish systems? We conjecture that RNFE helps buy labor to invest in multiple lucrative agricultural ventures. However, RNFE may reduce crop diversification if off-farm work competes with agricultural production for labor.

The rest of the paper is organized as follows. Section 2.2 introduces our data and reports descriptive statistics on key variables. Section 2.3 outlines our empirical strategy. Section 2.4 presents our results, and we offer concluding remarks in section 2.5.

2.2 Data and Descriptives

This study uses data on fish farming households in the seven most important fish producing districts in Southern Bangladesh (see Figure 2.1 for a map of the sampled districts). Aquaculture farms in the seven surveyed districts cover 275,970 ha, accounting for 41% of the national area of aquaculture farms and 24% of national aquaculture production. The seven selected districts account for 80% of aquaculture production in Southern Bangladesh (DoF, 2022).

Aquaculture in this zone is a mix of fish and shrimp, grown in traditional extensive and improved semi-intensive systems. Agriculture is dominated by rainfed monsoon and irrigated dry season rice, vegetable crops, and some off-season oilseeds and pulses. Rice production is oriented predominantly to subsistence, while vegetable cultivation and aquaculture tend to be market oriented. Aquaculture and agriculture may be distributed across separate plots on a given farm or integrated within a single plot. (Jahan et al., 2015; Ali et al., 2022). The average number of ponds operated per farm was 1.5 in 2020.

The survey was a panel, conducted in 2 years, 2014, and 2020. Sampling procedure was as follows:

In 2014, all upazilas with negligible fish production (per the 2008 agricultural census) were excluded from the initial sample frame, and the sample was drawn from among the remaining upazila by probability proportional to size, leaving 13 *upazila* (sub-districts) from a total of 56.

In each selected upazila, all *mouza* (the smallest administrative unit listed in the Bangladesh agricultural census), underwent a second stage of trimming to eliminate those with fewer than 20 aquaculture farms. Two to three of the remaining mouza were selected randomly in each of the 13 upazila for inclusion in the farm survey. Prior to the survey, a census of fish farmers

was conducted in all selected mouza. In each selected mouza, 20 farms were selected randomly from this list for interview.

In 2020, we conducted a resurvey of households from the 2014 survey. Prior to the survey, we conducted a census of all fish farming households in villages included in the 2014 survey and attempted to identify all households previously interviewed. 579 out the 721 households interviewed in 2014 showed up in the 2020 panel, implying an attrition rate of 20 percent. While the attrition rate is high, we find no significant differences between attrited and non-attrited subsamples across most variables, although more remote farms, and households with a higher non-farm income share were slightly more likely to attrite.

Table 2.1 describes the variables relevant to our analyses. We consider the following production input variables: the quantity of feed and non-feed inputs applied, total person-days of hired as well as family labor used, and the quantity of fish seed stocked.¹⁴ We also collected data on household demographics such as the household size, dependency ratio, household head's gender, and educational status. Other key variables include fish price per kilogram (kg), daily wage rate at the household-level, an off-farm participation indicator variable, crop farm income share, fish farm income share, non-farm income share, a crop diversification indicator variable, fish plots distance to the nearest road and from the household's residence. We also construct a Simpson diversification index (SDI), ranging between 0 and 1 to capture crop diversification at the intensive margin:

$$SDI_i = 1 - \sum_j \left(\frac{A_j}{\sum_j A_j} \right)^2 \quad (1)$$

where A_j denotes the acreage allocated to crop j .

¹⁴ The pond size (water area) was originally reported in decimals, where 100 decimals = 0.404648 hectares. We use this variable to scale the aforementioned input variables to produce per hectare measures.

Table 2.2 presents descriptive results pooled across both panel waves and disaggregated by crop diversification status. The following results are noteworthy. A substantial share of the sampled farm households (69%) are into crop production. The table also shows that crop diversified households used more family labor and almost twice as much hired labor on fish farms as households not growing any crops. The former also applied more feed. By contrast, undiversified farms utilized more nonfeed inputs and stocked more fish seed, on average. However, these differences are not statistically significant.

Turning to the household demographics, our results indicate that household size does not vary much by crop diversification status. The average household has approximately 4.6 members, with a dependency ratio of 0.6. Male household heads dominate our sample (96% male headship), and household heads in crop diversified households tend to have higher levels of education.

We also observe that a considerable share (60%) of the sampled households participated in off-farm activities. Off-farm participation rate was higher (63%) among crop diversified households and significantly so compared to those producing no crops (54%). Turning to the non-farm income share, the average estimate stands at 43%, ranging between 41% (crop diversified) and 46% (not diversified). This difference is significant at the 5 percent level and is consistent with similar estimates reported for Sub-Saharan Africa by Reardon (1997).

That said, aquaculture production remains the dominant source of earnings, accounting for 49% of total rural household income. This is true across both crop diversified (47%) and undiversified households (54%). Table 2.2 also shows that the difference in fish farm income share by crop diversification status is statistically significant. By contrast, crop production accounts for a relatively lower share (less than 10%) of total household income.

2.3 Empirical Strategy

2.3.1 Technical Efficiency Estimation

We derive technical efficiency estimates by specifying a stochastic frontier production function (SFPF) model in panel data setting as follows:

$$Y_{it} = f(x_{it}; \beta) \exp(v_{it} - u_{it}) \quad (2)$$

where Y_{it} denotes the fish output level; the deterministic portion of the model, $f(x_{it}; \beta)$ represents household i 's fish production frontier with input vector x_{it} at time t ; β denotes model parameters to be estimated; v_{it} is a symmetric disturbance term that captures statistical noise; and u_{it} is the inefficiency term. Following Stevenson (1980), we impose a normal-truncated-normal distributional assumption on the v — u error pair. In particular, the symmetric random error term, v_{it} is assumed to be *iid* normally distributed with zero mean and standard deviation, σ_v (that is, $v_{it} \sim N(0, \sigma_v^2)$). We also maintain that the one-sided inefficiency term is *iid* whose distribution derives from the truncation of $N(\mu, \sigma_u^2)$ at zero. Hence, the SFPF model reduces to the standard neoclassical production function if $u_{it} = 0$ (Kumbhakar, et al., 2020). Following Kumbhakar et al. (1991), we deploy the more efficient single-step approach which simultaneously accounts for the determinants of inefficiency while estimating the production frontier model parameters.

To address potential bias in the estimated production function parameters due to unobserved household heterogeneity, we apply the Mundlak-Chamberlain approach by including the means of the time-varying input variables as controls (Wooldridge, 2019). The Mundlak test of the null hypothesis that the unobserved heterogeneity can be ignored is rejected at the 1 percent level (p —*value* = 0.0005). Hence, we decide against the true random effects model, which

maintains the strong independence assumption between the production input covariates and the unobserved heterogeneity term.

We use a transcendental logarithmic (translog) specification for the SFPF estimation. Not only is the translog functional form more flexible for estimating production technology, but also supported by results from a likelihood ratio test for our sample. Hence, we estimate the following translog SFPF:

$$\ln Y_{it} = \beta_0 + \sum_{k=1}^5 \beta_k \ln x_{ikt} + \frac{1}{2} \sum_{j=1}^5 \sum_{k=1}^5 \beta_{jk} \ln x_{ikt} \ln x_{ijt} + \sum_{k=2}^5 \lambda_k D_{ikt} + \eta_i + \zeta_t + v_{it} - u_{it} \quad (3)$$

where i indexes the fish farm; Y_{it} represents the quantity of fish harvested in kilograms per hectare of pond water area; x_{ikt} denotes the quantity of input variable k used per hectare (see Table 2.1 for details on how the various input variables are defined), η_i is the unobserved heterogeneity term and ζ_t are time dummies; D_{ikt} are dummy variables which take the value 1 for zero-valued observations of the input variables, and 0 otherwise, which are included to facilitate the logarithmic transformation of the explanatory variables with zeros.¹⁵ The technical inefficiency term takes the following form:

$$\begin{aligned} u_{it} = & \alpha_0 + \alpha_1 CDI_{it} + \alpha_2 RNFE_{it} + \alpha_3 DEP_RATIO_{it} + \alpha_4 GENDER_i \\ & + \alpha_5 EDUC_i + \alpha_6 DIST_ROAD_{it} + \alpha_7 DIST_HH_{it} + \alpha_8 ANY_PRAWN_i \\ & + \alpha_9 SHRIMP_NOPRAWN_i \\ & + QUINT_COMM_{it} \beta + \zeta_d + \zeta_t + \varepsilon_{it} \end{aligned} \quad (4)$$

¹⁵ See Battese (1997) and Henderson (2015) for details. There is no such indicator variable for the fish seed variable since it has no zero values.

where the distribution of the disturbance term, ε_{it} derives from a normal distribution with zero mean and variance, σ_u^2 truncated at $-z_{it}\alpha$ with z_{it} denoting a vector of the aforementioned determinants of technical inefficiency (see below for specifics on how each of these variables are defined).

- CDI_{it} denotes the crop diversification indicator variable, which takes the value 1 if the farm produces any crops.
- $RNFE_{it}$ is the share of total household income from non-farm activities, ranging from 0 to 1.
- DEP_RATIO_{it} is the dependency ratio.
- $GENDER_i$ is a dummy variable, which takes the value of 1 if the household head is male.
- $EDUC_i$ denotes the household head's level of education in years.
- $DIST_ROAD_{it}$ is the mean distance of fishponds to the nearest road (in kilometers)
- $DIST_HH_{it}$ is the average distance of fishponds from the household's residence (in kilometers)
- ANY_PRAWN_i is a dummy variable for whether the household produces any prawn, where the "only fish" subgroup is the comparison category.
- $SHRIMP_NOPRAWN_i$ is a dummy variable indicating that the household cultivates shrimp but no prawn, with the "only fish" category as the base.
- $QUINT_COMM_{it}$ is a vector of dummies representing the commercialization quintiles, whose coefficients are interpreted relative to the lowest quintile (the omitted category). The 5th quintile is the most commercialized.

A negative and significant estimated coefficient signifies a decline in technical inefficiency in response to a marginal increase in the variable of interest. We include farmer district dummies, ζ_d to control for agroecological and infrastructural differences across districts and a time trend, ζ_t to capture secular effects common to all fish farmers in the inefficiency part of the model. While controlling for unobserved heterogeneity at a much granular level (i.e., the farm household level) is preferred, doing so worsened the model fit as indicated by the log-likelihood ratio statistic. Moreover, the estimated coefficients in the inefficiency part of the model appear inflated, which is symptomatic of underlying model convergence issues. Hence, we instead include farmer district dummies to partly control for this unobserved heterogeneity.

Following Coelli et al. (2005), we define technical efficiency as the ratio of observed fish output, Y_{it} to the maximum potential output that a fully efficient fish farmer can produce using the same set of inputs:

$$TE_{it} = \frac{Y_{it}}{Y_{it}^*} = \frac{f(x_{it}; \beta) e^{v_{it} - u_{it}}}{f(x_{it}; \beta) e^{v_{it}}} = \exp(-u_{it}) \quad (5)$$

Varying between 0 and 1, this unobservable technical efficiency score is approximated using the sample analog of the following conditional expectation term, $E[e^{-u_{it}} | \epsilon_{it}]$, where $\epsilon_{it} = v_{it} - u_{it}$ (Kumbhakar, et al., 2020).

2.3.2 Allocative Efficiency Estimation

Next, we obtain our allocative inefficiency measure as the natural logarithm of the ratio between observed wage rate and the estimated marginal revenue product of labor ($MRPL$), $\ln\left(\frac{w_{it}}{MRPL_{it}}\right)$ (Barrett, et al., 2008; Henderson, 2014). The $MRPL_{it}$ is derived from the stochastic frontier production function estimates as the product of per unit fish price and the marginal physical product of family labor employed on fish farms, which is estimated as follows:

$$\frac{\delta Y_{it}}{\delta L_{it}^f} = \frac{\delta \ln Y_{it}}{\delta \ln L_{it}^f} \times \frac{Y_{it}}{L_{it}^f} \quad (6)$$

An allocative inefficiency (*AI*) score of zero is what theory predicts in the absence of labor market frictions and other input or credit market failures. By contrast, $AI_{it} < 0$ ($AI_{it} > 0$) signifies an undersupply (oversupply) of on-farm labor than is optimal.

In line with Barrett et al. (2008) and Henderson (2014), for non-wage employed households, we impute their *AI* scores from the predicted values resulting from the regression of *AI* on select household- and farm-level covariates. In doing so, we account for sample selection using Heckman (1979)'s two-step approach. Among the covariates included in the selection equation are fish price, the dependency ratio, household size, the household head's gender, educational attainment as well as a time dummy. For consistent inference, we adjust the standard errors accordingly via bootstrapping with 1,000 replications in the second stage.

In what follows, we present our main finding from the two-step Heckman procedure. This method has the advantage of removing bias due to nonrandom selection as a result of missing wage data for households not involved in wage labor (Heckman, 1979). First, we estimate a probit model on the entire sample using households' participation in off-farm work for wages as the dependent variable. We then include the resulting inverse Mills ratio (*IMR*) in an augmented regression of the *AI* scores on select household- and farmer-level characteristics for the selected subsample. Next, we impute *AI* scores for the households without wage labor from the resulting predicted values. Results from the selectivity and *AI* regressions are summarized in Table 2.8 in the APPENDIX. Most importantly, the estimated coefficient on the *IMR* is statistically significant at the 1 percent level, implying that sample selection bias could be an issue if not accounted for.

Further, for ease of comparison with the TE results, we transform the AI variable into an allocative efficiency, AE equivalent such that $AE \rightarrow 1$ as $AI \rightarrow 0$ and $AE \rightarrow 0$ when $|AI| \rightarrow \infty$. Following Henderson (2014), we perform this transformation using the kernel of the normal density function:

$$AE_{it} = \exp\left(-\frac{AI_{it}^2}{2\sigma_{AI}^2}\right) \quad (7)$$

centered around the “ideal” AI mean of zero, where σ_{AI}^2 denotes the variance of AI around a mean of zero.

We specify the following regression equation to quantify the associations between the different diversification strategies and allocative efficiency:

$$AE_{it} = \lambda_0 + \lambda_1 CDI_{it} + \lambda_2 RNFE_{it} + \mathbf{X}_{it}\boldsymbol{\gamma} + \eta_i + \epsilon_{it} \quad (8)$$

using a fixed effects panel estimator, where the righthand-side variables are as previously defined; \mathbf{X}_{it} is a vector of the time-varying subset of controls included in the TE regression¹⁶; η_i is a household fixed effects term to control for unobserved household-specific heterogeneity and ϵ_{it} is the zero-valued disturbance term. We adjust for within-household correlation in the errors over time by clustering the standard errors at the household-level (Abadie, et al., 2023).

2.4 Regression Results

2.4.1 Determinants of Technical Efficiency

Table 2.3 reports the SFPF regression estimation results for both the Cobb-Douglas (C-D) and Translog specifications in columns (1) and (2), respectively. In both specifications, we control for unobserved household heterogeneity using the Mundlak-Chamberlain method and cluster

¹⁶ We also include the households’ value of aquaculture-related assets and its quadratic term. However, we do not include the production system dummies as standalone variables in this regression. This is because they do not vary over time; hence, are absorbed into the fixed effects term.

standard errors at the household-level. Column (1) shows that all the input variables are positive and statistically significant at the 1 percent level, indicating that fish yield is monotonically increasing in all inputs. However, as shown in Table 2.3, results from a Wald likelihood ratio test point towards a rejection of the joint test null that the coefficients on all quadratic and interaction terms are equal to zero (test statistic: $94.76 > 24.99 = \chi^2_{15}$). Hence, we base the rest of our analyses on the translog specification results.

A test of the null hypothesis that production technology is characterized by constant returns to scale (CRS) follows the derivation of output elasticities for the respective production input variables evaluated at the pooled sample means. We find no evidence to reject the null that the sum of the individual input elasticities equals 1, indicating constant returns to scale. Of the 5 inputs considered, the output elasticity of familial labor is the highest (0.634), followed by fish seed (0.497). These estimated elasticities are significant at the 5 percent level or better. By contrast, the output elasticities for hired labor, feed, and non-feed inputs are not statistically distinguishable from zero.

Before turning to our main regression results, we first present graphical evidence on the distribution of the *TE* and *AE* scores. The average *TE* score is 66%, which is lower than the median as indicated by the left-skewed distribution of the *TE* scores histogram plot (see Figure 2.2). By contrast, we observe a pile-up at zero for the *AE* scores, indicating that allocative inefficiency is rather the norm. In particular, the average *AE* score hovers around 34%.

Results from the inefficiency part of the SFPF estimation are shown in Table 2.4. Column (1) presents regression results from the most parsimonious specification. In column (2), we include a squared RNFE term to capture potential nonlinearities between RNFE and technical inefficiency.

Column (3) includes an interaction between the crop diversification dummy and RNFE variable to test for the extent to which RNFE conditions the crop diversification-technical efficiency relation.

As indicated earlier, a negative (positive) estimated coefficient as reported in Table 2.4 indicates that technical inefficiency is declining (increasing) in the said household- or farmer-level characteristic. We first discuss the results in column (1). The following results are of note. The coefficient of the crop diversification indicator is not statistically different from zero. This result suggests that, if anything, crop diversification does not compete with fish production on average, all else held constant. Similarly, we do not find a significant association between non-farm income diversification and technical efficiency. This result corroborates findings in other studies that also report an insignificant relationship between off-farm employment and technical efficiency (Chavas et al., 2005; Yang et al., 2016).

Table 2.4 also reports weak evidence of a nonlinear effect of RNFE on technical efficiency. The coefficient of the squared term is positive and statistically significant at the 10 percent level. This is consistent with evidence from Bangladesh, where Mondal et al. (2020) also report similar nonlinear effects of RNFE among crop producers.

In other results, we find that producing *any prawn* has a negative and statistically significant effect on technical efficiency. This effect is significant at the 10 percent level or better. By contrast, the coefficient on the *shrimp, but no prawn* variable is not statistically different from zero.

The results also show that technical inefficiency is negatively and significantly decreasing with aquaculture commercialization. Note that the parameter estimates on the commercialization quintile dummies are interpreted relative to the bottom quintile and are each significant at the 10 percent level or better, depending on the specification. We also find that technical inefficiency is

increasing with fishpond remoteness, suggesting that farms are more productive the closer they are to fish input and output markets.

To address our first research question, in column (3), we interact the crop diversification dummy with the non-farm income share variable. The coefficient on the interaction term is not significant, indicating that the association between non-farm income on technical efficiency does not differ by crop diversification status, on average. This result is slightly at odds with the results reported in Table 2.9 in the APPENDIX, where we instead use a continuous crop diversification variable, the Simpson diversification index.¹⁷ Table 2.9 shows that at higher levels of the non-farm income share, diversifying into crop production results in technical inefficiencies. This effect is significant at the 10 percent level. Perhaps, the Simpson index captures a richer variation in crop diversification than the crop diversification indicator variable; hence, the differences we observe.

2.4.2 Relationship between diversification and allocative efficiency

Next, we turn to estimating the association between livelihood diversification and allocative efficiency. Regression results are reported in Table 2.5. We begin with the results in column (1). We find a positive association between crop diversification and allocative efficiency on average, *ceteris paribus*. The estimated coefficient of the crop diversification dummy is significant at the 10 percent level. Since most fish farms (62%) in our sample tend to overuse family labor, crop diversification could be absorbing some of this surplus household labor, thereby improving allocative efficiency. By contrast, we do not detect any meaningful impact of RNFE on allocative efficiency. This may suggest that fish farms may not be buying labor with cash from RNFE to optimize household labor allocation between farm and off-farm activities.

