The University Of Sheffield.

Market- and household-level effects of economic shocks in Malawi

By

Stevier Kaiyatsa

A thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

The University of Sheffield
Faculty of Social Sciences
Department of Economics

September 2024

Abstract

This thesis has three empirical studies that are motivated by the impacts of economic shocks on food prices, household well-being, and schooling outcomes in Malawi.

The first empirical study systematically examines how transport costs are associated with spatial or regional inequalities in affordability of various foods across markets. The analysis reveals that the endogenous increase in transport costs is associated with a reduction in overall spatial inequalities in affordability of various food across markets in the short run, on average. This counterintuitive influence is driven by processed foods but not by perishable foods and nutrient-dense foods for which the changes in transport costs are associated with an increase in overall price differences across markets in the short run. Examining the relationship between transport costs and price differentials for each food item, we find that spatial inequality in food affordability widens for maize flour dehulled maize grain, maize grain (private), maize grain (ADMARC), brown beans, and eggs in the short run. Thus, an increase in transport costs lowers the incentive of traders to move these food items from a lower price market to a higher price one, sustaining spatial inequalities in prices across markets. Overall, the magnitudes of the relationships between transport costs and price differentials are smaller for market pairs that are closer to each other for all foods under consideration. In addition, we find that spatial inequality in food affordability widens for most foods under investigation, except for brown beans, usipa, utaka, and tomatoes in the long run. Therefore, there are both food security and nutritional implications of increases in transport costs, hence, the need to promote food affordability and nutrition.

The second empirical study investigates the extent to which the reform to Malawi's fuel policy adopted in 2012 increased or decreased agricultural production and consumption differentially among households that are net-sellers, net-buyers or self-sufficient in staple maize grain. Results show that households that are in autarky in remote areas increased maize production more than those closer to the market but had lower consumption due to the increase in transport costs of accessing markets. Households that are net buyers that reside closer to the market increased maize production, consumption, and became less prone to maize insecurity, while those that reside in remote locations had lower non-food consumption and became more prone to maize insecurity relative to households that are in autarky. Conversely, households that are net sellers that reside in remote locations had lower non-food

consumption and maize consumption, while those that reside closer to the market had a reduction in consumption, non-food consumption and non-maize food consumption relative to households that are in autarky. These findings have implications for other countries that are considering rescaling or removing fuel subsidies on household welfare.

The final empirical study examines the extent to which rainfall shocks differentially affect schooling outcomes in both primary and secondary education among boys and girls. The analysis shows that households allocate more resources to boys during the periods of flood shock, while resource allocation among girls is similar during the period of the rainfall shock and the normal rainfall. Turning to school attendance, the analysis reveals that the drought shock increases school attendance among younger boys and girls in lower primary school, but it reduces school attendance among older boys and girls in secondary school relative to the normal rainfall. Conversely, the flood shock increases school attendance among older boys and girls in upper primary and secondary school relative to a normal rainfall. Moving on to school progression, we find that the drought shock increases school progression among boys and younger girls in lower primary school, while the flood shock increases school progression among older boys in upper primary school and younger boys in lower secondary school, and among girls in secondary school relative to the normal rainfall. These findings have implications for education policy in other countries that are aiming at eliminating gender inequality in schooling.

Dedication

This thesis is dedicated to my wife, Tendai, for your love, support, endurance and for being actively involved in raising Lily and Liam when I was studying.

Acknowledgements

This thesis has been made possible by the generous support and persistent guidance from my supervisors, Bhavani Shankar, Nicolas Van de Sijpe, and Karl Taylor. Thank you so much for your invaluable academic and mental support, and various contributions throughout my time at TUoS. I would also like to acknowledge academic and support staff at the Department of Economics for their assistance and support.

In addition, I would like to thank my PhD examiners, Steven McIntosh and Arjan Verschoor, for their constructive and helpful comments, which have considerably improved my thesis.

My sincere gratitude goes to the Commonwealth Scholarship Commission in the United Kingdom for the award of the doctoral scholarship to undertake this research at TUoS.

I acknowledge Mr. Emmanuel Tsoka from the Office of National Statistical Office, and Mr. Maclean Kaluwa from the Ministry of Education in Malawi for your assistance with food prices and school census data, respectively. Thank you so much Kate Schneider for sharing your software code for the household consumption module of the Living Standards Measurement Study survey for Malawi.