As a robustness check and due to the pile up at zero for the allocative efficiency scores, we also present results from a Tobit regression for our most parsimonious specification. The estimated coefficient and marginal effects are reported in Table 2.11 in the APPENDIX. As can be seen, the results are very similar in magnitude to our FE-OLS regression estimates.

Other interesting results also emerge. We find that the dependency ratio is positively and significantly associated with allocative efficiency, all else held constant. Allocative efficiency was also found to be decreasing in the average distance of fish plots from the household's residence. This effect is significant at the 10 percent level.

In column (3), we include an interaction between the crop diversification dummy and the RNFE variable. We find a negative relationship between crop diversification and allocative efficiency at higher levels of non-farm income share. This effect is statistically significant at the 1 percent level. We suspect that undertaking both diversification strategies severely constrains household labor. Further, the toll on managerial ability from juggling both on-farm and off-farm diversification could undermine how efficiently farms use household labor. These results are qualitatively similar to alternative specifications, where we instead use the continuous crop diversification variable (see Table 2.10 in the APPENDIX).

2.4.3 Diversification and Fish Input Demand

Table 2.6 presents the effect of both crop diversification and RNFE on fish input demand. We estimate effects on demand for household labor, hired labor, fish seed, feed, and nonfeed inputs per hectare of pond water area as defined in Table 2.1. However, for the nonlabor input variables, we instead use expenditure data to better capture input quality.

Results indicate that crop diversification increases demand for family labor on fish farms, all else held constant. This can be due to the relatively higher demand for labor to manage sub-

systems on the farm in the face of labor market frictions. Moreover, crops are typically cultivated in close proximity to fishponds, and sometimes integrated within a single plot. Hence, more family labor use in crop production may imply more intense household labor use on fish farms. Also, this may reflect cash investment from crop sales into aquaculture, demanding more family labor for pond repair, guarding of fishponds, harvesting, among others. This explanation, however, is less plausible given that a relatively small share of household income comes from crop sales (less than 10%).

Similarly, we find that a higher non-farm income share increases household labor demand on average, *ceteris paribus*. This is a surprising result as we would expect higher non-farm incomes to increase hired labor use, substituting for family labor on fish farms. That said, this result aligns with Takahashi and Otsuka (2009)'s findings, which demonstrated increased utilization of family labor on rice farms when rice income was the primary source of earnings. By contrast, the coefficient on the interaction between RNFE and the crop diversification indicator is not statistically significant. The constraining effect of adopting both diversification strategies could explain this result as there may be no residual family labor pool to draw from.

Table 2.6 also shows that neither diversification strategies increased hired labor demand, which partly justifies the household labor supply effects we observe. This could also suggest large transaction costs in hiring in labor. In other results, we do not find any significant association between income diversification and fish seed, feed, and nonfeed input expenditure.

2.4.4 Association between RNFE and Crop Diversification

Table 2.7 presents regression results on the relationship between RNFE and crop diversification. As hypothesized, RNFE can compete with crop diversification for household labor time. Further, RNFE could substitute for crop diversification as a source of cash for fish input

purchase. On the other hand, RNFE can stimulate agricultural diversification such as into cash cropping from traditional food crops. Our results show that there is a negative and significant association between RNFE and crop diversification, suggesting a substitution between off-farm activities and diversification into crops.

2.5 Conclusions

Using panel survey data on fish farming households in southern Bangladesh, we examine fish efficiency and agricultural diversification, and how RNFE conditions each and their relation. We apply a fixed effects estimator to control for unobserved household-specific heterogeneity and derive technical efficiency estimates by fitting a stochastic frontier production function (SFPF). Following Barrett et al. (2008) and Henderson (2014), we also obtain a proxy for allocative efficiency using the imputation methods depicted therein. We derive the following key findings:

First, we do not find any significant relationship between crop diversification and technical efficiency. Similarly, there is no significant association between non-farm income diversification and technical efficiency. This result is supported by the finding that RNFE does not significantly increase non-family labor input purchase, on average.

Second, we do not find a significant interaction effect between crop diversification and RNFE on technical efficiency. By contrast, we find that higher levels of the non-farm income share result in a negative and significant (at the 10% level) association between crop diversification and technical efficiency, when we define crop diversification using the Simpson diversification index. We hypothesize that this may be due to the constraining effect of both diversification strategies on family labor. Indeed, we show that the effect of undertaking both diversification strategies on household labor demand is nil. Coupled with complexities of multitasking, adopting both strategies may place enormous strain on family labor, thereby negatively impacting aquaculture productivity.

Third, we find a positive and significant crop diversification effect on allocative efficiency. This may reflect a reallocation of surplus family labor from aquaculture to crop production, where family labor is overused. On the other hand, RNFE does not exert any meaningful impact on allocative efficiency, all else held constant.

Fourth, our results also indicate that for crop diversified households, higher levels of the non-farm income share reduces allocative efficiency. Following the same reasoning, allocative inefficiencies could result due to the strain on family labor resources. Moreover, we show that hired labor demand does not increase significantly with income diversification to relieve the pressure on familial labor.

Fifth, we find evidence of a substitution between crop diversification and RNFE. Perhaps, this points to the competing demand for household labor across both activities. On the other hand, this may reflect the view that cash from RNFE for input purchase may substitute for liquidity from cash cropping.

This study opens up other avenues for future research. To start with, research into the nature of crop diversification activities undertaken by the households could offer richer insights into which crops complement or compete with aquaculture.

Further, it will be useful to disentangle the household labor effects from the cash impact of undertaking both on-farm and off-farm diversification strategies. The general impression is that the cash effect is minimal given the limited impacts on non-household labor input expenditure.

Moreover, an investigation of the impact of interspecies diversification on fish efficiency is also of research interest. Such diversification can be seen as a practical way of diversifying risks associated with species-specific disease outbreaks and price volatility.

It should be noted that the results we present here are associational and should be interpreted with caution. Hence, additional work on the causal interpretation of the three-way relationship among crop diversification, RNFE, and farm efficiency is a valuable research pursuit and left to future research.

BIBLIOGRAPHY

- Abadie, A., Athey, S., Imbens, G. W. & Wooldridge, J. M., 2023. When should you adjust standard errors for clustering?. *The Quarterly Journal of Economics*, 138(1), pp. 1-35.
- Ali, H. et al., 2022. Economic performance characterization of intensive shrimp (*Penaeus monodon*) farming systems in Bangladesh. *Aquaculture, Fish and Fisheries*, 2(1), pp. 57-70.
- Barrett, C. B., Sherlund, S. M. & Adesina, A. A., 2008. Shadow wages, allocative inefficiency, and labor supply in smallholder agriculture. *Agricultural Economics*, 38(1), pp. 21-34.
- Begum, E. A., Hossain, M. I. & Papanagiotou, E., 2013. Technical Efficiency of Shrimp Farming in Bangladesh: An Application of the Stochastic Production Frontier Approach. *Journal of the World Aquaculture Society*, 44(5), pp. 641-654.
- Bezemer, D., Balcombe, K., Davis, J. & Fraser, I., 2005. Livelihoods and farm efficiency in rural Georgia. *Applied Economics*, Volume 37, p. 1737–1745.
- Chavas, J.-P., Petrie, R. & Roth, M., 2005. Farm Household Production Efficiency: Evidence from The Gambia. *American Journal of Agricultural Economics*, 87(1), p. 160–179.
- Coelli, T. & Fleming, E., 2004. Diversification economies and specialisation efficiencies in a mixed food and coffee smallholder farming system in Papua New Guinea. *Agricultural Economics*, 31(2-3), pp. 229-239.
- Deichmann, U., Shilpi, F. & Vakis, R., 2009. Urban Proximity, Agricultural Potential and Rural Non-farm Employment: Evidence from Bangladesh. *World Development*, 37(3), pp. 645-660.
- DoF, 2022. *Yearbook of Fisheries Statistics of Bangladesh, 2020-21*, Bangladesh: Fisheries Resources Survey System (FRSS), Department of Fisheries. Bangladesh: Ministry of Fisheries and Livestock.
- Emran, S.-A. et al., 2022. Impact of cropping system diversification on productivity and resource use efficiencies of smallholder farmers in south-central Bangladesh: a multi-criteria analysis. *Agronomy for Sustainable Development*, 42(4), p. 78.
- Faxon, H. O., 2020. *The Peasant and Her Smartphone: Agrarian Change and Land Politics in Myanmar*, s.l.: s.n.
- Heckman, J. J., 1979. Sample selection bias as a specification error. *Econometrica: Journal of the Econometric Society*, pp. 153-161.
- Henderson, H., 2014. Considering Technical and Allocative Efficiency in the Inverse Farm Size–Productivity Relationship. *Journal of Agricultural Economics*, 66(2), p. 442–469.
- Huang, J., Wu, Y. & Rozelle, S., 2009. Moving off the farm and intensifying agricultural production in Shandong: a case study of rural labor market linkages in China. *Agricultural Economics*, Volume 40, pp. 203-218.

- Islam, A. H. M. S., 2021. Dynamics and Determinants of Participation in Integrated Aquaculture– Agriculture Value Chain: Evidence from a Panel Data Analysis of Indigenous Smallholders in Bangladesh. *The Journal of Development Studies*, 57(11), pp. 1871-1892.
- Jahan, K. M. et al., 2015. *Aquaculture technologies in Bangladesh: An assessment of technical and economic performance and producer behavior*, Penang, Malaysia: WorldFish.
- Kennedy, E. T. & Cogill, B., 1987. *Income and nutritional effects of the commercialization of agriculture in southwestern Kenya*. s.l.:Intl Food Policy Res Inst.
- Kilic, T., Carletto, C., Miluka, J. & Savastano, S., 2009. Rural non-farm income and its impact on agriculture: evidence from Albania. *Agricultural Economics*, Volume 40, pp. 139-160.
- Kumbhakar, S. C., Parmeter, C. F. & Zelenyuk, V., 2020. Stochastic frontier analysis: Foundations and advances I. *Handbook of production economics*, pp. 1-40.
- Lanjouw, P. & Shariff, A., 2004. Rural non-farm employment in India: Access, incomes, and poverty impact. *Economic and Political Weekly*, pp. 4429-4446.
- Mondal, R. K., Selvanathan, E. A. & Selvanathan, S., 2020. Nexus between rural non-farm income and agricultural production in Bangladesh. *Applied Economics*, 53(10), pp. 1184-1199.
- Pfeiffer, L., López-Feldman, A. & Taylor, J. E., 2009. Is off-farm income reforming the farm? Evidence from Mexico. *Agricultural Economics*, Volume 40, pp. 125-138.
- Rahman, S., Kazal, M. M. H., Begum, I. A. & Alam, M. J., 2017. Exploring the future potential of jute in Bangladesh. *Agriculture*, 7(12), p. 96.
- Schreinemachers, P. et al., 2016. Farmer training in off-season vegetables: Effects on income and pesticide use in Bangladesh. *Food Policy*, Volume 61, pp. 132-140.
- Stevenson, R. E., 1980. Likelihood functions for generalized stochastic frontier estimation. *Journal of Econometrics*, 13(1), pp. 57-66.
- Takahashi, K. & Otsuka, K., 2009. The increasing importance of non-farm income and the changing use of labor and capital in rice farming: the case of Central Luzon, 1979–2003. *Agricultural Economics*, 40, pp. 231-242.
- Von Braun, J. & Kennedy, E. T., 1994. *Agricultural commercialization, economic development, and nutrition*, s.l.: Johns Hopkins University Press.
- Wooldridge, J. M., 2019. Correlated random effects models with unbalanced panels. *Journal of Econometrics*, 211(1), p. 137–150.
- Yang, J. et al., 2016. Migration, local off-farm employment, and agricultural production efficiency: evidence from China. *Journal of Productivity Analysis*, 45, pp. 247-259.

APPENDIX A: TABLES AND FIGURES

Table 2.1

<i>Description of key variables</i>	
Variable	Description
Yield (kg/ha)	Total quantity of harvested fish (including shrimp and prawns) from the whole farm in last production cycle per hectare of pond water area.
Familial labor (days/ha)	Total person-days of household labor employed per hectare spent on activities such as pond and dyke repair, stocking, feeding, fertilizer application, weeding, guarding, harvesting, and marketing.
Hired labor (days/ha)	Total person-days of hired labor per hectare.
Fish seed stocked (kg/ha)	Total quantity of fish seed stocked in the last production cycle per hectare.
Feed inputs (kg/ha)	Total quantity of both commercial pelleted and own farm-made feed applied over last cropping cycle per hectare.
Non-feed inputs (kg/ha)	Total quantity of urea, NPK ¹⁸ , TSP, DAP, cow dung, lime, salt, and other organic manure applied per hectare. ¹⁹
Fish price (BDT/kg)	Price per kilogram of fish sold in Bangladeshi Taka
Wage rate (BDT/day)	Daily wage rate in Bangladeshi Taka
Fish farm income share	Share of total household income from fish sales
Crop farm income share	Share of total household income from crop sales
Non-farm income share	Share of total household income from non-farm sources (i.e., wage and self-employment as well as remittances)
Crop diversification dummy (0/1)	An indicator variable which takes the value 1 if household produces any crop, mostly rice and vegetables
Simpson Diversification Index ²⁰	Degree of crop diversification (0 indicates no crop diversification)
Marketed surplus share (%)	Share of harvested fish that is sold
Distance to nearest road (km)	Mean distance of fish plots from the nearest road in kilometers

¹⁸ NPK denotes nitrogen, phosphorus, and potassium; TSP denotes triple super phosphate; DAP denotes diammonium phosphate.

¹⁹ We also explored the sensitivity of our main results to alternative feed and nonfeed input use variables. In particular, we use feed and nonfeed input values in place of the quantity measures. While these alternative input variables may better reflect input quality, the model fit for our stochastic frontier production function (SFPF) regressions are slightly worse. Further, the input parameter estimates are attenuated. This result could be partially attributed to a worsening of the measurement error problem especially for the fixed effects estimator since the input values incorporate self-reported price information. Hence, we prefer the quantity-based feed and nonfeed input measures.

Distance to household (km)	Mean distance of fish plots from the household's residence in kilometers
----------------------------	--

Table 2.1 (cont'd)

Household size	Total number of adults and children living in the household
Dependency ratio	Number of household members aged < 15 plus 65+ divided by those aged 15 - 64 years old
Gender (Male = 1)	Household head's gender
Education (years)	Household head's years of schooling
Off-farm (0/1)	Dummy for off-farm work participation

Table 2.2*Summary statistics on key variables by crop diversification status*

	Total	Crop diversification status	
		Diversified	Not diversified
<i>Production variables</i> ²¹			
Yield (kg/ha)	2189.46	2274.81	1997.65
Familial labor (days/ha)*	805.33	882.97	630.85
Hired labor (days/ha)	54.78	64.17	33.68
Fish seed (kg/ha)	1626.11	1594.43	1697.32
Feed inputs (kg/ha)*	2206.24	2484.90	1580.03
Nonfeed inputs (kg/ha)	1269.09	1071.59	1712.90
<i>Price variables</i>			
Fish price (BDT/kg)*	234.32	217.88	272.07
Wage rate (BDT/day)	311.13	314.88	301.34
<i>Household characteristics</i>			
Gender (Male = 1)	0.96	0.97	0.95
Education (years)	5.4	5.6	5.0
Household size	4.6	4.7	4.5
Dependency ratio	0.60	0.61	0.56
Off-farm (0/1)*	0.60	0.63	0.54
<i>Income Diversification variables</i>			
Fish farm income share (%)*	49	47	54
Crop farm income share (%)*	8	12	
Non-farm income share (%)*	43	41	46
Crop diversification (0/1)*	0.69	1.00	0.00
Simpson index*	0.17	0.24	0.00
<i>Plot-level variables</i>			
Distance to nearest road (km)	0.66	0.61	0.77
Distance to household (km)	0.55	0.57	0.51
Observations	1156	800	356

Notes: Values reported are means unless otherwise stated. Monetary values are expressed in 2014 constant prices to account for inflation. Simpson index ranges from 0 to 1, where 0 represents no diversification. * indicates significant difference in means by crop diversification status at the 5% level of better.

Table 2.3*Stochastic Frontier Production Function (SFPP) estimation results*

Dependent variable: $\ln(\text{Yield})$		
Variable	(1) Cobb-Douglas	(2) Translog
$\ln(\text{Familial labor})$	0.180*** (0.025)	-0.164 (0.128)
$\ln(\text{Hired labor})$	0.131*** (0.026)	0.147 (0.144)
$\ln(\text{Fish seed})$	0.118*** (0.031)	-0.273** (0.132)
$\ln(\text{Feed})$	0.182*** (0.024)	0.270* (0.157)
$\ln(\text{Non} - \text{feed})$	0.132*** (0.020)	0.248* (0.134)
$\frac{1}{2} \times \ln(\text{Familial labor})^2$		0.061*** (0.015)
$\frac{1}{2} \times \ln(\text{Hired labor})^2$		0.008 (0.028)
$\frac{1}{2} \times \ln(\text{Fish seed})^2$		0.033 (0.023)
$\frac{1}{2} \times \ln(\text{Feed})^2$		-0.011 (0.020)
$\frac{1}{2} \times \ln(\text{Non} - \text{feed})^2$		-0.039** (0.015)
$\ln(\text{Familial}) \times \ln(\text{Hired})$		-0.001 (0.009)
$\ln(\text{Familial}) \times \ln(\text{Fish seed})$		0.028* (0.016)
$\ln(\text{Familial}) \times \ln(\text{Feed})$		-0.013* (0.007)
$\ln(\text{Familial}) \times \ln(\text{Non} - \text{feed})$		-0.016* (0.009)
$\ln(\text{Hired}) \times \ln(\text{Fish seed})$		-0.017 (0.014)
$\ln(\text{Hired}) \times \ln(\text{Feed})$		0.002 (0.006)
$\ln(\text{Hired}) \times \ln(\text{Non} - \text{feed})$		0.009 (0.009)

Table 2.3 (cont'd)

$\ln(\text{Fish seed}) \times \ln(\text{Feed})$		-0.001 (0.010)
$\ln(\text{Fish seed}) \times \ln(\text{Non} - \text{feed})$		0.023 (0.014)
$\ln(\text{Feed}) \times \ln(\text{Non} - \text{feed})$		0.001 (0.006)
<i>yr2020</i>	-0.245** (0.111)	-0.042 (0.119)
<i>Constant</i>	2.888*** (0.239)	4.681*** (0.923)
Output elasticity wrt		
<i>Familial labor</i>		0.634*** (0.116)
<i>Hired labor</i>		0.155 (0.123)
<i>Fish seed</i>		0.497** (0.202)
<i>Feed</i>		0.023 (0.174)
<i>Non – feed</i>		-0.188 (0.120)
H_o : Constant returns to scale		
<i>LR-statistic (p-value)</i>		0.12 (0.731)
Joint test of significance $\beta_{jk} = 0$		
<i>LR-statistic</i>		94.76***
Household FE	✓	✓
Log pseudolikelihood	-1305.26	-1257.88
Observations	1,158	1,158

Notes: Standard errors are clustered at the household-level and are reported in parentheses. LR-statistic denotes the Likelihood ratio statistic. We control for household fixed effects using the Mundlak-Chamberlain approach. $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. wrt denotes “with respect to.”

Table 2.4

<i>Determinants of Technical inefficiency from Translog SFPF</i>			
Variables	Dependent variable: Technical inefficiency		
	Coefficient estimates		
	(S.E.)		
	(1)	(2)	(3)
<i>CDI</i>	0.225 (0.684)	0.542 (0.987)	-1.469 (1.138)
<i>RNFE</i>	1.543 (1.068)	-10.255 (6.479)	-0.792 (1.590)
<i>RNFE</i> ²		13.143* (7.636)	
<i>CDI</i> × <i>RNFE</i>			3.337 (2.091)
<i>ANY_PRAWN</i>	1.815** (0.812)	2.466* (1.384)	1.672* (0.935)
<i>SHRIMP_NO_PRAWN</i>	1.712 (1.217)	2.077 (1.713)	1.696 (1.354)
<i>QUINT_COMM:</i>			
<i>Q2</i>	-3.012** (1.243)	-2.549 (1.858)	-3.150* (1.668)
<i>Q3</i>	-4.874*** (1.714)	-5.382* (3.186)	-5.072** (2.390)
<i>Q4</i>	-4.684*** (1.643)	-4.928* (2.860)	-4.896** (2.248)
<i>Q5</i>	-4.301*** (1.484)	-4.700* (2.615)	-4.381** (1.985)
<i>DEP_RATIO</i>	-0.910 (0.662)	-1.323 (1.110)	-0.969 (0.783)
<i>GENDER</i>	-2.529 (1.907)	-2.637 (2.538)	-2.909 (2.379)
<i>EDUC</i>	-0.222** (0.104)	-0.286 (0.187)	-0.235* (0.138)
<i>DIST_ROAD</i>	1.149*** (0.362)	1.509* (0.809)	1.142** (0.517)
<i>DIST_HH</i>	-0.021 (0.145)	0.006 (0.208)	-0.022 (0.158)
<i>yr2020</i>	5.277 (4.727)	9.169 (8.465)	5.669 (5.683)

Table 2.4 (cont'd)

<i>Constant</i>	-7.140 (4.864)	-11.516 (9.194)	-6.291 (5.842)
σ_u	2.162*** (0.278)	2.393*** (0.569)	2.224*** (0.442)
σ_v	0.520*** (0.024)	0.527*** (0.029)	0.519*** (0.026)
$\lambda = \frac{\sigma_u}{\sigma_v}$	4.159*** (0.274)	4.539*** (0.557)	4.286*** (0.433)
District FE	✓	✓	✓
Observations	1,158	1,158	1,158

Notes: Negative coefficients indicate a decline in technical inefficiency with a marginal increase in the variable of interest. Standard errors are reported in parenthesis and are clustered at the household level * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2.5

Allocative efficiency regression using FE-OLS

Variables	Dependent variable: Allocative efficiency		
	Coefficient estimates		
	(S.E.)		
	(1)	(2)	(3)
<i>CDI</i>	0.059* (0.033)	0.058* (0.033)	0.145*** (0.043)
<i>RNFE</i>	0.054 (0.049)	0.164 (0.158)	0.198*** (0.074)
<i>RNFE</i> ²		-0.128 (0.176)	
<i>CDI</i> × <i>RNFE</i>			-0.217*** (0.083)
<i>QUINT_COMM</i> :			
<i>Q2</i>	-0.056 (0.047)	-0.060 (0.048)	-0.059 (0.046)
<i>Q3</i>	-0.054 (0.049)	-0.060 (0.050)	-0.052 (0.048)
<i>Q4</i>	-0.168*** (0.049)	-0.174*** (0.050)	-0.159*** (0.049)
<i>Q5</i>	-0.095* (0.050)	-0.100** (0.050)	-0.091* (0.049)
<i>DEP_RATIO</i>	0.091*** (0.031)	0.091*** (0.031)	0.089*** (0.031)
<i>EDUC</i>	0.006 (0.005)	0.006 (0.005)	0.006 (0.005)
<i>DIST_ROAD</i>	0.030** (0.014)	0.029** (0.014)	0.032** (0.014)
<i>DIST_HH</i>	-0.022* (0.013)	-0.021* (0.013)	-0.022* (0.012)
<i>ASSET_VALUE</i> /10 ³	1.24e-03*** (4.42e-04)	1.22e-03*** (4.48e-04)	1.22e-03*** (4.36e-04)
<i>ASSET_VALUE</i> ² /10 ⁶	-3.83e-07*** (1.44e-07)	-3.77e-07** (1.46e-07)	-3.79e-07** (1.42e-07)
<i>yr2020</i>	-0.333*** (0.023)	-0.337*** (0.023)	-0.324*** (0.023)
<i>Constant</i>	0.418*** (0.061)	0.417*** (0.061)	0.352*** (0.065)
Household FE	✓	✓	✓
Observations	1,109	1,109	1,109

Notes: Standard errors are reported in parenthesis and are clustered at the household level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2.6

Effect of diversification on fish input demand

Variables	Dependent variables				
	ln(familial labor)	ln(hired labor)	ln(fish seed in BDT)	ln(feed in BDT)	ln(nonfeed in BDT)
	Coefficient estimates (S.E.)				
<i>CDI</i>	0.470** (0.202)	1.236 (1.772)	0.459 (0.246)	-0.758 (0.568)	-0.078 (0.335)
<i>RNFE</i>	0.670** (0.323)	1.029 (2.661)	-0.751 (0.599)	-1.599* (0.863)	-0.692 (0.626)
<i>CDI × RNFE</i>	-0.422 (0.337)	-2.145 (3.060)	-0.205 (0.546)	1.189 (0.995)	0.277 (0.722)
<i>QUINT_COMM:</i>					
<i>Q2</i>	-0.190 (0.223)	1.136 (1.669)	0.511* (0.291)	0.773 (0.592)	1.349*** (0.413)
<i>Q3</i>	-0.210 (0.197)	4.185** (1.641)	0.725** (0.294)	0.508 (0.568)	1.350*** (0.415)
<i>Q4</i>	-0.359* (0.196)	3.782** (1.624)	0.405 (0.290)	0.764 (0.540)	1.517*** (0.388)
<i>Q5</i>	-0.594*** (0.206)	4.146** (1.719)	0.665** (0.280)	0.394 (0.574)	1.739*** (0.405)
<i>DEP_RATIO</i>	0.092 (0.103)	2.840*** (1.077)	-0.054 (0.226)	-0.206 (0.292)	-0.174 (0.182)
<i>EDUC</i>	-0.079*** (0.022)	-0.066 (0.168)	-0.011 (0.020)	-0.022 (0.059)	0.028 (0.040)
<i>DIST_ROAD</i>	0.019 (0.058)	0.293 (0.452)	-0.012 (0.056)	0.091 (0.151)	0.056 (0.089)
<i>DIST_HH</i>	-0.129 (0.084)	0.235 (0.388)	-0.039 (0.052)	-0.014 (0.149)	0.099 (0.086)
<i>ASSET_VALUE/10³</i>	-7.51e-04* (3.93e-04)	-6.73e-03*** (1.39e-03)	-6.50e-04** (3.04e-04)	1.16e-03 (8.35e-04)	-1.38e-04 (6.85e-04)
<i>yr2020</i>	2.205*** (0.101)	16.502*** (0.773)	-5.336*** (0.143)	7.481*** (0.274)	2.750*** (0.165)
<i>Constant</i>	4.565*** (0.260)	-0.637 (2.236)	16.449*** (0.377)	4.131*** (0.767)	5.341*** (0.529)
Household FE	✓	✓	✓	✓	✓
Observations	1,158	1,158	1,158	1,158	1,158

Notes: Standard errors are reported in parenthesis and are clustered at the household level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2.7*Effect of RNFE on crop diversification*

Dependent variable: Crop diversification (0/1)	
Variables	Coefficient estimates (S.E.)
<i>RNFE</i>	-0.137** (0.065)
<i>DEP_RATIO</i>	0.009 (0.034)
<i>EDUC</i>	0.006 (0.008)
<i>DIST_ROAD</i>	-0.001 (0.026)
<i>DIST_HH</i>	0.013 (0.017)
<i>ASSET_VALUE</i> /10 ³	5.35e-04 (7.65e-04)
<i>ASSET_VALUE</i> ² /10 ⁶	-1.14e-07 (2.50e-07)
<i>yr2020</i>	0.141*** (0.031)
<i>Constant</i>	0.640*** (0.062)
Household FE	✓
Observations	1,160

Notes: Standard errors are reported in parenthesis and are clustered at the household level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure 2.1: *Map of sampled Bangladeshi districts*

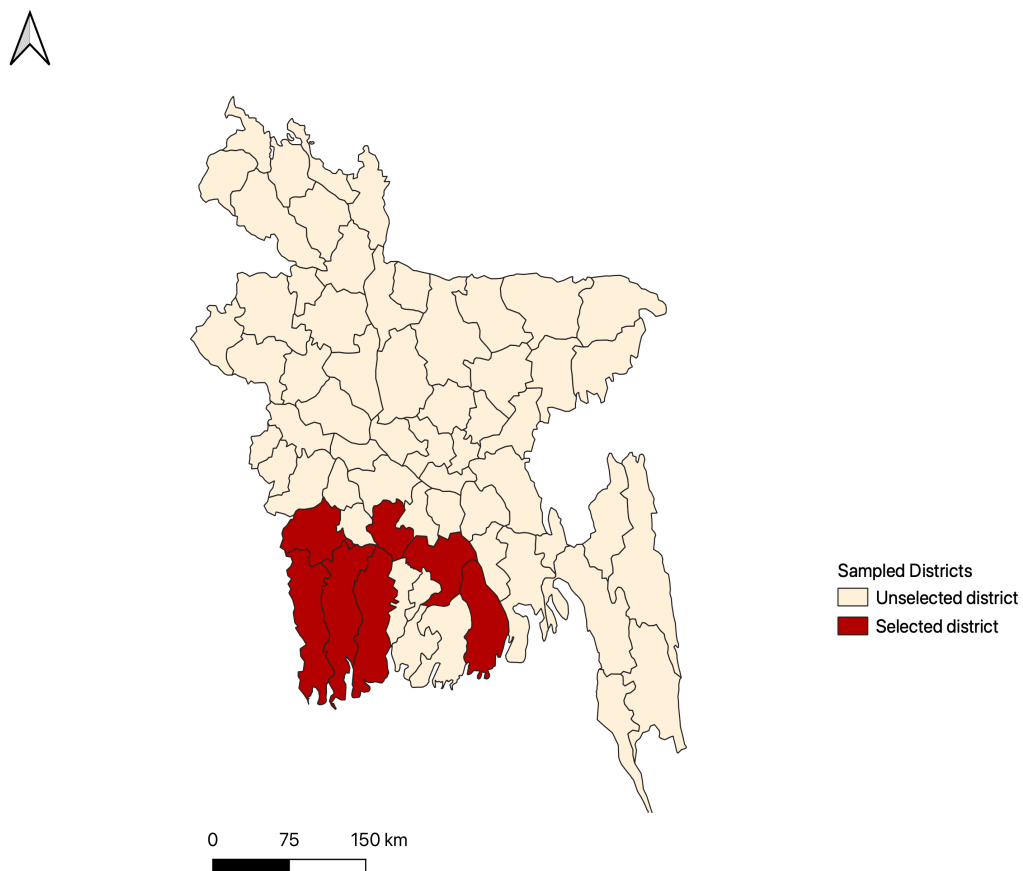
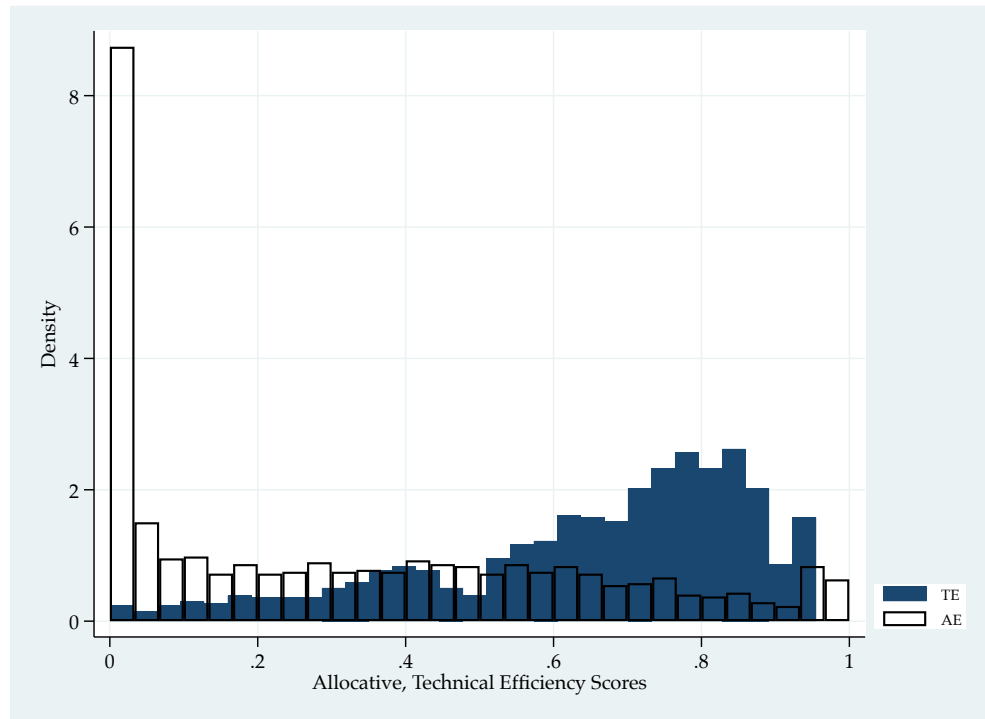


Figure 2.2: *Distribution of Technical and Allocative Efficiency estimates*



APPENDIX B: SUPPLEMENTARY TABLES

Table 2.8

<i>Two-step Heckman Correction</i>		
Dependent variable: Off-farm work participation (0/1)		
Variable	Coeff. estimates	S.E.
<i>Panel A: Probit selection equation</i>		
<i>FISH_PRICE</i>	0.0006**	0.0002
<i>HHSIZE</i>	0.029	0.027
<i>DEP_RATIO</i>	-0.184**	0.091
<i>EDUC</i>	-0.020*	0.011
<i>GENDER</i>	0.699**	0.278
<i>yr2020</i>	2.229***	0.108
<i>Constant</i>	-1.532***	0.304
Observations	1,138	
Dependent variable: Allocative inefficiency (AI)		
Variable	Coeff. estimates	S.E.
<i>Panel B: AI equation</i>		
<i>HHSIZE</i>	0.060***	0.015
<i>EDUC</i>	-0.021***	0.007
<i>GENDER</i>	0.255	0.157
<i>EXPER</i>	0.001	0.003
<i>DIST_ROAD</i>	0.039*	0.021
<i>DIST_HH</i>	-0.160***	0.025
<i>IMR</i>	1.952***	0.151
<i>yr2020</i>	4.675***	0.192
<i>Constant</i>	-3.696***	0.335
Observations	1,138	

Notes: IMR denotes the inverse mills ratio. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2.9

<i>Determinants of Technical inefficiency using alternative crop diversification variable</i>			
Variables	Dependent variable: Technical inefficiency		
	Coefficient estimates		
	(S.E.)		
	(1)	(2)	(3)
<i>Simpson index</i>	-1.099 (1.228)	-0.960 (1.313)	-5.471* (2.974)
<i>RNFE</i>	1.392 (1.075)	-8.984** (3.736)	0.120 (0.953)
<i>RNFE</i> ²		11.384*** (4.069)	
<i>Simpson index</i> × <i>RNFE</i>			8.599* (5.044)
<i>ANY_PRAWN</i>	1.758** (0.881)	2.165*** (0.781)	1.579* (0.857)
<i>SHRIMP_NO_PRAWN</i>	1.636 (1.251)	1.791 (1.264)	1.703 (1.257)
<i>QUINT_COMM:</i>			
<i>Q2</i>	-2.913** (1.439)	-2.263** (1.132)	-2.838* (1.548)
<i>Q3</i>	-4.751** (2.072)	-4.785*** (1.441)	-4.484** (2.241)
<i>Q4</i>	-4.606** (1.961)	-4.424*** (1.426)	-4.402** (2.092)
<i>Q5</i>	-4.226** (1.742)	-4.217*** (1.288)	-4.074** (1.882)
<i>DEP_RATIO</i>	-0.865 (0.703)	-1.142 (0.700)	-0.810 (0.701)
<i>GENDER</i>	-2.467 (1.995)	-2.314 (1.789)	-2.090 (2.005)
<i>EDUC</i>	-0.217* (0.117)	-0.253*** (0.095)	-0.209 (0.127)
<i>DIST_ROAD</i>	1.114*** (0.427)	1.317*** (0.274)	0.983** (0.448)
<i>DIST_HH</i>	-0.022 (0.141)	0.013 (0.169)	-0.034 (0.131)
<i>yr2020</i>	5.408 (4.834)	8.077 (5.142)	4.371 (4.455)

Table 2.9 (cont'd)

<i>Constant</i>	-6.675 (4.975)	-9.284* (4.840)	-4.975 (4.560)
σ_u	2.143*** (0.372)	2.295*** (0.153)	2.078*** (0.448)
σ_v	0.519*** (0.025)	0.527*** (0.023)	0.513*** (0.026)
$\lambda = \frac{\sigma_u}{\sigma_v}$	4.132*** (0.365)	4.359*** (0.157)	4.053*** (0.437)
District FE	✓	✓	✓
Observations	1,158	1,158	1,158

Notes: Negative coefficients indicate a decline in technical inefficiency with a marginal increase in the variable of interest. Standard errors are reported in parenthesis and are clustered at the household level * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2.10

<i>Allocative efficiency regression using alternative crop diversification variable</i>			
Variables	Dependent variable: Allocative efficiency		
	Coefficient estimates		
	(S.E.)		
	(1)	(2)	(3)
<i>Simpson index</i>	0.055 (0.054)	0.054 (0.054)	0.150* (0.083)
<i>RNFE</i>	0.048 (0.049)	0.169 (0.158)	0.084 (0.055)
<i>RNFE</i> ²		-0.140 (0.175)	
<i>Simpson index</i> × <i>RNFE</i>			-0.220 (0.146)
<i>QUINT_COMM</i> :			
<i>Q2</i>	-0.059 (0.047)	-0.063 (0.048)	-0.056 (0.047)
<i>Q3</i>	-0.052 (0.049)	-0.058 (0.051)	-0.049 (0.049)
<i>Q4</i>	-0.171*** (0.049)	-0.177*** (0.050)	-0.165*** (0.049)
<i>Q5</i>	-0.099** (0.050)	-0.104** (0.050)	-0.095* (0.050)
<i>DEP_RATIO</i>	0.090*** (0.031)	0.091*** (0.031)	0.090*** (0.031)
<i>EDUC</i>	0.006 (0.005)	0.006 (0.005)	0.007 (0.005)
<i>DIST_ROAD</i>	0.031** (0.014)	0.030** (0.014)	0.031** (0.014)
<i>DIST_HH</i>	-0.022* (0.012)	-0.021* (0.012)	-0.020 (0.012)
<i>ASSET_VALUE</i> /10 ³	1.25e-03*** (4.42e-04)	1.23e-03*** (4.47e-04)	1.24e-03*** (4.49e-04)
<i>ASSET_VALUE</i> ² /10 ⁶	-3.82e-07*** (1.44e-07)	-3.77e-07** (1.46e-07)	-3.76e-07** (1.47e-07)
<i>yr2020</i>	-0.333*** (0.025)	-0.338*** (0.025)	-0.335*** (0.025)
<i>Constant</i>	0.453*** (0.057)	0.450*** (0.057)	0.431*** (0.059)
Household FE	✓	✓	✓

Table 2.10 (cont'd)

Observations	1,109	1,109	1,109
<i>Notes:</i> Standard errors are reported in parenthesis and are clustered at the household level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.			