To my colleagues, Lucy Ward, Thomas Siddall, Tingli Shen, and unmentioned colleagues, thanks for the support and encouragement. To my friends, Gregory Cooper, Morris Oleng and Takondwa Chauma, thanks for the encouragement and making this journey an enjoyable experience.

Table of contents

List of tables

List of figures

Chapter 1

Introduction

1.1 Motivation and aims

Malawi's food system is changing as farming households are increasingly becoming dependent on markets to access food, farm inputs, and better output prices. However, Malawi is one of the landlocked countries in Sub-Saharan Africa (SSA), which heavily relies on road transportation for its domestic trade and largely depends on rain-fed agriculture. The road transportation is heavily dependent on imported liquid fuels (Kaunda, 2016; Lall et al., 2009; Robinson & Wakeford, 2013) and has been significantly affected by fuel price shocks leading to higher transport costs. Due to higher transport costs, either food markets do not deliver various foods across space efficiently or farming households are unable to access inputs and better output prices, which lead to spatial inequalities in affordability of foods across markets, and reduce agricultural productivity and household welfare, respectively (see, for example, Aggarwal, 2018; Aker, 2010a; Christiaensen & Demery, 2018; Filmer et al., 2021; Headey et al., 2019; Jones, 2017; Nkegbe & Abdul Mumin, 2022; Olabisi et al., 2021; Stifel & Minten, 2017). Conversely, rain-fed agriculture has recently become more vulnerable to rainfall shocks such as droughts and floods (Benson & Weerdt, 2023; Chirwa, 2006; Government of Malawi, 2016a). Rainfall shocks reduce agricultural revenue and household consumption, which may affect schooling outcomes as households strive to cope with the negative effects (Aguilar & Vicarelli, 2022; Baez et al., 2017; Baquie & Fuje, 2020; McCarthy et al., 2017; McLaughlin et al., 2023; Rosales-Rueda, 2018). Therefore, this thesis is motivated by the effects of economic shocks arising from the change in fuel prices on functionality of food markets and household well-being, and rainfall shocks on schooling outcomes in SSA using an empirical application for Malawi.

The first empirical chapter aims at examining how transport costs are associated with spatial or regional inequalities in affordability of various foods across markets in Malawi. There is little research that systematically look at the implications of fuel price changes on food affordability and nutrition (Abbott et al., 2008; Dillon & Barrett, 2016; Fuje, 2019; Roehner, 1996; Volpe et al., 2013). Thus, transport costs as a driver of regional variation in food affordability and nutrition is underexplored. The reason is that availability of data on transport rates or charges across various routes is limited in most developing countries (Roehner, 1996; Salazar et al., 2019). Therefore, we follow Storeygard (2016) to systematically examine the relationship between fuel price changes and price dispersion of various foods across markets in Malawi. Further, most of the previous work on price dispersion of agricultural commodities has focused on most storable grains that are either semi-perishable (*e.g.*, cowpea, beans, and groundnuts) or less perishable (*e.g.*, millet, maize, rice, and sorghum) in developing countries. A further contribution we make to the literature is to examine how transport costs are associated with price dispersion of processed foods, more perishable foods (*e.g*., eggs, meat, fresh fruits, and vegetables), or nutrient-dense foods across markets in a developing country setting. Nutrient-dense foods are foods with high concentration of essential vitamins and minerals that are important for nutrition and long-term health (Darmon & Drewnowski, 2015; Dittus et al., 1995; Drewnowski et al., 2004; Drewnowski & Specter, 2004; Rideout et al., 2015). Although it is not always the case, more nutritious foods are often more perishable, have higher value to weight ratios, and are difficult to aggregate in large volumes but require a shorter time for traders to get them into the market.