Table 2.11

<i>Tobit allocative efficiency regression</i>		
Dependent variable: Allocative efficiency		
Variable	Coefficient estimate (S.E)	Avg. marginal effect
<i>CDI</i>	0.056* (0.032)	0.047* (0.026)
<i>RNFE</i>	0.075 (0.052)	0.063 (0.044)
<i>QUINT_COMM:</i>		
<i>Q2</i>	-0.043 (0.028)	-0.037 (0.024)
<i>Q3</i>	-0.008 (0.028)	-0.007 (0.024)
<i>Q4</i>	-0.077*** (0.029)	-0.065*** (0.025)
<i>Q5</i>	-0.036 (0.028)	-0.031 (0.024)
<i>DEP_RATIO</i>	0.079*** (0.028)	0.067*** (0.024)
<i>EDUC</i>	0.004 (0.004)	0.003 (0.004)
<i>DIST_ROAD</i>	0.022 (0.014)	0.018 (0.012)
<i>DIST_HH</i>	-0.017 (0.017)	-0.014 (0.015)
<i>ASSET_VALUE</i> /10 ³	1.26e-03** (6.23e-04)	1.07e-03** (5.21e-04)
<i>ASSET_VALUE</i> ² /10 ⁶	-3.92e-07 (2.12e-06)	
<i>yr2020</i>	-0.346*** (0.022)	
<i>Constant</i>	0.450*** (0.038)	
Household FE	✓	
Observations	1,109	

Notes: Standard errors are reported in parenthesis and are obtained via bootstrapping with 500 replication. We control for household-specific heterogeneity using Mundlak-Chamberlain's Correlated Random Effects (CRE) approach by including the time averages of all the righthand-side variables in the regression equation. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

CHAPTER 3: PARENTAL EDUCATIONAL ATTAINMENT AND CHILD LABOR: EVIDENCE FROM MALAWI

3.1 Introduction

Does child labor respond inversely to parental education? If so, whose education matters more, and for which forms of child labor? Curbing child labor in all forms remains an elusive undertaking especially in low- and middle-income settings. For the first time in nearly two decades, the global effort against child labor has stalled (ILO, 2021). Recent International [Labor] Organization (ILO) statistics indicate that 160 million children aged 5 – 17 were involved in child labor—up from 152 million in 2016 (ILO, 2017). This trend, however, masks significant heterogeneity in child labor prevalence at the sub-regional level. At present, child labor mitigation appears more challenging across many sub-Saharan African, South Asian, and Latin American countries (see Figure 3.1). Child labor is especially pronounced in sub-Saharan Africa, where 1 in every 5 children aged 5 – 17 is a child laborer (ILO, 2017, 2021).

At the same time, one of the factors widely acknowledged to hold promise for child labor mitigation on the sub-continent is human capital acquisition. Studies on the cascading intergenerational effects of parental education on child labor and schooling outcomes are well-documented in the development economics literature (Patrinos and Psacharopoulos, 1995; Rosati and Tzannatos, 2000; Das and Mukherjee, 2007; Emerson and Souza, 2007; Cigno et al., 2001; Kurosaki et al., 2006). While there exists overwhelming evidence, suggesting a negative association between parental education and child labor participation, studies that aim to establish a causal link are rare. Potential confounders that jointly determine child labor decisions and parental education such as cultural inclinations, prevailing local economic activity levels, *inter alia*, could limit the extent to which previous research findings can inform policy.

To overcome this identification challenge, I draw on insights from the demography literature, wherein findings suggest that the direct influence of grandparents or lack thereof on grandchildren's socioeconomic outcomes hinges crucially on familial living arrangements. Consistent with this finding, Zeng and Xie (2014) note that non-co-resident grandparents' educational attainment has no bearing on grandchildren's schooling outcomes conditional on parental characteristics.

While some studies provide support for this result (Warren and Hauser, 1997; Erola and Moio, 2007), others find evidence to the contrary (Jæger, 2012; Chan and Boliver, 2013). Despite controlling for parents' education, income, and wealth, Chan and Boliver (2013) note that grandparents exert a significant direct effect on grandchildren's occupational classes in Britain. This result, however, does not account for multigenerational co-residence (Chan and Boliver, 2013). Hence, conditional on multigenerational co-residence and a range of parental and household-level characteristics, I use as a set of instruments grandparents' educational attainment to exploit plausibly exogenous variation in parents' schooling.²² Further, to explore the robustness of my results, I apply practical methods that relax the exclusion restriction assumption in an imperfect instruments framework (Conley et al., 2012). In particular, I derive bounds for the causal parameters of interest when the instruments are allowed to violate the exclusion restriction.

Using quasi-random access to education in Nyasaland (now Malawi) during the colonial period (mostly between 1859 and 1964), I present evidence on the effect of parental education on child labor outcomes while taking advantage of this opportune setting. First, the early stages of this period coincided with the peak of slave trade in Malawi, at the height of which men, women,

²² That is, I use maternal grandparents' education as instruments for mother's schooling, whereas the paternal grandparents' educational levels serve as instrumental variables for the father's schooling conditional on multigenerational co-residence as well as parental and household-specific characteristics.

and sometimes children were abducted in organized raids. Not long after the Christian missionaries had settled in Nyasaland did they realize that these slave raids were more than just distractions to their evangelism mandate. With mostly firearms and sometimes ransom payments, the missionaries gradually brought these slave raiding forays under control. As the freed ex-slave populations became the responsibility of the missionaries, the former were educated in missionary schools as part of their rehabilitation (Allen, 2008).

Second, even after missionary education was further democratized to accommodate non-slave populations, lack of colonial governmental support and its attendant resource constraint challenges meant that missionaries had to literally turn away scores of students due to limited capacity. Scarce qualified teaching personnel and school infrastructure prompted a form of rationing of missionary educational access during this time. Hence, in my instrumental variables (IV) analysis, I use both grandparents' education (represented by indicator variables taking the value 1 if a grandparent at least completed primary school) to instrument for a given parent's level of education. Given that there are more instruments than potentially endogenous variables, this strategy also allows for checking the credibility of my exclusion restriction—the over-identification test.

The nexus between parental educational attainment and child labor outcomes warrants significant research attention given the growing evidence of strong intergenerational persistence in child labor among households in low-income countries (Emerson and Souza, 2003; Aransiola and Justus, 2017). Educated parents typically demonstrate a proclivity for investing in their children's education, which could rival alternative uses of the child's time such as child labor work. Moreover, the productivity of the parents' time as an input in the child's education may increase with parental schooling (Behrman et al., 1999; Cigno et al., 2002; Andrabi et al., 2012).

For instance, Andrabi et al. (2012) observed that children with educated mothers spend more time on school-related activities at home, which competes with time spent working within and outside the home.

This paper focuses on another channel—parental engagement in non-farm employment. Using a nationally representative survey data set on household demographics and child time use in Malawi, this study’s contributions to the broader child labor literature are threefold. First, it provides deeper insights into parental educational effects on child labor outcomes with a focus on non-farm employment as a key mechanism. Second, by employing an instrumental variables (IV) strategy, this study attempts to address endogeneity issues inherent in the standard estimation of the relationships of interest. Moreover, this study leverages a relatively longer reference period to classify child laborers—an improvement upon earlier approaches, where a one-week reference period is routinely used. Finally, it also accounts for child labor heterogeneity by considering two categories of child labor work: (1) household farm work, and (2) casual, part-time or “ganyu” labor.²³

The IV estimation results indicate that there is a negative, and statistically significant relationship between parental education and “ganyu” labor participation. By contrast, we do not find a significant effect of maternal education on household farm work, while the father’s education is significantly negatively associated with this child labor measure. In other results, I find that maternal education significantly improves school attendance. On the other hand, I do not find a meaningful impact of paternal education on school attendance. These results are robust to relaxing the exclusion restriction assumption. I also find that engagement in non-farm income employment pursuits among educated parents might play a role in mediating these effects.

²³ Where “ganyu” labor refers to any form of low wage, short-term labor arrangement outside the household.

The rest of the paper is organized as follows. Section 3.2 presents a review of the related literature. Section 3.3 presents the data and some descriptive statistics. Section 3.4 illustrates the empirical strategy, and section 3.5 addresses some endogeneity concerns. Section 3.6 summarizes the results, section 3.7 presents a sensitivity analysis, section 3.8 tests some possible mechanisms and section 3.9 concludes.

3.2 Related Literature

Parental education has the potential to reduce child labor participation and improve school attendance (Canagarajah and Coulombe, 1997; Grootaert, 1998; Bhalotra and Heady, 1998; Canagarajah and Nielsen, 1999; Tzannatos, 2003; Kurosaki et al., 2006; Hsin, 2007; Emerson and Souza, 2007; Das and Mukherjee, 2007). While not considered a policy variable per se (Grootaert, 1998), parental education—as a mitigation strategy—can be appealing as it is less intrusive compared to overt child labor bans or prohibitions and has potentially longer-lasting effects (Cigno et al., 2002). Besides altering parental preferences for/against child labor, education also affects parents' labor market choices. Moreover, even among non-altruistic parents, making their children work might no longer be in their best interest if the return to childhood education is sufficiently high. As such, most studies analyzing the correlates of child labor participation often examine both schooling and child labor work as these decisions are interlinked. While estimating a multi-stages sequential probit model, Grootaert (1998) notes that parental education improves the odds of exclusively attending school as well as combining school and work in Côte D'Ivoire. In a similar setting, Canagarah and Coulombe (1997) discuss the influence of factors that jointly determine schooling and child labor decisions among Ghanaian children and found that school participation appear more responsive to parental education. I build on these early contributions to the literature while focusing on parental education as the key variable of interest.

Also, relevant to this line of research is the extent to which the unitary household model in lieu of the collective model adequately captures critical intra-household power dynamics in the decision-making process (Thomas, 1990; Browning et al., 1994; Thomas, 1994; Duflo, 2003). More importantly, the motivation for the collective model raises relevant questions about whose education matters most? This paper also contributes to a growing literature on the heterogeneous impacts of the mother's and father's education on child labor and schooling outcomes. In South Asia, Kurosaki et al. (2006) find direct empirical evidence, suggesting that the mother's education is more important in reducing child labor and improving school attendance in Andhra Pradesh. By contrast, Emerson and Souza (2007) stress that paternal education has a stronger negative influence on the child labor status of sons than the mother's education in Brazil. With respect to school attendance, the authors find that maternal schooling exerts a stronger positive impact on girls' school attendance, whereas the father's education positively predicts higher school attendance for sons.

Consistent with Kurosaki et al. (2006), Das and Mukherjee (2007) also reveal that despite the influence of the father's education, maternal education significantly reduces school dropout rates and child labor incidence among boys in urban India. Similarly, Patrinos and Psacharopoulos (1995) also note that maternal schooling has a strong and negative influence on future employment prospects. Further, Bhalotra and Heady (1998) review evidence of a negative and significant relationship between maternal schooling and household farm work among children in Pakistan and Ghana.

An important gap in the existing literature that remains under-explored pertains to the heterogeneity of child labor work itself. Some child labor activities are undoubtedly more harmful than others. As a consequence, we might expect such forms of child labor to decline dramatically

with improvements in household socioeconomic conditions. Ali (2019) provides some evidence on the importance of accounting for child labor heterogeneity by showing that only the worst forms of child work experienced significant declines with increasing levels of household income. The author interprets this result as perhaps reflecting parents' non-pecuniary motivations behind engaging their children in non-hazardous forms of child work such as unpaid family work. I extend the scope of my research question to also account for this heterogeneity by considering both household farm work and casual, part-time or "ganyu" employment.

3.3 Data

Analysis for this study leverages data from the Malawian Integrated Household Survey (IHS) Program. The IHS is a product of collaborative work between the World Bank and the Malawian National Statistical Office as part of the LSMS - ISA (Living Standards Measurement Study - Integrated Surveys on Agriculture) household survey project. Extending across multiple rounds, the survey started off in 2010 with the implementation of the Third Integrated Household Survey (IHS3). The IHS3 sample was designed to be representative at the national-, regional-, and urban/rural levels. Following the IHS3, the Integrated Household Panel Survey (IHPS) 2013 was administered to follow-up on the 3,246 households initially interviewed in 2010. The tracking of split-off individuals during follow-up resulted in a final IHPS 2013 sample of 4,000 households that could be linked back to 3,104 previously interviewed households during baseline.

The two most recent rounds of the panel survey: the Fourth (IHS4) and Fifth (IHS5) Integrated Household Surveys were conducted in 2016/17 and 2019/20, respectively. Due to funding challenges, the initial sampling frame was halved from 204 enumeration areas (EAs) to 102 during these latter rounds. The IHS4 ended with 2,508 households tracking an original target of 1,989 households in 102 EAs from the IHPS 2013. Following a similar tracking guideline, the

IHS5 grew to include 3,245 households who were interviewed to collect detailed information on individual and household demographic variables, agricultural production, other socioeconomic activities, as well as community-level characteristics. The data were collected using survey questionnaires via interviews with chiefly, the household head. Ultimately, for my analyses, I use the two latest IHPS waves (that is, the IHS4 and IHS5) for reasons I will specify shortly.

A major data limitation that often plagues many child labor studies is finding a suitable measure of child labor outcomes. In many developing country studies, a one-week reference period has been widely used to characterize child labor participation. Such a short reference period, however, may induce very little variation in child labor outcomes, which could be further exacerbated by measurement error resulting in imprecise estimates (Dorman, 2008). As an example, the main child labor measures in the 2010 and 2013 IHPS include: (1) the number of hours in the last 7 days (before the survey) the child spent on agricultural activities, (2) the number of hours in the past week the child run or did any kind of non-agricultural work, and (3) the number of hours spent yesterday collecting water. Since interviewer visits mostly occurred between March and November, which overlaps with the peak season,²⁴ this might predict more child involvement in agricultural work relative other child labor activities (e.g., non-farm work).

Further, to the extent that the one-week reference period is too short to capture any meaningful variation in child labor activities, results may fail to fully reflect the true intensity of child labor work.²⁵ Hence, to partially obviate the threat of finding an insignificant relationship between parental education and child labor for this reason, I use the two latest rounds of the IHPS

²⁴ Peak season refers to a time of the year when crop harvest reaches its maximum. Agricultural labor demand tends to be highest during this time as farmers strive to harvest and get their produce to the market in time to avoid spoilage and loss of quality.

²⁵ Nevertheless, see Andrabi et al. (2012), Kazianga et al. (2012) and Ali (2019) for the use of a similar time frame for the collection of child time use data.

for my analysis. Unlike the earlier waves, the child labor measures represented in the IHS4 and IHS5 are over a relatively longer reference period (specifically, over the past 12 months), circumventing the aforementioned data limitations. Therefore, my two child labor measures include two indicator variables: one for whether the child contributed to household farming activities in the past year and another for whether the child engaged in any casual, part-time or “ganyu” labor in the last 12 months.

For the child level analysis, I restrict the sample to children aged 5 – 17 as that is typically the age range over which the ILO reports child labor statistics. Moreover, since a majority of children within this age bracket are of primary school-going age, this allows for studying the potential trade-off between school attendance and child labor work. Summary statistics on both child- and household-level characteristics for the resulting sample are reported in Table 3.1. In the first column (the “All” column), I pool observations across both years and report descriptive statistics at both the child- and household-level in panels A and B, respectively. The average age of a child in the sample is about 11 years and there is an even split in gender representation. As panel A depicts, roughly half of all observation-years identify as female with females slightly overrepresented in the 2019 panel. A substantively high proportion of children in the sample were reported to be currently attending school or did attend the just ended school session. It can also be inferred from the rather high school attendance rate that few of these children can be considered full-time workers. Roughly 40% of the children in the sample contributed to any household farm work, and about 15% of them engaged in casual, part-time employment. Turning attention to panel B, women appear to receive less education relative to men.²⁶ This result is also reflected in the

²⁶ For the rest of the analysis, for mother’s and father’s education variables with missing values, the missing values are substituted with the value zero and this imputation is controlled for by including indicator variables that take a value of 1 if an observation is missing and zero otherwise as additional controls. This approach is widely utilized in

relatively lower educational attainment rate among both paternal and maternal grandmothers relative to grandfathers in the sample.

Table 3.2 presents evidence on child labor incidence and school attendance for the full sample by child gender and age. As expected, the household farm work incidence ratio is higher for boys (44%) compared to girls (39%). Also, older children (74%) are more likely to be involved in household farming activities relative to younger children (33%). This result seems intuitive given that older children are more likely to be out of school, signaling better availability to support household farm work. Moreover, farm work can be intense; hence, the sturdier build of boys and older children makes them better suited to working on-farm. The data also reveals a gender disparity in casual, part-time or “ganyu” employment. Boys are disproportionately more involved in this type of child labor work. The incidence ratio for casual, part-time employment is roughly 18% for boys, and 13% for girls. Again, the summary statistics indicate that older children have a greater incidence of casual, part-time labor participation. Further, while a substantial proportion of the children in the sample indicated that they do attend school, school attendance is rather low among older children. By contrast, the younger cohort are significantly more likely to attend school. This result partly explains the disproportionately greater fraction of older children participating in the various forms of child work especially household farm work.

3.4 Empirical Strategy

To quantify the effect of parental education on child labor participation, I first estimate the following linear effects model:

$$y_{iht} = \alpha Educ_{ih} + \mathbf{x}_{ht}\boldsymbol{\beta} + \delta_d + \delta_t + \epsilon_{iht} \quad (1)$$

the development economics literature to reduce the number of dropped observations due to missing data (Kurosaki et al., 2006).

where y_{iht} is an indicator variable, which takes the value 1 if child i in household h in time period t such that $t \in \{2016, 2019\}$: (1) was engaged in household farm work, (2) was involved in any casual, part-time or “ganyu” employment in the past 12 months, and zero otherwise; $Educ$ denotes the mother’s or father’s years of schooling; α denotes the effect of parental education on child labor; \mathbf{x}_{ht} is a vector of time-varying controls including household size, a wealth index, number of male and female household members below age 6, total area of cultivated land, household distance from the nearest road; I also include controls for female-headship status, and religious affiliation; β is a vector of parameters on the time-varying household-level covariates to be estimated; δ_d is a district fixed effects term; δ_t is a time dummy; and ϵ_{iht} is an idiosyncratic error term with zero mean and standard deviation, σ_ϵ . Standard errors are clustered at the child-level to allow for correlation of errors for a child across years but not across children.

It is important to note that parents’ education will likely be predetermined by 2016 for a large fraction of children in the sample; hence, the key explanatory variables of interest should not vary much over the survey years, if at all. That is, parental education is more or else a time-constant variable. As such, estimating the relationship between parental education and child labor using the traditional fixed effects approach will likely result in the estimated coefficients of interest getting dropped. This empirical challenge presents a justification for the pooled ordinary least squares (POLS) estimator. However, while the linear probability model (LPM) yields estimates that are easy to interpret, the estimated probabilities can sometimes lie outside the unit interval (that is, above one or below zero). Hence, as a robustness check, I also estimate equation (1) using a non-linear model.