The second empirical study investigates the extent to which the reform to Malawi's fuel policy adopted in 2012 increased or decreased agricultural production and consumption differentially among households that are net-sellers, net-buyers or self-sufficient in staple maize grain. The Government re-introduced automatic price adjustment mechanism in 2012 that was abandoned in 2004 to sustain fuel supply, which led to an increase in fuel prices. Prior to the fuel price reform, liquid fuels were being sold at subsidised prices and events of fuel shortages were common across the country. This chapter adds an important dimension to the literature on transport costs in SSA by estimating how the fuel subsidy removal differentially affected staple maize production and consumption. There are only a limited number of studies that investigate the effects of transport costs on production and consumption of farm produce by households in developing countries using observational cross-sectional data via econometric estimation (see, for example, Damania et al., 2016; Fuje, 2019; Minten et al., 2013; Omamo, 1998; Vandercasteelen et al., 2018). The closest study to ours is Fuje (2019) that examines the impact of fuel subsidy reforms on real incomes for net buyers and net sellers of grains (*i.e*., "teff", wheat, maize, sorghum, and barley) via a non-parametric approach in Ethiopia. Building on Fuje (2019) and the previous studies, ours is the first study to estimate the differential impacts of the fuel policy reform on households using nationally representative observational panel data from SSA via a parametric approach.

The third empirical chapter investigates how rainfall shocks differentially affect schooling outcomes in both primary and secondary education among boys and girls in Malawi. There is a small but growing literature that investigates the effects of weather shocks or changing climate on child schooling outcomes in developing countries. However, empirical evidence on how weather shocks affect child schooling outcomes in SSA is limited (Björkman-Nyqvist, 2013; Randell & Gray, 2016). We add to this literature by examining how rainfall shocks differentially affect schooling outcomes in both primary and secondary schools among boys and girls in Malawi. Björkman-Nyqvist (2013) is the closest study to ours, examining gender differential effects of weather shocks on child schooling outcomes. However, the aforementioned study used pooled cross-sectional household survey data and administrative primary school census data aggregated at the district level. We build on Björkman-Nyqvist (2013) and the previous studies to examine how rainfall shocks differentially affect schooling outcomes in both primary and secondary education among boys and girls using household panel survey data and administrative school-level census data for Malawi.

1.2 Overview of the thesis

1.2.1 Overview of chapter 2

This second chapter systematically examines how transport costs are associated with spatial or regional inequalities in affordability of various foods across markets in Malawi. Given that not all foods are produced within locations markets operate, market traders transport food from areas of high production (*i.e.,* surplus locations*)* to areas of low production (*i.e.,* deficit locations*)*, which involves transport costs. However, the larger share of marketing costs that market traders incur is transport costs (Fafchamps & Gabre-Madhin, 2006), which impedes market traders to transport food items from markets in surplus areas to markets in deficit areas. This increases price dispersion of various foods across markets, which increases food prices and reduces food affordability in markets located in deficit areas. Therefore, we anticipate transport costs to have positive associations with price dispersion of various foods across markets, but the magnitude of the relationship is smaller for market pairs that are closer to each other than market pairs that are farther from each other.

We use monthly consumer price monitoring data that the Malawi National Statistical Office collects to compute the Consumer Price Index. Further, we obtained monthly average diesel pump prices from the Malawi Energy Regulatory Authority and the route distance over paved roads between the market pairs from Google Maps. In combination with other data from various sources, we first estimate how the changes in transport costs are associated with overall price dispersion of various foods across markets using the panel non-linear dyadic regression model via the Poisson pseudo-likelihood regression estimator with multiple levels of fixed effects in the short run. Then, we investigate whether processed foods, perishable foods, or nutrient-dense foods modify the relationship between transport costs and overall price dispersion of various foods across markets in the short run. Finally, we estimate separately the panel non-linear dyadic regression model via the Poisson pseudo-likelihood regression estimator with multiple levels of fixed effects, and the instrumental variable Poisson (ivpoisson) estimator for each food item to examine the association between transport costs and price dispersion across markets in the short and long runs, respectively.

Contrary to our expectation, the results from our analysis show that the endogenous increase in transport costs is associated with the reduction in overall price dispersion of various foods under investigation across markets in the short run, on average. We find that this counterintuitive influence is driven by processed foods but not by perishable foods and nutrient-dense foods for which the changes in transport costs are associated with an increase in overall price differences across markets in the short run. Moving on to separate results for each food item, we find that the association between the changes in transport costs and price differences remain negative and significant for rice grain, beef, goat meat, powdered milk, and bananas but become positive and significant for maize flour dehulled, maize grain (private), maize grain (ADMARC), brown beans, and eggs providing additional evidence that the counterintuitive influence of the changes in transport costs on overall price differentials are driven by food items that have lower search costs and are easy to aggregate in the short run. Turning to spatial heterogeneity of the association between transport costs and price differences, we find that the magnitudes of the relationships are smaller for market pairs that are closer to each other for all foods under consideration. Moving on to the influence of transport costs on price differences for each food item in the long run, we find that the price differences in the previous period play a small role in predicting the price differences in the current period, and the increase in transport costs is associated with the increase in price differentials for most foods under investigation, except for brown beans, usipa, utaka, and tomatoes.

1.2.2 Overview of chapter 3

This third chapter builds on the previous chapter to investigate the extent to which the reform to Malawi's fuel policy adopted in 2012 led to either an increase or a decrease agricultural production and consumption of staple maize grain differentially among households. The fuel price reform increased the cost of transporting produce from the farm to the market or consumption centre, and inputs from the market to the farm using motorised transportation. We hypothesise that the policy reform has immediate differential effects on staple maize production and consumption, but the differential effects do not persist over time once the policy is adopted. Thus, we expect the differential effects of the policy reform to deplete from its initial impact as households cope to dampen off its effects over time. Consistent with the previous literature, we anticipate a heterogeneous differential impact of the fuel price reform on households that vary with household status (*i.e*., net-seller, net-buyer or self-sufficient in staple maize grain), and market access. Net sellers are households whose quantity of staple maize grain sold on the market is greater than the quantity of staple maize grain purchased, net buyers are households whose quantity of staple maize grain purchased on the market is greater than the quantity of staple maize grain sold, and self-sufficient households are those that do not purchase or sell any staple maize grain on the market. We expect the policy reform to have a larger effect on net sellers and net buyers relative to self-sufficient households of staple maize grain that varies with the level of market access.

We use three waves of nationally representative panel data from the Integrated Household Panel Survey (IHPS), which were implemented in 2010, 2013, and 2016 as part of the Living Standards Measurement Study survey (LSMS-ISA) for Malawi. The policy reform in 2012 provides us with a natural experimental setting to conduct our analysis, where the 2010 data represents the period before the reform while the 2013 and 2016 data represent the period after the reform. We explicitly examine whether any pre-existing differences on staple maize production and consumption across household groups persisted after the policy reform. Any break in pre-existing differences in the level or trend of staple maize production and consumption closer to the time of the reform in 2012 and then further away from the reform period is our estimate of the causal impact of the policy reform (Finkelstein, 2007; Sun & Shapiro, 2022). We use the data between 2010 and 2013 (*i.e.*, one year after the policy reform) to estimate immediate differential effects of the policy reform, and the data between 2010 and 2016 (*i.e.*, four years after the policy reform) to estimate persistent differential effects on maize production and consumption among households using a fixed effects estimator.

Results confirm that there are heterogeneous differential impacts of the fuel price reform on households that vary with household status and market access. We find that there are both short- and long-term consequences of the fuel policy reform on staple maize production and consumption. Thus, households do not dampen off the effects of increasing transport costs in the long run. Overall, our results indicate that households that are in autarky in remote areas increased maize production more than those closer to the market, but their consumption fell arising from the increase in transport costs of accessing markets. Households that are net buyers that reside closer to the market increased maize production, consumption, and became less prone to maize insecurity, while those that reside in remote locations had lower non-food consumption and became more prone to maize insecurity relative to households that are in autarky. Conversely, households that are net sellers that reside in remote locations saw a reduction in non-food consumption and maize consumption, while those that reside closer to the market lost in consumption, non-food consumption and non-maize food consumption relative to households that are in autarky. These differential effects are less sensitive to how market access is measured.

1.2.3 Overview of chapter 4

This fourth chapter investigates the extent to which rainfall shocks differentially affect schooling outcomes in both primary and secondary education among boys and girls in Malawi. Rainfall shocks affect household income and consumption through its effects on agricultural production. However, households may treat boys and girls differently to respond to the negative effects of rainfall shocks, which may lead to differential effects on schooling outcomes among boys and girls. Consistent with the previous literature, we anticipate the effects of drought and flood shocks on schooling outcomes to vary by child age. Thus, we anticipate older girls to be affected by drought and flood shocks, while boys and young girls to be insulated from the effects of drought and flood shocks.