3.5 Addressing Endogeneity

The key explanatory variables of interest—mother’s and father’s education—might fail to satisfy the strict exogeneity assumption in the linear effects model. Reverse causality does not seem to be a problem here as the data set suggests that parental educational attainment would have been determined prior to “future” child labor supply decisions for most children in the sample. That said, correlation between parental education and confounders residing in the idiosyncratic error term that also predict child labor participation will yet yield inconsistent estimates. For example, to the extent that higher household agricultural productivity leads to increased demand for labor, child labor could worsen with higher productivity if hired agricultural labor is scarce. As a consequence, a positive correlation between parental education and household agricultural productivity could result in an underestimation of a negative parental educational effect on child labor.

To address this endogeneity concern, I use an IV strategy. The IV method requires that we have an instrument or set of instruments that are strongly correlated with parental education (the relevance condition) but affect child labor outcomes only through the key explanatory variable(s) of interest (the exclusion restriction). I use grandparents’ literacy as instruments for parental educational attainment.²⁷ In particular, the mother’s education is instrumented by each of her parents’ level of education, which are measured as indicator variables taking the value 1 if they at least completed primary school and 0, otherwise.²⁸ The choice of these instruments is motivated in large part by Malawian grandparents limited discretion in their educational attainment predominantly in the colonial era. In the absence of a clear educational policy, combined with a

²⁷ Variation in the grandparents’ education is partly driven by the quasi-random nature of missionary educational access as explained in the introduction.

²⁸ I instrument for the father’s education in a similar manner.

lack of governmental support, Christian missionaries became the primary custodians of education in Nyasaland pre-independence (McCracken, 2012). For enslaved persons (mostly children) who were freed by the missions, their path to educational attainment was arguably due to chance. Upon rescue, these ex-slaves were trained in missionary schools, with some going on to become priests (McCracken, 2012). Further, even after missionary schools were open to non-slave populations, the financial toll on these missions meant that not all who sought missionary education could be admitted.

As I show later in the results section, it is rather straightforward to see how the chosen set of instruments satisfies the relevance assumption. However, what remains obscure is whether the necessary exclusion restrictions are satisfied since this assumption is not directly testable. That is, is it indeed the case that grandparents' literacy is not correlated with other factors beyond the parents' education that might also influence child labor participation? A potential concern about the IVs due to which the exclusion restriction might be violated is that an educated grandparent can increase the returns to the child's school attendance perhaps by helping out with schoolwork at home, which can impact child labor participation decisions.

As a result of the potential violation of the exclusion restriction for this reason and other related concerns, I control for multigenerational co-residence in all my IV regressions. This strategy is motivated by recent findings in the demography literature, arguing that non-co-resident grandparents' educational attainment exerts little to no influence on grandchildren's schooling outcomes conditional of parental characteristics (Warren and Hauser, 1997; Erola and Moio, 2007; Zeng and Xie, 2014). The multigenerational co-residence variables are measured as two indicator variables: one for whether the grandfather is dead or lives away from the child and another for whether the grandmother is deceased or lives away from the household. I include both

variables in all my IV regressions, but in cases where these two variables are strongly correlated, I control for one or the other due to multicollinearity.

Further, intrinsically linked to the missionaries' educational curricula was the mandate to produce native purveyors of the Christian faith. As such, for the predominantly Islamic share of the population, self-selection out of missionary education will be common (Bone, 1982). Hence, I also include religion dummies as additional controls. Nevertheless, as a robustness check, I also obtain 2SLS estimates while dropping parents who indicated to be Muslim to investigate the stability of my results.²⁹

3.6 Results

3.6.1 Descriptive Statistics

In Figures 3.2 and 3.3, I plot raw means for the two child labor outcomes against parental educational status for both survey years. A few important patterns emerge. First, Figure 3.2 shows that household farm labor participation is relatively common among children with uneducated parents. This pattern holds irrespective of the parent's gender. Second, I find that household farm labor participation is rather remarkably stable over time among children with educated parents. By contrast, I observe an uptick in this outcome variable over time when parents are uneducated. Turning attention to Figure 3.3, we observe patterns that diverge somewhat from the trends reported in Figure 3.2. In relative terms, participation in this form of child labor work appears more pervasive among children with uneducated parents. However, the figure shows that participation rates in casual, part-time or "ganyu" employment worsens over time for both sub-groups

²⁹ While data on grandparents' religious affiliations would have been most beneficial for this exercise, such information on the grandparents in the data set is rather sparse. Hence, I use parents' religious affiliations as a proxy. Insofar as children adopt their parents' religion in this setting, this exercise still serves its intended purpose.

irrespective of parental educational status. From a policy standpoint, this finding reflects in part the importance of accounting for child labor heterogeneity.

Figures 3.4 and 3.5 illustrate the relationship between child labor outcomes and household wealth graphically. While Figure 3.5 shows a consistently negative relationship between casual, part-time or “ganyu” employment and household wealth quintiles across survey waves, the pattern suggested by Figure 3.4 is somewhat non-linear. Figure 3.4 indicates that child household farm work participation initially worsens with household wealth then begins to improve at extremely high levels of wealth. This finding is in line with Bhalotra and Heady (1998)’s discovery that child labor on household farms could worsen with wealth in the presence of multiple factor market failures.

3.6.2 Regression Results

Linear probability model estimates for equation (1) are reported in the first two columns of Table 3.3. The columns present estimated coefficients of parental educational effects on household farm labor participation, and casual, part-time employment, respectively. Column (3) shows how school attendance responds to changes in parental education. Across all columns, I include controls for household-level covariates including the household size, a wealth index, number of male and female household members under age 6, household area of cultivated land, distance to the nearest road, religious affiliation of the household head, and female headship status. A few results stand out.

The estimated coefficients for the mother’s educational attainment variable are negative and statistically significant for both child labor measures as reported in columns (1) and (2). In particular, an additional year of maternal schooling is associated with a 0.4 (0.7) percentage points decline in the likelihood of child labor involvement in household farm work (casual, part-time

employment), on average. There is also a strong and negative association between paternal education and “ganyu” labor. By contrast, the father’s education does not exert a significant effect on household farm work. Similarly, results in column (2) indicate that an additional year of the father’s schooling decreases child labor participation in casual, part-time labor employment by 0.9 percentage points, on average, *ceteris paribus*.

Table 3.3 also reports the effect of parental educational attainment on school attendance. Results are presented in column (3). Consistent with Kurosaki et al. (2006), school attendance appears more responsive to maternal schooling. The point estimate of the coefficient on maternal education is 0.01, while the estimated coefficient for the father’s education is 0.003—both estimated coefficients are statistically significant. In Table 3.4, I re-estimate equation (1) for both child labor outcomes and the school attendance dependent variable using a probit model. The estimated average partial effects from these probit models appear remarkably similar to the LPM estimates. Hence, in what follows, I prioritize the LPM estimates for ease of interpretation.

Table 3.5 reports the estimated effects of parental education on child labor outcomes and school attendance by the child’s gender. While girls appear less likely to engage in “ganyu” employment, I do not find any significant heterogeneous effects of parental education on child labor and school attendance by child gender. The fact that we do not find significant sex-specific parental educational effects on child time use suggests waning discrimination in human capital investments against girls.

Next, I present results from the two-stage least squares (2SLS) estimator. First, I report estimates from the first stage of the IV analysis in Table 3.6. In column 1 (2), I report estimates from the regression of the mother’s (father’s) years of education on the maternal (paternal) grandparents’ literacy variables and other “exogenous” covariates for the full sample. Columns (3)

through (6) show first stage results disaggregated by child gender. I broadly find evidence of a strong correlation between parental educational attainment and grandparents' literacy.

Table 3.7 reports the 2SLS estimation results. Columns (1)–(3) present the estimated coefficients for the two child labor measures and school attendance for the full sample in that order. The 2SLS estimates for the mother's educational attainment variable are reported in panel A, while panel B presents the 2SLS estimates for the father's education variable. Following Olea and Pflueger (2013), I report the effective F statistic from a heteroskedastic, and cluster-robust test of the null of weak instruments across my IV specifications. A rejection of the null hypothesis signals a strong first stage. Further, I also report the Hansen J statistic with its corresponding p-value from the test of the null that the over-identifying restrictions are indeed valid. Failure to reject the null in favor of the alternative hypothesis lends credence to the assumption that the necessary exclusion restrictions are satisfied.

The weak IV tests reveal that the reported effective F statistics exceed the critical value for the $\tau = 30\%$ weak instrument threshold across all specifications. That is, we can conclude that the instruments are strong. Second, the over-identification tests are reassuring as I broadly fail to reject the null hypothesis that the over-identifying restrictions are valid. I now turn to the results.

First, I do not find a significant association between maternal education and child household farm work participation. That is, after instrumentation, the effect of the mother's education on household farm work is attenuated (that is, it tends toward zero). By contrast, there is a strong negative association between maternal education and "ganyu" labor involvement. In particular, an additional year of maternal schooling is associated with a 1.5 percentage points decrease in casual, part-time employment.

Second, I find a strong positive association between the mother's education and school attendance. Turning attention to panel B, the 2SLS estimates indicate a negative and statistically significant impact of paternal education on both child labor measures. Specifically, an additional year of paternal schooling is associated with a 2.6 (2.3) percentage points decline in the incidence of household farm work (casual, part-time employment), on average. On the other hand, the estimated coefficient for the school attendance outcome variable is not statistically different from zero.

Table 3.8 presents 2SLS estimates of the impact of parental education on child time use by child gender. Panel A reports similar effects of maternal education on "ganyu" labor across gender. By contrast, I do not find a significant effect of maternal schooling on female school attendance, while the mother's education strongly improves male school attendance. Similarly, the negative effect of paternal education on "ganyu" labor is only significant for the male subsample.

As a robustness check, I re-run my IV analysis while restricting the sample to non-Islamic parents. One might be concerned about selection of Islamic grandparents out of formal education due to Christian bias in the missionary educational curriculum. Results are reported in Table 3.9. The 2SLS estimates using this restricted sample are remarkably similar to my main results in Table 3.7. The insensitivity of the main results to this robustness check suggests that potential selection of Islamic grandparents out of missionary education does not bias my main findings in any meaningful way.

Next, I explore how child labor outcomes respond to parental educational attainment for children of differing age groups. To obtain these estimates, I interact the parental education variables with age dummies to estimate these heterogeneous effects. Results are presented in Figures 3.6 and 3.7 for the maternal educational attainment effect while Figures 3.8 and 3.9 present

the 2SLS estimates for the father's education by child age. Some general patterns emerge. I do find some evidence of heterogeneous parental educational attainment effects by child age. In particular, I find that the estimated coefficients are statistically indistinguishable from zero for relatively younger children. By contrast, the estimated effects of parental education on the child labor outcomes for older children is negative and statistically significant although less precisely estimated (that is, the confidence intervals are larger). This finding could be rationalized in part by the relatively lower child labor participation rate among younger children to begin with.

3.7 Imperfect Instruments Sensitivity Analysis

In this sub-section, I examine the sensitivity of my IV results to a relaxation of the exclusion restriction. Following Conley et al. (2012), I obtain bounds on the causal effect of parental education, while allowing for a direct effect of grandparents' literacy on child time use. While the over-identification tests suggest that the instruments may be valid, they are only necessary, but not sufficient conditions for instrument validity (Clarke and Matta, 2018).

Consider the IV model below:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\epsilon} \quad (2)$$

$$\mathbf{X} = \mathbf{Z}\boldsymbol{\Pi} + \mathbf{V} \quad (3)$$

where \mathbf{Y} is a vector of the child time use variables; \mathbf{X} is a vector of the parental education variables; \mathbf{Z} are the instruments (grandparents' literacy); $\boldsymbol{\Pi}$ is a vector of first-stage coefficients; $\boldsymbol{\gamma}$ captures the direct effect of the instruments on the outcome variables. The exclusion restriction implies that $\boldsymbol{\gamma} = 0$, signaling that the instruments affect child time use only through parental education.

The imperfect instruments framework allows for relaxing the $\boldsymbol{\gamma} = 0$ assumption. In particular, I assume that there is a direct negative association between grandparents' literacy and

child labor. In doing so, I set priors such that γ falls within the range $[\gamma_{min}, 0]$, where $\gamma_{min} \in \{-0.001, -0.002, -0.003\}$ to capture the degree of violation of the exclusion restriction.³⁰ Bounds are then obtained as the union of all confidence intervals for γ inside the assumed range $[\gamma_{min}, 0]$.³¹ Results are presented in Tables 3.10 and 3.11 for the maternal and paternal education effects, respectively. The results indicate that the estimated bounds are relatively robust to worsening violations of the exclusion restriction.³² Reassuringly, the 2SLS estimates fall within the estimated bounds, which do not include zero for the significant 2SLS results. Hence, despite substantial deviations from perfect exogeneity, my 2SLS results are robust to varying degrees of violation of the exclusion restriction.³³

3.8 Potential Mechanisms

Higher educational attainment is typically associated with greater non-farm labor force participation. Strong pull factors such as the relatively higher expected returns from non-farm employment can induce a preference for non-farm engagements among the educated. Hence, in identifying the effect of parental education on child labor outcomes, the role of the parents' occupation cannot be ignored. In what follows, I explore whether educated parents are more likely to participate in non-farm employment activities. A positive result from this investigation will partly explain the strong and negative effect of parental education on child labor work, especially household farm work. Indeed, I find strong evidence indicative of positive sorting among educated parents into non-farm business engagements and wage employment. I test for this evidence using the 2SLS estimates from instrumenting for parents' education with the corresponding

³⁰ For the school attendance variable, γ_{min} becomes $\gamma_{max} \in \{0.001, 0.002, 0.003\}$.

³¹ See Clarke and Matta (2018) for details on the union of confidence intervals (UCI) procedure.

³² Note that the priors on γ need not be the same for both instruments and may be extended to differing violations of perfect exogeneity.

³³ Some of the priors on γ are as high as 90% of the POLS estimates of parental education on child time use.

grandparents' literacy indicators. Errors are clustered at the parent-level and results are reported in Table 3.12.

Column (1) reports that an additional year of the mother's (father's) education is associated with a 4.2 (7.9) percentage points increase in non-farm business participation. I equally find strong positive effects of parental education on the likelihood of wage employment for both mothers and fathers. Taken together, these results have noteworthy implications. First, given that education improves the odds of non-farm engagement and wage employment, which usually requires parents to be away from home, we might expect child labor to decline if child labor work typically requires close parental supervision. Second, for non-farm households with relatively younger children, there are good reasons to expect that keeping their children in school as they (the parents) work is typically preferred. Hence, these children by design are exempt from any form of child labor work. However, in a "full" parent household, this mechanism will depend on whether both parents are engaged in non-farm and/or wage employment. One can envision a scenario where the mother is a salaried employee while the father attends to the household farm. In that case, child labor might not fall with parental education if multiple factor markets such as land and labor are missing (Bhalotra and Heady, 1998), as the child's services will be needed on-farm.

3.9 Conclusions

Child labor remains a pervasive phenomenon in sub-Saharan Africa. Given laws at both national and international levels to minimize child labor, the innocuous nature of household child labor participation makes it less noticeable and challenging to eradicate. In this paper, I revisit an important empirical question: does child labor respond inversely to parental education? There is a wide scope of anecdotal evidence suggesting that parental education reduces child labor participation; however, studies that attempt to address possible endogeneity issues as well as child

labor work heterogeneity are rare. Moreover, very few studies have attempted to explore parental engagement in non-farm employment as a potential mechanism driving these effects. To assess the sensitivity of my results to violations of the exclusion restriction, I employ the imperfect instruments method proposed by Conley et al. (2012).

Using a nationally representative Malawian panel data set, I find that parental education is generally child labor mitigating. There is a strong and negative effect of maternal schooling on “gangyu” labor involvement, but no effect on household farm work. In particular, an additional year of maternal (paternal) schooling is roughly associated with a 1.5 (2.3) percentage points decline in casual, part-time or “ganyu” employment, on average. Similarly, the return for an additional year of paternal schooling is a 2.6 percentage points decrease in household farm work. I find limited evidence of differing estimated effects by child gender for my LPM estimates; however, the estimated effects appear more pronounced for boys and older children for the 2SLS estimates. Results suggest that the impact of parental education on both child labor measures are mostly driven by older children, who are more likely to work on household farms at that age. The study’s findings also indicate that child school attendance improves especially with higher maternal education. This finding is consistent with Das and Mukherjee (2007) and Kurosaki et al. (2006), who also find strong and positive effects of maternal schooling on child school attendance. Nevertheless, evidence of such effects on child school attendance is weak for the paternal education variable.

Finally, I also show that parental engagement in non-farm employment pursuits could be a mechanism underlying the negative effect of parental education on child labor outcomes. I find strong evidence that educated parents are more likely to engage in non-farm businesses and wage employment. Nonetheless, there are a few caveats to consider. Obviously, there could be other

pathways through which the effect of parental education on child labor could be mediated. In addition, further analysis is required to uncover how parental engagement in the non-farm economy directly impacts child time use. Is it earned non-farm income or the transition to the non-farm sector per se that predicts lower child labor participation? Supplementary qualitative data via interviews can provide additional insights into this empirical question.

BIBLIOGRAPHY

- Ali, F. R. M. (2019). In the same boat, but not equals: The heterogeneous effects of parental income on child labour. *The Journal of Development Studies*, 55(5):845–858.
- Allen, J. (2008). Slavery, Colonialism, and the Pursuit of Community Life: Anglican Mission Education in Zanzibar and Northern Rhodesia 1864–1940. *History of Education*.
- Andrabi, T., Das, J., and Khwaja, A. I. (2012). What did you do all day? Maternal education and child outcomes. *Journal of Human Resources*, 47(4):873–912.
- Appleton, S. and Balihuta, A. (1996). Education and agricultural productivity: evidence from Uganda. *Journal of International Development*, 8(3):415–444.
- Aransiola, T. J. and Justus, M. (2017). Intergenerational persistence of child labor in Brazil. In *International Conference on Applied Economics*, pages 613–630. Springer.
- Behrman, J. R., Foster, A. D., Rosenweig, M. R., and Vashishtha, P. (1999). Women’s schooling, home teaching, and economic growth. *Journal of Political Economy*, 107(4):682–714.
- Bhalotra, S. and Heady, C. (1998). Child labour in rural Pakistan and Ghana. University of Bristol and University of Bath, Bristol, mimeo.
- Bone, D. S. (1982). Islam in Malawi. *Journal of Religion in Africa*, 13:126–138.
- Browning, M., Bourguignon, F., Chiappori, P.-A., and Lechene, V. (1994). Income and outcomes: A structural model of intrahousehold allocation. *Journal of Political Economy*, 102(6):1067–1096.
- Canagarajah, S. and Coulombe, H. (1997). Child labor and schooling in Ghana. Available at SSRN 620598.
- Canagarajah, S. and Nielsen, H. S. (1999). Child labor and schooling in Africa: A comparative study. World Bank, Social Protection Team.
- Chan, T. W. and Boliver, V. (2013). The grandparents’ effect in social mobility: Evidence from British birth cohort studies. *American Sociological Review*, 78(4):662–678.
- Cigno, A., Rosati, F. C., and Tzannatos, Z. (2001). Child labor, nutrition, and education in rural India: An economic analysis of parental choice and policy options. Washington, DC: The World Bank.
- Cigno, A., Rosati, F. C., and Tzannatos, Z. (2002). *Child Labor Handbook*. Washington: The World Bank.
- Clarke, D. and Matta, B. (2018). Practical considerations for questionable IVs. *The Stata Journal*, 18(3):663–691.