Our measures of a rainfall shock is a binary drought indicator that takes on a value of one if the negative standardised deviation of rainfall from historical mean precipitation in the community is equal to or less than negative one, and zero otherwise, and a flood binary indicator that takes on a value of one if the positive standardised deviation of rainfall from historical mean precipitation in the community is equal to or greater than positive one, and zero otherwise. We use both household level panel data (2010, 2013, and 2016) from IHPS and the school census administrative data (2010 – 2016) from the Ministry of Education in Malawi and apply the fixed effects estimator separately for boys and girls.

The analysis reveals that there is differential treatment in children's education whereby households allocate more resources in boys' education during the periods of the flood shock, while resource allocation in girls' education is similar during the periods of the rainfall shock and the normal rainfall. As we expected, we find that the effects of the rainfall shock on school attendance and progression vary with child age. However, the effects of rainfall shocks on school attendance are similar between boys and girls, while the effects on school progression are different among boys and girls. For example, we find that the drought shock increases school attendance among younger boys and girls in lower primary school, but it reduces school attendance among older boys and girls in secondary school relative to the normal rainfall. Conversely, the flood shock increases school attendance among older boys and girls in upper primary and secondary school relative to a normal rainfall. Moving on to school progression, we find that the drought shock increases school progression among boys and younger girls in lower primary school, while the flood shock increases school progression among older boys in upper primary school and younger boys in lower secondary school, and among girls in secondary school relative to the normal rainfall. Overall, these findings are consistent at the school level.

1.3 Organisation of the thesis

This thesis has three chapters that have been written as standalone articles to facilitate Journal submission. The remainder of this thesis is organised as follows.

The second chapter systematically examines how transport costs influence spatial or regional inequalities in affordability of various foods across markets in Malawi. The chapter begins by reviewing the literature related to the role of markets in the food system, market integration, price differentials and transport costs. This is followed by the overview of food marketing in Malawi. The following section reviews various measure of price dispersion and findings from the previous empirical studies from developing countries. The proceeding section describes the conceptual framework, empirical strategy to establish the association between the changes in fuel costs with distance and price differentials, and data used in this chapter. Then, we present the main findings from the analysis and discuss results from several robustness checks where possible. The final section of this chapter concludes and provides policy implications of the changes in fuel costs on market integration in developing countries.

The third chapter investigates the extent to which the reform to Malawi's fuel policy adopted in 2012 increase or decrease agricultural production and consumption of staple maize grain differentially among households. The chapter starts by introducing the literature related to market participation, household well-being, and transport costs. The following section presents an overview of agriculture in Malawi. This is followed by an overview of fuel pricing in Malawi. The following section discusses the empirical method that is adopted in this chapter to examine distributional effects of fuel policy reforms on household welfare, and reviews findings from the previous empirical studies from developing countries. The proceeding section describes the theoretical framework, empirical strategy to establish the causal effect of the fuel policy reform on household welfare, data, classification of household groups, and construction of welfare-related indicators used in this chapter. Then, we present the main findings from the analysis along with the results from robustness tests using alternative measures of market access. The final section of this chapter concludes and provides policy implications of the effects of removing fuel subsidies on household welfare in developing countries.

The fourth chapter investigates the extent to which rainfall shocks differentially affect schooling outcomes in both primary and secondary education among boys and girls in Malawi. The chapter begins by introducing the literature related to inequality in schooling among boys and girls, and effects of rainfall shocks on households. The next section presents an overview of schooling in Malawi. This is followed by an overview of rainfall shocks in Malawi. The proceeding section reviews approaches that are often used in the literature to identify locations affected by rainfall shocks, empirical methods used to identify a causal relationship between the weather shock and outcome of interest, and findings from some of the previous studies from developing countries. The next section describes the conceptual framework, empirical strategy to establish the causal effect of rainfall shocks on schooling outcomes, data, and construction of schooling-related indicators used in this chapter. Then, we present the main findings from the analysis along with the results from robustness tests using alternative measures of rainfall shocks. The final section of this chapter concludes and provides policy implications of the effects of rainfall shocks on schooling outcomes in developing countries.

The firth chapter provides a conclusion to this thesis. The chapter starts by providing the summary of the main findings along with policy implications for the three empirical chapters. Further, the chapter discuss the limitations of the analyses in some of the empirical chapters that might provide opportunities for future research.

Chapter 2

How do transport costs influence price dispersion of various foods across markets in Malawi.

2.1 Introduction

Food markets play a critical role in improving access to food even amongst farm households who partially produce their own food in rural areas (Food and Agriculture Organization of the United Nations, 2016; Headey et al., 2019; Jones, 2017; Matita et al., 2021; Nandi et al., 2021; Rideout et al., 2015; Zanello et al., 2019). However, food markets in rural areas are often poorly equipped, inefficient, incomplete, and disjointed (de Janvry et al., 1991; Filmer et al., 2021; Headey et al., 2019; Hoddinott et al., 2015; Sibhatu & Qaim, 2017). Evidence shows that food markets are fragmented because of higher transport costs between them, which increases price differentials of food commodities (Aggarwal, 2018; Aker, 2010a; Atkin & Donaldson, 2015; Filmer et al., 2021; Jensen, 2007; Mu & van de Walle, 2011; Roehner, 1996). According to Roehner (1996), it becomes easier for spatially separated markets to engage in trade as transport costs decline, which lead to price convergence and better market integration. Lack of price convergence across markets increases costs of food and reduces the affordability of healthy diets when market purchase is an important mechanism to access certain food items in deficit locations (Filmer et al., 2021; Jensen, 2007; Roehner, 1996; Zant, 2018). For instance, Filmer et al. (2021) find that the Philippines' cash transfer raised local demand for food such as eggs and fish among program beneficiaries, which raised food prices and worsened nutrition among non-recipients in more remote locations where markets are disjointed. Thus, transport costs impede the ability of markets to deliver various foods across space efficiently leading to spatial or regional inequalities in affordability of various foods across markets.

The objective of this chapter is to systematically examine how the changes in transport costs influence spatial or regional inequalities in affordability of various foods across markets in Malawi. Although there is a large literature on food market integration, the extent to which transport costs affect price dispersion of food commodities across markets has received little attention in the literature (Abbott et al., 2008; Dillon & Barrett, 2016; Fuje, 2019; Roehner, 1996; Volpe et al., 2013). According to Roehner (1996) and Salazar et al. (2019), availability of data on transport rates or charges across various routes is limited in most developing countries. If this data is available, it will be incomplete with a shorter frequency for few market pairs (Salazar et al., 2019). As a result, most studies use fuel costs (Jensen, 2007), distance (see, for example, Aker et al., 2014; Roehner, 1996), proximity to a railway line (Zant, 2018), and removal of a fuel subsidy (Fuje, 2019, 2020) as direct proxies for transport costs. Fewer studies have used transport costs data across market pairs when examining how information communication technology and climatic shocks affect price dispersion of grains across markets in developing countries (Aker, 2010a, 2010b; Salazar et al., 2019). These studies find that higher transport costs increase price dispersion of storable grains across markets. Thus, there is limited empirical evidence on how transport costs affect price dispersion of various foods across markets in developing countries where transport costs are estimated to be high (Amjadi & Yeats, 1995; Limão & Venables, 2001; Rizet & Gwet, 1998; Rizet & Hine, 1993; Teravaninthorn & Raballand, 2009). To fill this knowledge gap, we systematically examine whether the changes in fuel price are associated with price dispersion of various foods across markets in Malawi using a novel measure of transport costs as in Storeygard (2016). The variation in fuel price in Malawi is potentially endogenous because the government may decide whether and when to pass on changes in the global fuel price to local diesel prices to avoid sudden increases in food prices thereby introducing a potential reverse causality. Thus, the government of Malawi sets local fuel prices, which are uniform across the country. We hypothesise that transport costs are positively associated with price dispersion of various foods across markets, but the magnitudes of the associations are smaller for market pairs that are closer to each other than market pairs that are farther from each other.