- Conley, T. G., Hansen, C. B., and Rossi, P. E. (2012). Plausibly exogenous. *Review of Economics and Statistics*, 94(1):260–272.
- Das, S. and Mukherjee, D. (2007). Role of women in schooling and child labour decision: The case of urban boys in India. *Social Indicators Research*, 82(3):463–486.
- Dorman, P. (2008). Child labour, education, and health: A review of the literature. ILO Geneva.
- Duflo, E. (2003). Grandmothers and granddaughters: old-age pensions and intrahousehold allocation in South Africa. *The World Bank Economic Review*, 17(1):1–25.
- Dumas, C. (2020). Productivity Shocks and Child Labor: The Role of Credit and Agricultural Labor Markets. *Economic Development and Cultural Change*, 68(3):763–812.
- Emerson, P. M. and Souza, A. P. (2003). Is there a child labor trap? Intergenerational persistence of child labor in Brazil. *Economic Development and Cultural Change*, 51(2):375–398.
- Emerson, P. M. and Souza, A. P. (2007). Child labor, School Attendance, and Intrahousehold Gender Bias in Brazil. *The World Bank Economic Review*, 21(2):301–316.
- Erola, J. and Moisio, P. (2007). Social mobility over three generations in Finland, 1950–2000. *European Sociological Review*, 23(2):169–183.
- Grootaert, C. (1998). Child labor in Cote d’Ivoire: incidence and determinants, volume 1905. World Bank Publications.
- Hayami, Y. and Ruttan, V. W. (1970). Agricultural productivity differences among countries. *The American Economic Review*, 60(5):895–911.
- Hsin, A. (2007). Children’s time use: Labor divisions and schooling in Indonesia. *Journal of Marriage and Family*, 69(5):1297–1306.
- ILO (2017). Global Estimates of Child Labour: Results and Trends, 2012–2016. ILO (2021). Child Labour Global Estimates 2020, Trends and the Road Forward.
- Jæger, M. M. (2012). The extended family and children’s educational success. *American Sociological Review*, 77(6):903–922.
- Kazianga, H., De Walque, D., and Alderman, H. (2012). Educational and child labour impacts of two food-for-education schemes: Evidence from a randomised trial in rural Burkina Faso. *Journal of African Economies*, 21(5):723–760.
- Kurosaki, T., Ito, S., Fuwa, N., Kubo, K., and Sawada, Y. (2006). Child labor and school enrollment in rural India: Whose education matters? *The Developing Economies*, 44(4):440–464.
- McCracken, J. (2012). A History of Malawi: 1859 - 1966. Boydell & Brewer Inc.

- Olea, J. L. M. and Pflueger, C. (2013). A robust test for weak instruments. *Journal of Business & Economic Statistics*, 31(3):358–369.
- Patrinos, H. A. and Psacharopoulos, G. (1995). Educational performance and child labor in Paraguay. *International Journal of Educational Development*, 15(1):47–60.
- Reardon, T., Stamoulis, K., Balisacan, A., Cruz, M., Berdegúe, J., and Banks, B. (1998). Rural non-farm income in developing countries. *The State of Food and Agriculture*, 1998:283–356.
- Reggio, I. (2011). The influence of the mother’s power on her child’s labor in Mexico. *Journal of Development Economics*, 96(1):95–105.
- Reimers, M. and Klasen, S. (2013). Revisiting the role of education for agricultural productivity. *American Journal of Agricultural Economics*, 95(1):131–152.
- Rosati, F. C. and Tzannatos, Z. (2000). Child labor in Vietnam: An Economic Analysis. The World Bank: mimeo.
- Singh, I., Squire, L., and Strauss, J. (1986). Agricultural Household Models: Extensions, Applications, and Policy. Number 11179. The World Bank.
- Thomas, D. (1990). Intra-household resource allocation: An inferential approach. *Journal of Human Resources*, 635–664.
- Thomas, D. (1994). Like father, like son; like mother, like daughter: Parental resources and child height. *Journal of Human Resources*, 950–988.
- Tzannatos, Z. (2003). Child labor and school enrollment in Thailand in the 1990s. *Economics of Education Review*, 22(5):523–536.
- Warren, J. R. and Hauser, R. M. (1997). Social stratification across three generations: New evidence from the Wisconsin Longitudinal Study. *American Sociological Review*, pages 561–572. 20
- Zeng, Z. and Xie, Y. (2014). The effects of grandparents on children’s schooling: Evidence from rural China. *Demography*, 51(2):599–617.

APPENDIX A: TABLES AND FIGURES

Table 3.1

Summary Statistics			
		Mean (S.E)	
Variable	All	2016	2019
Panel A. Child Characteristics			
Age (in years)	10.82 (3.60)	10.69 (3.59)	10.93 (3.61)
Female (0/1)	0.51	0.50	0.51
Attends school (0/1)	0.92	0.92	0.92
Contributes to household farm work (0/1)	0.42	0.41	0.42
Engaged in casual, part-time employ. (0/1)	0.15	0.13	0.17
Observations	7,133	3,193	3,940
Panel B. Household Characteristics			
Mother's education (in years)	5.35 (3.68)	5.04 (3.53)	5.59 (3.78)
Father's education (in years)	6.71 (3.97)	6.44 (3.88)	6.92 (4.04)
Mother's age (in years)	37.95 (9.96)	37.87 (10.01)	38.00 (9.93)
Father's age (in years)	43.20 (10.87)	43.19 (10.59)	43.21 (11.09)
Household size	5.79 (1.94)	5.95 (1.93)	5.68 (1.93)
Number of male HH members under 6	0.53 (0.70)	0.52 (0.70)	0.53 (0.70)
Number of female household members under 6	0.54 (0.70)	0.55 (0.73)	0.52 (0.68)
Area of cultivated land (in acres)	1.83 (1.88)	1.88 (2.00)	1.80 (1.77)
Maternal grandmother is educated (0/1)	0.07	0.06	0.07
Maternal grandfather is educated (0/1)	0.16	0.17	0.15
Paternal grandmother is educated (0/1)	0.05	0.04	0.05
Paternal grandfather is educated (0/1)	0.11	0.11	0.12
Female head (0/1)	0.27	0.24	0.28
Religion			
% No religion	1.96	2.44	1.59
% Traditional	0.63	0.03	1.10
% Christian	79.44	79.26	79.58
% Islam	15.29	15.76	14.93
% Other religion	0.39	0.32	0.44
Observations	2,635	1,109	1,526

Notes: Summary statistics are reported on households with children aged 5 – 17 with observations weighted using the 2016 panel sampling weights. Standard errors are reported in parentheses.

Table 3.2*Child labor incidence and school attendance by gender and age*

Variable	Gender				Age Cohort		
	Total	Girls	Boys	<i>p-value:</i> Δ	5 - 14 yrs	15 - 17 yrs	<i>p-value:</i> Δ
Household farm work	41.6	39.4	43.8	0.000	33.1	74.2	0.000
Casual, part-time or "ganyu" labor	15.4	13.0	17.8	0.000	10.1	35.7	0.000
Attends school	92.1	91.7	92.5	0.203	96.6	76.2	0.000

Notes: Sample means are reported as percentages using the pooled sample across the 2016 and 2019 panel waves.

Source: Author's own calculations.

Table 3.3*Estimates of parental education effects on child time use*

	(1)	(2)	(3)
	Linear Probability Model		
	Dependent Variables		
Variable	HH Farm Work (0/1)	“Ganyu” labor (0/1)	Attends school (0/1)
Mother’s education (years)	-0.004* (0.002)	-0.006*** (0.002)	0.009*** (0.001)
Father’s education (years)	-0.001 (0.002)	-0.009*** (0.002)	0.003** (0.001)
Controls	✓	✓	✓
District FE	✓	✓	✓
Year dummy	✓	✓	✓
Observations	6,951	6,951	6,334

Notes: Standard errors are reported in parentheses and are clustered at the child level. Control variables include household size, wealth index, size of household cultivated land, number of female household members under age 6, number of male household members under age 6, child’s gender, religion, female headship status, and household distance from nearest road. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.4*Average partial effects of parental education on child time use from probit model*

	(1)	(2)	(3)
	Dependent Variables		
Variable	HH Farm Work (0/1)	"Ganyu" labor (0/1)	Attends school (0/1)
Mother's education (years)	-0.004* (0.002)	-0.007*** (0.002)	0.010*** (0.001)
Father's education (years)	-0.001 (0.002)	-0.009*** (0.002)	0.003** (0.001)
Controls	✓	✓	✓
District FE	✓	✓	✓
Year dummy	✓	✓	✓
Observations	6,951	6,951	6,334

Notes: Control variables include household size, wealth index, size of household cultivated land, number of female household members under age 6, number of male household members under age 6, child's gender, religion, female headship status, and household distance from nearest road. Standard errors are reported in parentheses and are clustered at the child level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.5*Effect of parental education on child time use by gender - LPM*

Variable	(1)	(2)	(3)
	Dependent Variables		
	HH Farm Work (0/1)	"Ganyu" labor (0/1)	Attends school (0/1)
Mother's education (years)	-0.003 (0.003)	-0.006*** (0.002)	0.008*** (0.002)
Father's education (years)	-0.002 (0.003)	-0.010*** (0.002)	0.003** (0.001)
1[<i>Girl</i> = 1]	-0.022 (0.022)	-0.048*** (0.018)	-0.023 (0.015)
Mother's educ \times 1[<i>Girl</i> = 1]	-0.002 (0.004)	-0.001 (0.003)	0.003 (0.002)
Father's educ \times 1[<i>Girl</i> = 1]	0.001 (0.004)	0.003 (0.003)	-0.001 (0.002)
Controls	✓	✓	✓
District FE	✓	✓	✓
Year dummy	✓	✓	✓
Observations	6,951	6,951	6,334

Notes: Standard errors are reported in parentheses and are clustered at the child level. Control variables include household size, wealth index, size of household cultivated land, number of female household members under age 6, number of male household members under age 6, religion, female headship status, and household distance from nearest road. 1[*Girl* = 1] is an indicator for whether the child is a female. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.6

First stage regression results – LPM estimates

	(1)	(2)	(3)	(4)	(5)	(6)
	Full sample		Females		Males	
	Dependent variables					
Variable	Mother's education	Father's education	Mother's education	Father's education	Mother's education	Father's education
Grandmother's literacy (0/1)	1.494*** (0.193)	0.817*** (0.245)	1.301*** (0.265)	0.839*** (0.320)	1.719*** (0.282)	0.782** (0.369)
Grandfather's literacy (0/1)	1.800*** (0.144)	1.536*** (0.178)	2.097*** (0.216)	1.732*** (0.243)	1.476*** (0.187)	1.312*** (0.262)
Household size	-0.244*** (0.027)	-0.060*** (0.023)	-0.213*** (0.039)	-0.068** (0.031)	-0.271*** (0.037)	-0.045 (0.033)
Cultivated plot area (acres)	0.063** (0.026)	0.022 (0.027)	0.055 (0.035)	-0.009 (0.037)	0.068* (0.038)	0.052 (0.040)
Wealth index	-0.006*** (0.001)	0.008*** (0.001)	0.006*** (0.001)	0.008*** (0.001)	0.007*** (0.001)	0.008*** (0.001)
No. of female HH members < age 6	0.255*** (0.059)	0.040 (0.061)	0.313*** (0.081)	0.129 (0.080)	0.366*** (0.090)	-0.080 (0.095)
No. of male HH members < age 6	0.336*** (0.061)	0.027 (0.061)	0.221** (0.090)	-0.022 (0.090)	0.282*** (0.078)	0.051 (0.083)
Female household head (0/1)	-0.298 (0.229)	-0.148 (0.254)	-0.569 (0.322)	0.184 (0.371)	-0.144 (0.320)	-0.427 (0.350)
Distance to nearest road (km)	-0.024*** (0.006)	-0.019*** (0.007)	-0.011 (0.009)	-0.023** (0.009)	-0.039*** (0.009)	-0.014 (0.009)
Mother's education (years)		0.305*** (0.015)		0.295*** (0.021)		0.315*** (0.022)
Father's education (years)	0.301*** (0.014)		0.295*** (0.020)		0.308*** (0.021)	
Constant	4.290*** (1.094)	8.8086*** (0.511)	4.551*** (1.498)	7.919*** (0.716)	2.791*** (0.799)	7.893*** (0.720)
Multigenerational co- residence controls	✓	✓	✓	✓	✓	✓

Table 3.6 (cont'd)

Religion dummies	✓	✓	✓	✓	✓	✓
District FE	✓	✓	✓	✓	✓	✓
Year dummy	✓	✓	✓	✓	✓	✓
Observations	6,790	6,951	3,536	2,800	3,415	2,801

Notes: Standard errors are reported in parentheses and are clustered at the child level. *p<0.10, **p<0.05, ***p<0.01

Table 3.7*2SLS estimates of the impact of parental education on child time use*

Variable	(1)	(2)	(3)
	Dependent Variables		
	HH Farm work (0/1)	"Ganyu" labor (0/1)	Attends school (0/1)
Panel A: Mother's education instrumented			
Mother's education	0.005 (0.008)	-0.015*** (0.005)	0.010*** (0.004)
Multigenerational co-residence controls	✓	✓	✓
Other Controls	✓	✓	✓
Religion dummies	✓	✓	✓
District FE	✓	✓	✓
Year dummy	✓	✓	✓
Montiel Olea & Pflueger <i>F</i> stat	195.19	195.19	185.40
Hansen <i>J</i> stat (p-value)	1.89 (0.17)	0.00 (0.99)	2.83 (0.09)
Observations	6,951	6,951	6,334
Panel B: Father's education instrumented			
Father's education	-0.026** (0.013)	-0.023*** (0.008)	-0.005 (0.006)
Multigenerational co-residence controls	✓	✓	✓
Other Controls	✓	✓	✓
District FE	✓	✓	✓
Religion dummies	✓	✓	✓
Year dummy	✓	✓	✓
Montiel Olea & Pflueger <i>F</i> stat	63.11	63.11	61.60
Hansen <i>J</i> stat (p-value)	0.02 (0.90)	1.16 (0.28)	2.14 (0.14)
Observations	5,736	5,736	5,229

Notes: Standard errors are reported in parentheses and are clustered at the child level. Other control variables include household size, wealth index, size of household cultivated land, number of female household members under age 6, number of male household members under age 6, religion, female headship status, and household distance from nearest road. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.8*2SLS estimates of the impact of parental education on child time use by gender*

	(1)	(2)	(3)	(4)	(5)	(6)
	Females			Males		
Variable	HH Farm work (0/1)	“Ganyu” labor (0/1)	Attends school (0/1)	HH Farm work (0/1)	“Ganyu” labor (0/1)	Attends school (0/1)
Panel A: Mother’s education instrumented						
Mother’s education	0.000 (0.010)	-0.012* (0.007)	0.007 (0.005)	0.010 (0.011)	-0.015* (0.008)	0.014*** (0.005)
Multi. co-residence	✓	✓	✓	✓	✓	✓
Other controls	✓	✓	✓	✓	✓	✓
Religion dummies	✓	✓	✓	✓	✓	✓
District FE	✓	✓	✓	✓	✓	✓
Year dummy	✓	✓	✓	✓	✓	✓
Montiel Olea & Pflueger F stat	92.26	92.26	107.40	93.74	93.74	81.05
Hansen J stat (p-value)	0.01 (0.92)	0.00 (0.95)	2.57 (0.11)	1.63 (0.20)	0.03 (0.86)	0.70 (0.40)
Observations	3,536	3,536	3,245	3,415	3,415	3,089
Panel B: Father’s education instrumented						
Father’s education	-0.019 (0.016)	-0.000 (0.009)	-0.011 (0.009)	-0.021 (0.018)	-0.032*** (0.012)	0.001 (0.008)
Multi. co-residence	✓	✓	✓	✓	✓	✓
Other controls	✓	✓	✓	✓	✓	✓
Religion dummies	✓	✓	✓	✓	✓	✓
District FE	✓	✓	✓	✓	✓	✓
Year dummy	✓	✓	✓	✓	✓	✓
Montiel Olea & Pflueger F stat	43.12	43.12	38.82	31.55	31.55	24.04
Hansen J stat (p-value)	0.14 (0.71)	0.55 (0.46)	1.54 (0.21)	0.02 (0.89)	1.54 (0.22)	0.57 (0.45)
Observations	2,868	2,868	2,637	2,868	2,868	2,592

Notes: Standard errors are reported in parentheses and are clustered at the child level. Control variables include household size, wealth index, size of household cultivated land, number of female household members under age 6, number of male household members under age 6, religion, female headship status, and household distance from nearest road. *p<0.10, **p<0.05, ***p<0.01

Table 3.9*2SLS estimates of the impact of parental education on child time use – robustness check*

	(1)	(2)	(3)
	Omitted Muslim sample		
	Dependent Variables		
Variable	HH Farm work (0/1)	“Ganyu” labor (0/1)	Attends school (0/1)
Panel A: Mother’s education instrumented			
Mother’s education	0.009 (0.009)	-0.014** (0.006)	0.012*** (0.004)
Multigenerational co-residence controls	✓	✓	✓
Other Controls	✓	✓	✓
Religion dummies	✓	✓	✓
District FE	✓	✓	✓
Year dummy	✓	✓	✓
Montiel Olea & Pflueger <i>F</i> stat	142.96	142.96	136.84
Hansen <i>J</i> stat (p-value)	1.24 (0.26)	0.07 (0.79)	6.92 (0.01)
Observations	5,852	5,852	5,341
Panel B: Father’s education instrumented			
Father’s education	-0.025* (0.015)	-0.028*** (0.009)	-0.006 (0.007)
Multigenerational co-residence controls	✓	✓	✓
Other Controls	✓	✓	✓
Religion dummies	✓	✓	✓
District FE	✓	✓	✓
Year dummy	✓	✓	✓
Montiel Olea & Pflueger <i>F</i> stat	53.59	53.59	51.68
Hansen <i>J</i> stat (p-value)	0.06 (0.81)	2.01 (0.16)	2.97 (0.08)
Observations	4,845	4,845	4,425

Notes: Standard errors are reported in parentheses and are clustered at the child level. Other control variables include household size, wealth index, size of household cultivated land, number of female household members under age 6, number of male household members under age 6, religion, female headship status, and household distance from nearest road. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.10*2SLS maternal educational impact - relaxing $\gamma = \mathbf{0}$ assumption*

	Estimated coefficient	Lower bound	Upper bound
Panel A: HH Farm Work			
$\gamma_1 = \gamma_2 = -0.001$	0.005	-0.012	0.014
$\gamma_1 = \gamma_2 = -0.002$	0.005	-0.012	0.015
$\gamma_1 = \gamma_2 = -0.003$	0.005	-0.012	0.015
Panel B: "Ganyu" labor			
$\gamma_1 = \gamma_2 = -0.001$	-0.015***	-0.024	-0.007
$\gamma_1 = \gamma_2 = -0.002$	-0.015***	-0.024	-0.006
$\gamma_1 = \gamma_2 = -0.003$	-0.015***	-0.024	-0.006
Panel C: Attends school			
$\gamma_1 = \gamma_2 = 0.001$	0.010***	0.002	0.018
$\gamma_1 = \gamma_2 = 0.002$	0.010***	0.002	0.018
$\gamma_1 = \gamma_2 = 0.003$	0.010***	0.001	0.018

Notes: γ_1 and γ_2 represent the direct effect of the maternal grandparents' literacy variables on child time use. Bounds derived from Conley et al. (2012)'s union of confidence intervals method. * p < 0.10, ** p < 0.05, *** p < 0.01

Table 3.11*2SLS paternal educational impact - relaxing $\gamma = \mathbf{0}$ assumption*

	Estimated coefficient	Lower bound	Upper bound
Panel A: HH Farm Work			
$\gamma_1 = \gamma_2 = -0.001$	-0.026**	-0.048	-0.005
$\gamma_1 = \gamma_2 = -0.002$	-0.026**	-0.048	-0.004
$\gamma_1 = \gamma_2 = -0.003$	-0.026**	-0.048	-0.003
Panel B: "Ganyu" labor			
$\gamma_1 = \gamma_2 = -0.001$	-0.023***	-0.036	-0.009
$\gamma_1 = \gamma_2 = -0.002$	-0.023***	-0.036	-0.008
$\gamma_1 = \gamma_2 = -0.003$	-0.023***	-0.036	-0.007
Panel C: Attends school			
$\gamma_1 = \gamma_2 = 0.001$	-0.005	-0.014	0.007
$\gamma_1 = \gamma_2 = 0.002$	-0.005	-0.015	0.007
$\gamma_1 = \gamma_2 = 0.003$	-0.005	-0.015	0.007

Notes: γ_1 and γ_2 represent the direct effect of the paternal grandparents' literacy variables on child time use. Bounds derived from Conley et al. (2012)'s union of confidence intervals method. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.12*2SLS estimates of the effect of parental education on non-farm employment*

Variable	(1)	(2)
	Dependent Variables	
	Non-farm business (0/1)	Wage employment(0/1)
Panel A: Mother's education instrumented		
Mother's education	0.042*** (0.012)	0.027*** (0.009)
Multigenerational co-residence controls	✓	✓
Other Controls	✓	✓
Religion dummies	✓	✓
District FE	✓	✓
Year dummy	✓	✓
Montiel Olea & Pflueger <i>F</i> stat	56.12	57.01
Hansen <i>J</i> stat (p-value)	1.06 (0.30)	2.31 (0.13)
Observations	6,564	6,940
Panel B: Father's education instrumented		
Father's education	0.079*** (0.021)	0.081*** (0.022)
Multigenerational co-residence controls	✓	✓
Other Controls	✓	✓
Religion dummies	✓	✓
District FE	✓	✓
Year dummy	✓	✓
Montiel Olea & Pflueger <i>F</i> stat	21.79	22.02
Hansen <i>J</i> stat (p-value)	0.22 (0.64)	1.92 (0.17)
Observations	5,548	5,736

Notes: Standard errors are reported in parentheses and are clustered at the parent level. Wage employment is measured as an indicator variable for whether the mother worked as an employee for wages or salary in the past year in Panel A, while it is measured as a dummy variable taking the value one if the father's primary economic activity over the past 12 months was wage employment in Panel B. Other control variables include household size, wealth index, size of household cultivated land, number of female household members under age 6, number of male household members under age 6, religion, female headship status, and household distance from nearest road. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure 3.1: *Regional Prevalence of Child Labor*

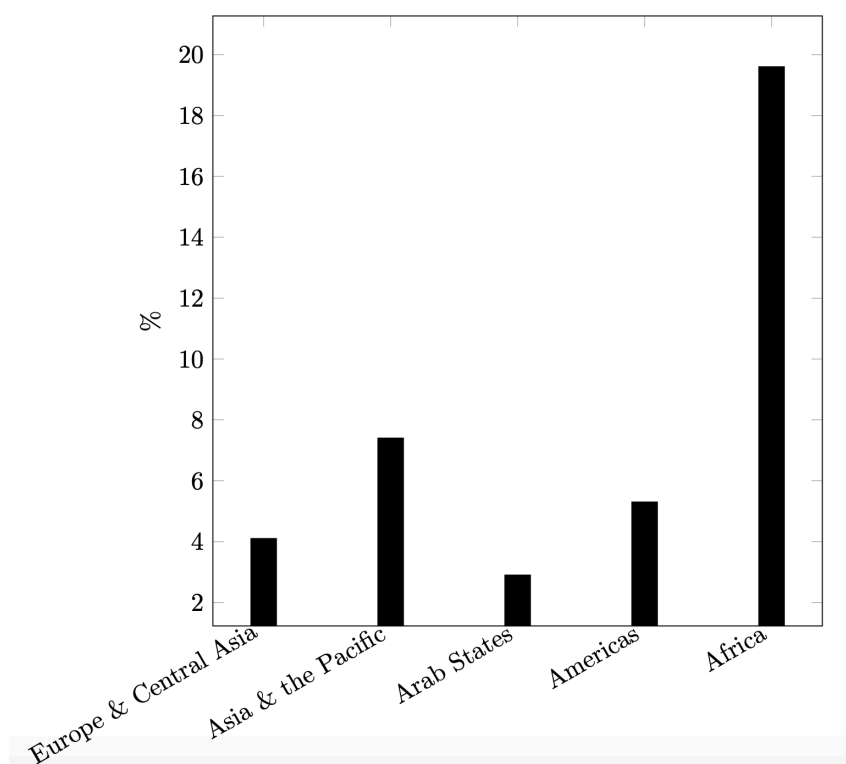
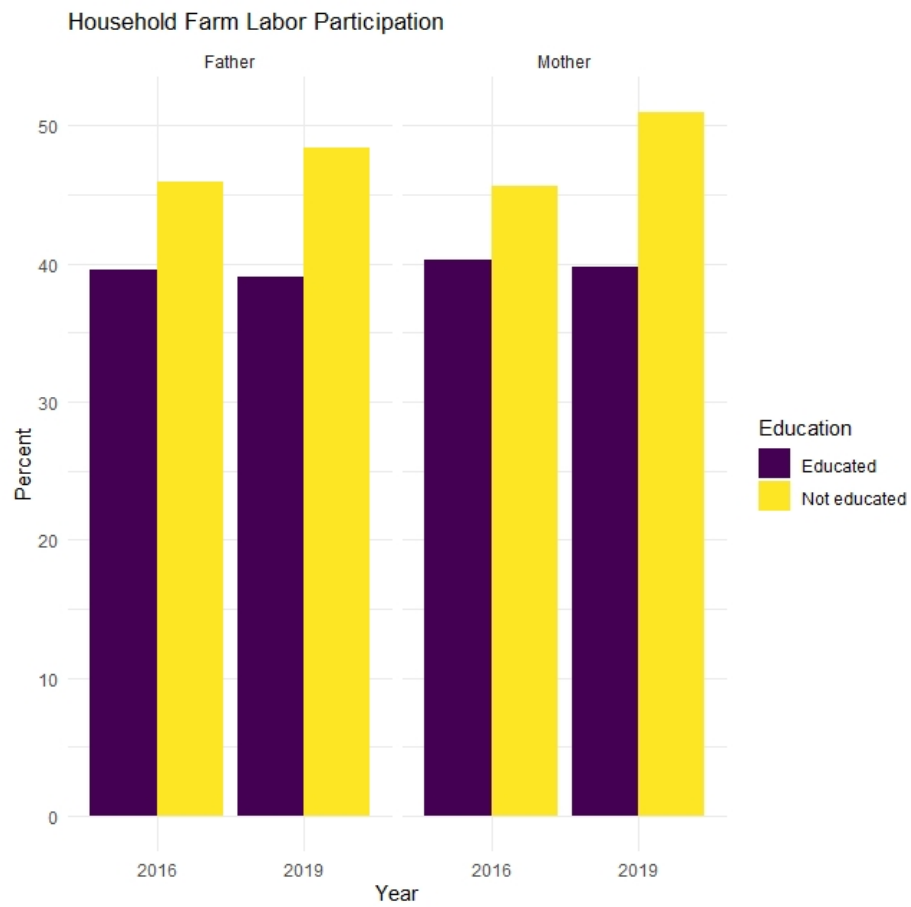
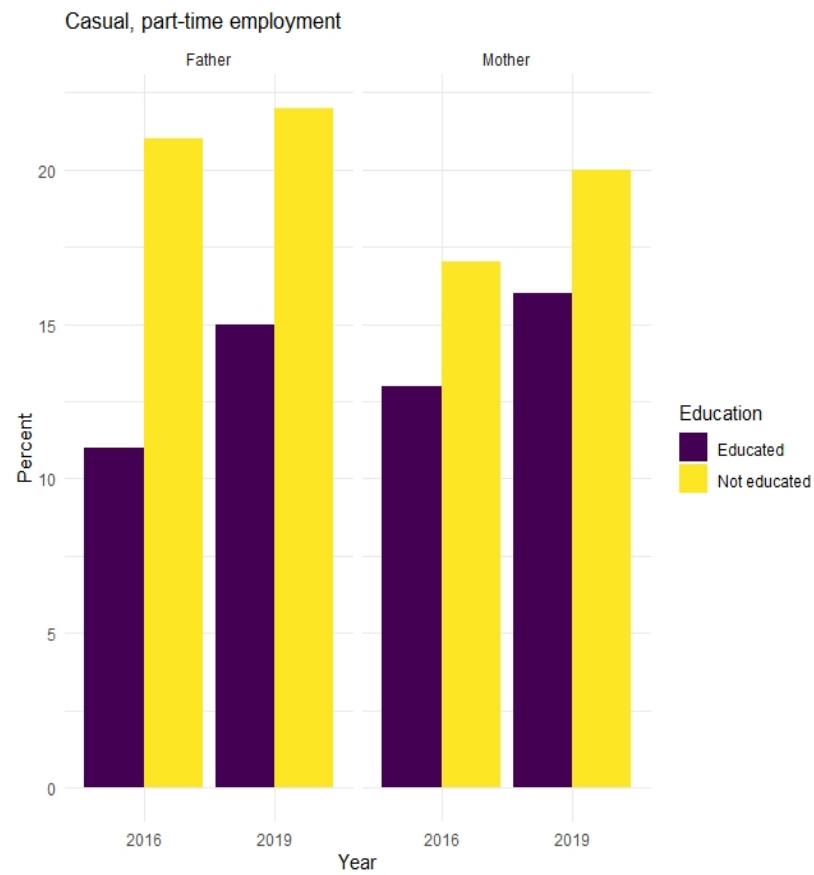


Figure 3.2: *Household Farm Labor Participation by parents' education status*



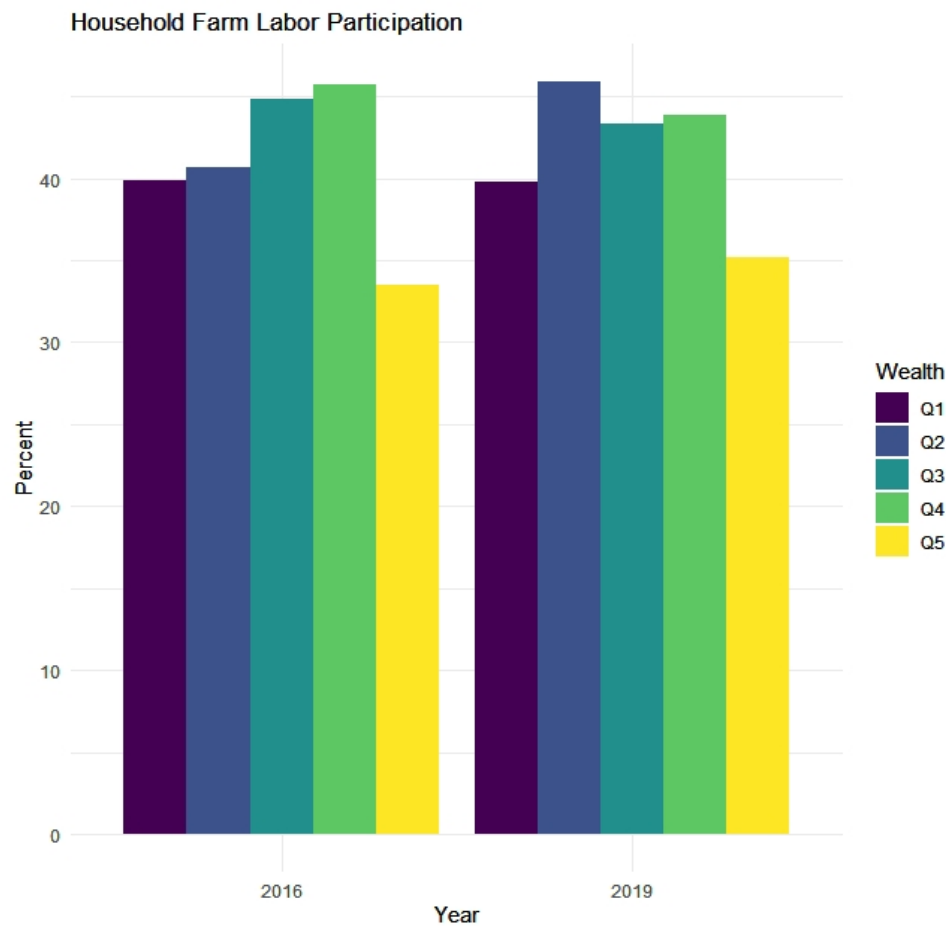
Notes: No education implies zero years of education. Observations are weighted using 2016 panel weights.

Figure 3.3: *Casual, part-time employment by parents' education status*



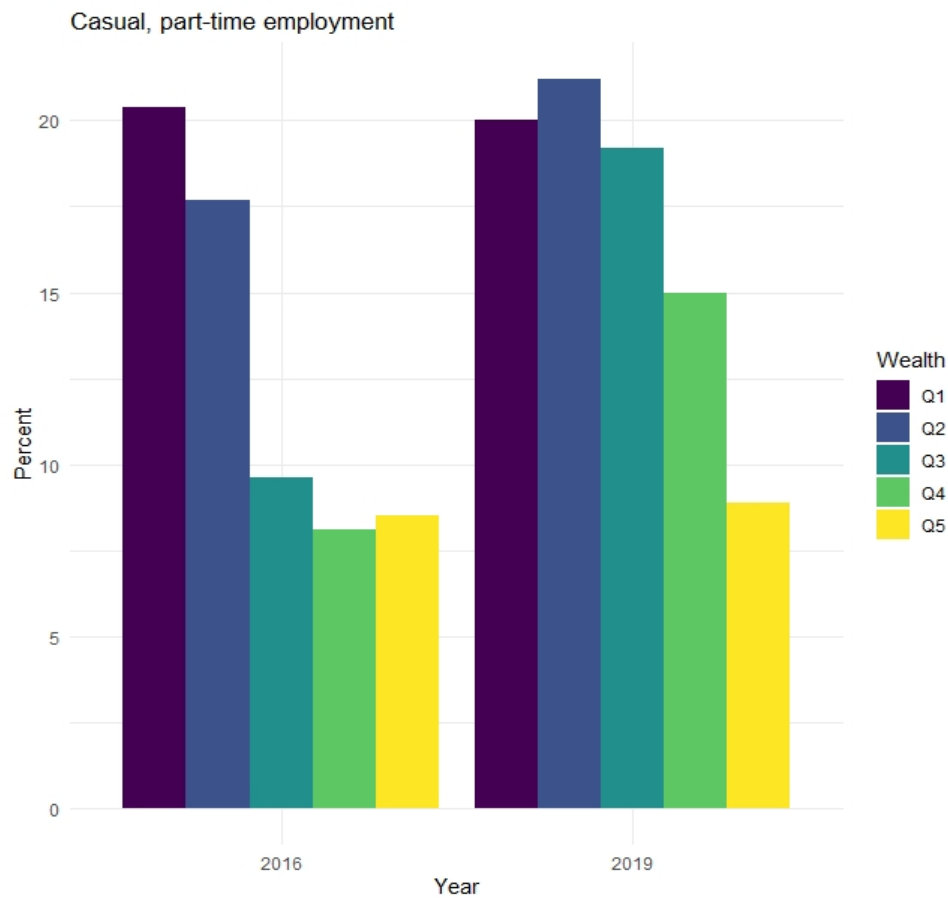
Notes: No education implies zero years of education. Observations are weighted using 2016 panel weights.

Figure 3.4: *Household farm labor participation by wealth quintiles*



Notes: Q1 denotes lowest wealth quintile. The wealth index is measured using household assets based on Principal Component Analysis. The wealth index variable was constructed using a principal component analysis where assets such as cars, motorcycles, bicycles, televisions, electric or gas stove, generators, washing machines, air conditioner, fan, radio, among others are given varying weights depending on the rarity of ownership among the sampled households. Observations are weighted using 2016 panel weights.

Figure 3.5: *Casual, part-time employment by wealth quintiles*



Notes: Q1 denotes lowest wealth quintile. The wealth index is measured using household assets based on Principal Component Analysis. The wealth index variable was constructed using a principal component analysis where assets such as cars, motorcycles, bicycles, televisions, electric or gas stove, generators, washing machines, air conditioner, fan, radio, among others are given varying weights depending on the rarity of ownership among the sampled households. Observations are weighted using 2016 panel weights.

Figure 3.6: *Effect of maternal education on household farm work - instrumented*

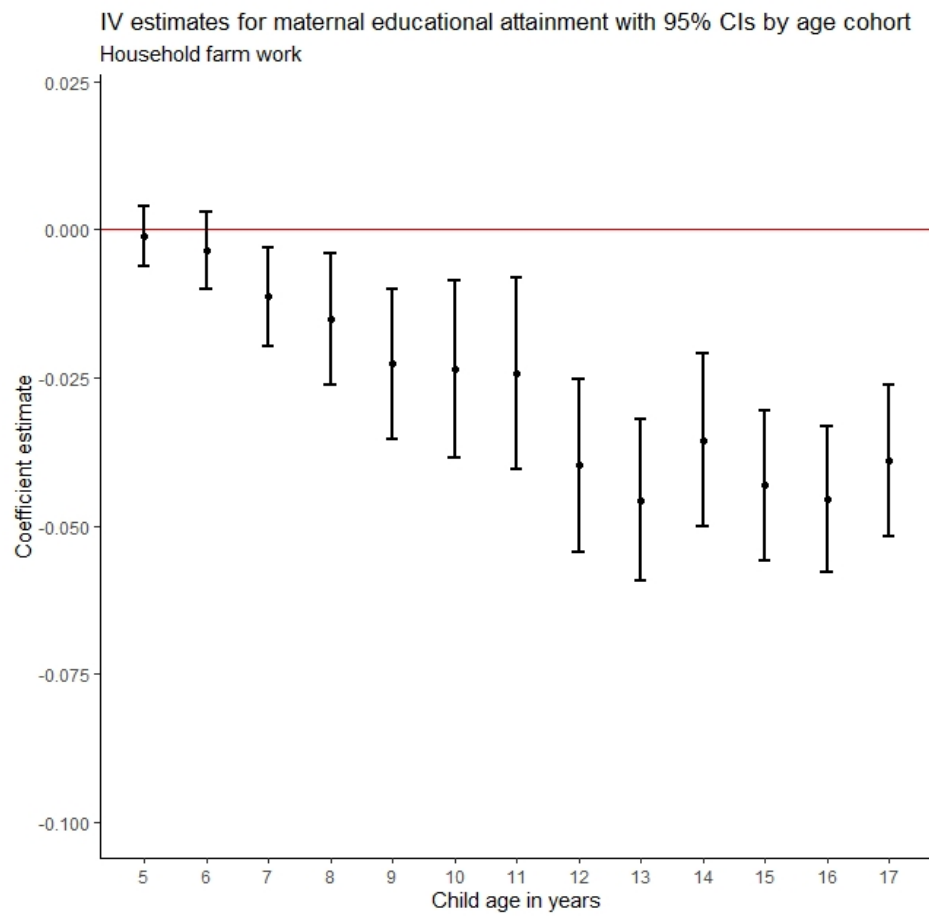


Figure 3.7: *Effect of maternal education on casual, part-time or “ganyu” labor employment – instrumented*

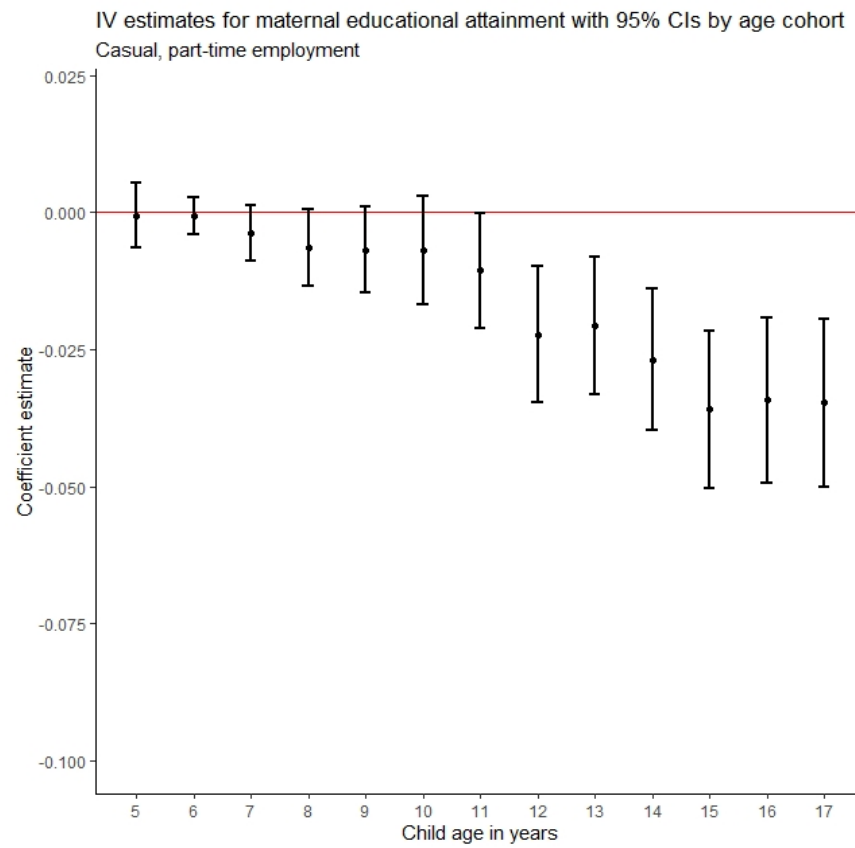


Figure 3.8: *Effect of paternal education on household farm work - instrumented*

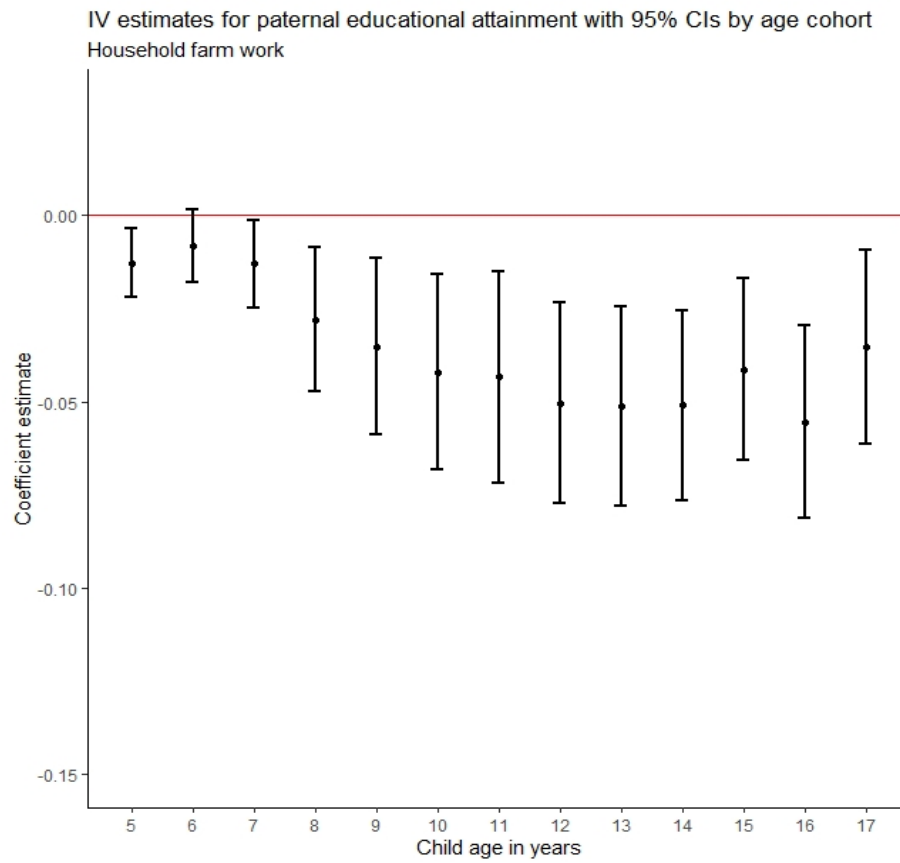
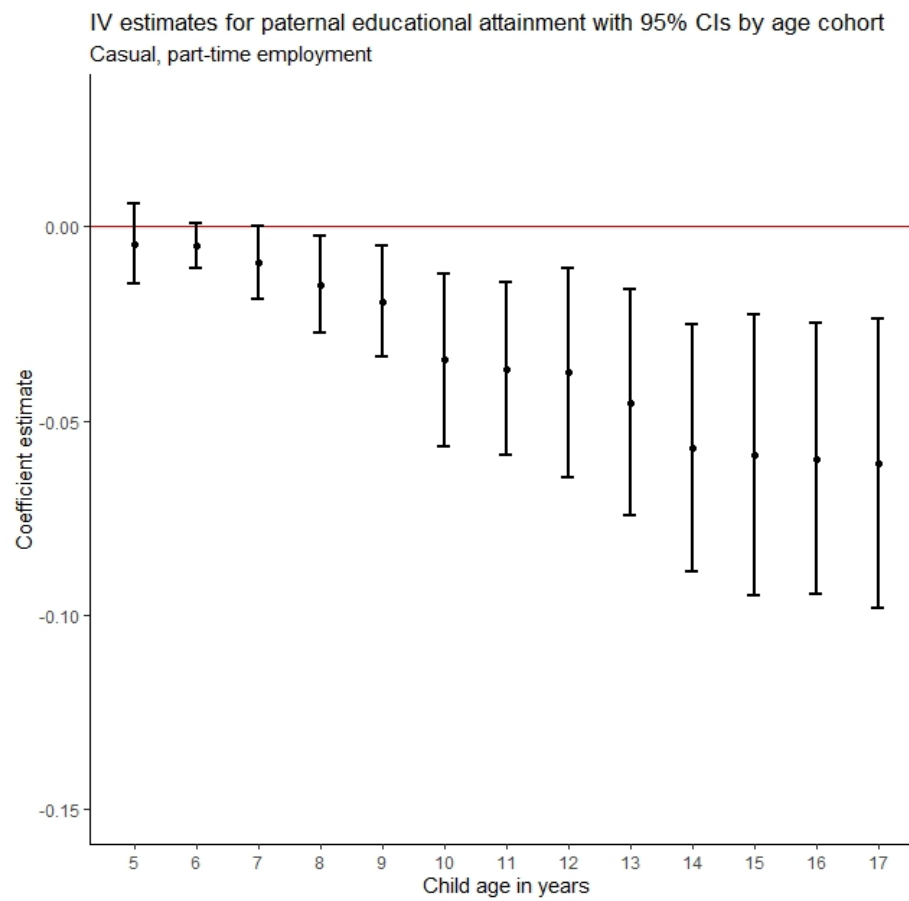


Figure 3.9: *Effect of paternal education on casual, part-time or “ganyu” labor - instrumented*



APPENDIX B: THEORETICAL MODEL

In this section, I present a simple conceptual framework to formally model child labor supply response to parental educational attainment. I use a version of the well-known Singh et al. (1986)'s agricultural household model wherein households are simultaneously involved in both consumption and production.³⁴ In principle, parental education can influence child labor participation through a variety of pathways. First, education can induce higher agricultural productivity (Hayami and Ruttan, 1970; Appleton and Balihuta, 1996; Reimers and Klasen, 2013). For instance, in settings with rapid or accelerating rate of technical change, educated farmers can capitalize on the availability of new technological innovations to expand production scale. In a cross-country study, Reimers and Klasen (2013) find a highly significant, and positive relationship between education and agricultural productivity using panel data on 95 countries. The authors show that this effect is robust to alternative specifications, data sets, and estimation strategies. The resulting rise in agricultural incomes due to positive agricultural productivity shifts can relax a household's liquidity constraint, inducing a reorientation of the child's time towards school.

Second, education drives up the economic returns from non-farm work (that is, either in wage or self-employment) which often requires skilled labor (Reardon et al., 1998). Reardon et al. (1998) revealed that education is a strong determinant of non-farm employment participation, projected to overtake landholdings as the major driver of non-farm income at least among rural households. Consequently, educated parents might find their skills better suited to non-farm activities especially in urban areas, where non-farm jobs are relatively plentiful. For children in such households, their parent's inter-sectoral mobility—that is, from on-farm to non-farm work—

³⁴ Ideally, and in keeping with the discussions above, we should employ a collective household model as the basis for the theoretical micro-foundations. However, for expositional clarity and given that the intent of the model is not to recover estimates for underlying parameters such as the intra-household bargaining weights, I rather use a unitary household model.

might reduce their involvement in household farm work altogether if children tend to work on farms side-by-side with their parents. The model presented here focuses on the latter pathway. This model borrows from by Reggio (2011), but it is adapted to account for how parental education shapes child labor participation decisions through non-farm employment.

Consider a household comprising of two agents: a parent (agent p), and a child (agent c). To model the household's child labor supply decision, we will assume that we are in a setting where only child, and adult (parent) family labor can be used in the production process, with the child serving as a source of on-farm labor.³⁵ The total time available to the parent is 1. We assume that the utility functions are twice continuously differentiable, strictly quasi-concave, and increasing in consumption, and leisure but decreasing in child labor. Another simplification of this model is that the decision-making family member (the parent) derives utility from aggregate household consumption, and we further assume that the cross partial derivative of the parent's utility function between consumption and child labor is non-negative, $u_{c_1,h}^1 \geq 0$. This assumption implies that the marginal utility of consumption is non-decreasing in child labor. That is, household consumption and exempting the child from work are not complementary (Reggio, 2011). Given that this assumption is a staple in the existing theoretical child labor literature, it seems at least standard to maintain (Dumas, 2020).

Further, we also assume that the child cannot sell her labor hours on the labor market, but the parent can. This assumption aligns with the type of child labor activities we consider in this study given that most child laborers serve as contributing household workers without pay. Besides,

³⁵ A significant contribution of this study is the investigation of parental educational attainment on different forms of child work including casual, part-time or "ganyu" labor. However, the theoretical model presented here focuses on the child's involvement in household farm work for the sake of brevity and tractability.

less than 0.3% of the children in the sample reported to have engaged in any full-time salaried or wage employment during the 2016 survey year.

The household solves the following optimization problem:

$$\max_{c_1, c_2, l_p, s_c, h, b_1} u^1(c_1, l_p, h) + \beta u^2(c_2) \quad (1)$$

$$s. t. \quad c_1 + \rho s_c = b_1 + p_a F \left[(1 - \gamma) \left((1 - l_p) + \lambda h \right) \right] + \gamma w_p (1 - l_p) \quad (2)$$

$$c_2 + R b_1 = \underline{w} + (\bar{w} - \underline{w}) s_c \quad (3)$$

$$h + s_c + l_c = 1 \quad (4)$$

where u^t denotes the parent's utility at time $t \in \{1, 2\}$ ³⁶ The parent's utility is a function of aggregate consumption c_t in each time period, and she derives utility from her own leisure (l_p), and disutility from child labor, h in period 1. In the first period, the household can allocate the child's total time endowment, 1 across schooling (s_c), work (h), or leisure (l_c). The parent, on the other hand, can either work or enjoy leisure in period 1, but does not work in period 2 (that is, $l_p = 1$ in the second period). The household's consumption expenditure in period 1 is met with farm income from agricultural production ($F(\cdot)$), borrowing (b_1), and non-farm income if the parent's probability of engaging in non-farm work, γ is non-zero. An important assumption underlying the period 1 budget constraint is that child labor complements adult family labor in household agricultural production. Hence, movement of the parent completely out of farm work drives the child's contribution to household farm work down to zero. Household expenditure in period 1 consists of direct consumption, c_1 and the cost of child education, ρ if she attends school; $F(\cdot)$ denotes the household's production function which we assume is increasing in total household labor but exhibits diminishing marginal returns (that is, $F'(\cdot) > 0$, and $F''(\cdot) < 0$); λ is the labor

³⁶ Notice that all household decisions rest with the parent.

productivity ratio between adult and child labor, and w_p is the prevailing non-farm employment wage rate; p_a is the price of the production good, and we normalize the price of the consumption good to 1.

In period 2, the child is now a working adult, and her income depends on her educational status in the first period. Hence, the child's income in period 2 is given by a base income, \underline{w} plus a schooling premium, $(\bar{w} - \underline{w})s_c$ depending on the amount of schooling she received in period 1. The household's consumption in period 2, c_2 is covered by the child's earnings as a working adult and any outstanding household debt incurred in period 1 must be paid-off at a gross interest rate, R . Consolidating the budget constraints for the two time periods, and the child time constraint yields the following inter-temporal budget constraint:

$$c_1 + \frac{c_2}{R} = p_a F[(1 - \gamma)((1 - l_m) + \lambda h)] + \frac{w}{R} - \left(\rho - \frac{\bar{w} - w}{R} \right) (1 - h - l_c) + \gamma w_p (1 - l_p) \quad (5)$$

which indicates that the household's total consumption (in period 1 units) across the two time periods must equal the total amount of resources available to them (also expressed in period 1 units). Given this consolidated inter-temporal budget constraint, we can solve the household's problem above for the optimal levels of consumption, leisure, schooling, and child labor. The generic functional form of our utility function precludes the derivation of closed form solutions for these choice variables, though we can show that these variables can be derived as functions of prices, wages, and the parameters, γ , and λ .

Next, we investigate the influence of parental education on child labor through non-farm employment. Assuming we have an interior solution, the first order necessary condition with respect to child labor yields:

$$u_h^1 = -\mu \left[p_a(1 - \gamma)\lambda F'(\cdot) + \left(\rho - \frac{\bar{w} - w}{R} \right) \right] \quad (6)$$

where μ denotes the Lagrange multiplier. The FOC above can be interpreted as the marginal utility or disutility from child labor for the parent must equal the marginal benefit of child work captured by the marginal revenue product of child labor from agricultural production, plus the difference between the direct cost of school attendance, and the discounted premium from having an educated child. Similarly, the first order condition with respect to period 1 aggregate consumption produces:

$$\mu = u_{c_1}^1 \quad (7)$$

Substituting (7) into (6), it follows that,

$$-\frac{u_h^1}{u_{c_1}^1} = \left[p_a(1 - \gamma)\lambda F'(\cdot) + \left(\rho - \frac{\bar{w} - w}{R} \right) \right] \quad (8)$$

which represents the trade-off between child labor and household consumption.

To derive the effect of parental education on child labor, we can rewrite (8) as follows:

$$G = u_h^1 + u_{c_1}^1 \left[p_a(1 - \gamma)\lambda F'(\cdot) + \left(\rho - \frac{\bar{w} - w}{R} \right) \right] = 0 \quad (9)$$

Recall that—as one of our assumptions stipulates—parental education affects child labor by driving up the likelihood of non-farm employment. Therefore, by implicitly differentiating (9) with respect to the probability of non-farm work participation, γ , we have:

$$\frac{dh}{d\gamma} = -\frac{\frac{\delta G}{\delta \gamma}}{\frac{\delta G}{\delta h}} \quad (10)$$

$$= -\frac{-\lambda p_a u_{c_1}^1 \left[F'(\cdot) - \lambda(1 - \gamma) \left((1 - l_p^*) + \lambda h^* \right) F''(\cdot) \right]}{u_{h,h}^1 + u_{c_1,h}^1 \left[p_a(1 - \gamma)\lambda F'(\cdot) + \rho - \left(\frac{\bar{w} - w}{R} \right) \right] + p_a u_{c_1}^1 \lambda^2 (1 - \gamma)^2 F''(\cdot)} \quad (11)$$

where $u_{c_1}^1$ denotes the parent's marginal utility of period 1 consumption. Notice that given our assumptions above, the numerator is unambiguously negative (that is, since $F'(\cdot) > 0$ and $F''(\cdot) < 0$). By contrast, the sign of the denominator is ambiguous and will depend on the relative magnitudes of the immediate benefits of not enrolling the child in school and the discounted premium from having an educated child working in period 2, as well as the restriction on the cross partial derivative between consumption and child labor. This result can be expressed mathematically as follows:³⁷

$$\text{sign} \left[\frac{dh}{d\gamma} \right] = \begin{cases} - & \text{if } u_{c_1,h}^1 = 0 \\ - & \text{if } u_{c_1,h}^1 > 0 \text{ and } p_a(1-\gamma)\lambda F'(\cdot) + \rho \leq \left(\frac{\bar{w} - w}{R} \right) \\ +/- & \text{if } u_{c_1,h}^1 > 0 \text{ and } p_a(1-\gamma)\lambda F'(\cdot) + \rho > \left(\frac{\bar{w} - w}{R} \right) \end{cases} \quad (12)$$

In sum, our theoretical model can be distilled into three alternative predictions. First, an increase in the probability of non-farm employment induced by higher parental education reduces child labor if the cross partial derivative between consumption and child labor, $u_{c_1,h}^1$ is zero. Specifically, if a marginal increase (or decrease) in child labor has no effect on the marginal utility of consumption for the parent, child labor should fall with an increase in γ . Because an increase in the likelihood of the parent's non-farm engagement diminishes the child's contribution to household agricultural production, we expect child labor to diminish given that the parent derives disutility from child labor and the marginal utility of consumption remains unaffected.

³⁷ Similarly, we can express the comparative statics in terms of the effect on child labor due to a change in parental education as follows:

$$\frac{dh}{dE} = \frac{dh}{d\gamma} \cdot \frac{d\gamma}{dE}$$

$(-/+)$ $(-/+)$ $(+)$

, where E represents parental schooling.

Second, if we assume that the marginal utility of consumption increases with child labor, then for a marginally higher propensity of parental non-farm employment, a reduction in child labor incidence is yet optimal if the future gain of having an educated child at least exceeds the immediate benefit of child work. Put differently, given that the parent's valuation of aggregate consumption in period 2 offsets the benefit of additional child effort in agricultural production in period 1, child labor will still decline. On the other hand, under the assumption that the marginal utility of consumption with respect to child labor is positive, and the immediate benefit of child work is relatively higher, then the effect of a change in γ on child labor will depend on the magnitudes of the second order effects of child labor on the parent's utility and household agricultural production. The overall effect on child labor is therefore ambiguous in this case.