To estimate the influence of transport costs on price dispersion of various foods across markets, we use monthly consumer price monitoring panel data that the National Statistical Office (NSO) collects to compute the monthly Consumer Price Index (CPI) in Malawi. The dataset contains monthly retail prices for 26 various foods that were consistently collected from January 2007 through to July 2021 across 32 markets. We classified food items into animal source foods, vegetables, fruits, legumes and nuts, and staples including roots and tubers (FAO and FHI 360, 2016). Further, we obtained the route distances over paved roads between the market pairs from Google Maps on Malawi's paved road network and local diesel prices from the Malawi Energy Regulatory Authority (MERA). In combination with other data from various sources, we first estimate how the changes in transport costs are associated with overall price dispersion of various foods across markets using the panel non-linear dyadic regression model via the Poisson pseudo-likelihood regression estimator with multiple levels of fixed effects in the short run. Then, we investigate whether processed foods, perishable foods, or nutrient-dense foods modify the relationship between transport costs and overall price dispersion of various foods across markets in the short run. Finally, we estimate separately the panel non-linear dyadic regression model via the Poisson pseudo-likelihood regression estimator with multiple levels of fixed effects, and the instrumental variable Poisson (ivpoisson) estimator for each food item to examine the association between transport costs and price dispersion across markets in the short and long runs, respectively.

Most of the previous work on price dispersion of agricultural commodities has focused on most storable grains that are either semi-perishable (*e.g.*, cowpea, beans, and groundnuts) or less perishable (*e.g.*, millet, maize, rice, and sorghum) in developing countries.[1] To our knowledge, Jensen (2007) is the only study that has investigated how investment in information and communication technologies affect price dispersion of a highly perishable food commodity (*i.e*., fish) in India. A further contribution we make to the literature is to examine whether transport costs are associated with price dispersion of processed foods, more perishable foods (*e.g*., eggs, meat, fresh fruits, and vegetables), or nutrient-dense foods across markets in a developing country setting.[2] Nutrient-dense foods are foods with high concentration of essential vitamins and minerals that are important for nutrition and long-term health (Darmon & Drewnowski, 2015; Dittus et al., 1995; Drewnowski et al., 2004; Drewnowski & Specter, 2004; Rideout et al., 2015). Animal source foods, vegetables, and fruits are more nutrient-dense than legumes and nuts, and staples including roots and tubers. Although it is not always the case, more nutritious foods are often more perishable, have higher value to weight ratios, and are difficult to aggregate in large volumes but require a shorter time for traders to get them into the market. However, trade of highly perishable foods over longer distances is restricted in most developing countries including Malawi due to absence of cooled transport system (Fafchamps & Gabre-Madhin, 2006; Kachule & Franzel,

[1] Perishability is less of a problem in storable grains. In the absence of cooled transport system in most developing countries, more perishable foods require a shorter time to get into the market than storable grains.
[2] It is important to note that perishability is relevant for price integration, while nutrient-density is key for nutrition.

2009; Zant, 2018). Transport costs per unit volume reduce when market traders transport either larger loads (Fafchamps & Gabre-Madhin, 2006), which depends on how quickly they can be organised such as processed foods or food items with higher weight value ratios such as fruits and vegetables (Roehner, 1996), which depends on local demand in the destination market. Therefore, we hypothesise that the association between transport costs and price dispersion of food commodities is smaller for processed, less perishable, or less nutritious foods that can easily be organised into larger volumes than for unprocessed, more perishable, or more nutritious foods that take a lot of time to aggregate across markets.

Contrary to our expectation, the results from our analysis show that the endogenous increase in transport costs is associated with the reduction in overall price dispersion of various foods under investigation across markets in the short run, on average. We find that this counterintuitive influence is driven by processed foods but not by perishable foods and nutrient-dense foods for which the changes in transport costs are associated with the increase in overall price differences across markets in the short run. Moving on to separate results for each food item, we find that the association between the changes in transport costs and price differences remains negative and significant for rice grain, beef, goat meat, powdered milk, and bananas but become positive and significant for maize flour dehulled, maize grain (private), maize grain (ADMARC), brown beans, and eggs providing additional evidence that the counterintuitive influence of the changes in transport costs on overall price differentials are driven by food items that have lower search costs and are easy to aggregate in the short run. Turning to spatial heterogeneity of the association between transport costs and price differences, we find that the magnitudes of the relationships are smaller for market pairs that are closer to each other for all foods under consideration. Moving on to the influence of transport costs on price differences for each food item in the long run, we find that the price differences in the previous period play a small role in predicting the price differences in the current period, and the increase in transport costs is associated with the increase in price differentials for most foods under investigation, except for brown beans, usipa, utaka, and tomatoes.

The rest of the chapter is organised as follows. The next section presents the overview of food marketing in Malawi. Section 2.3 reviews related literature, while section 2.4 describes the data and methodology used in the study. The results are presented in section 2.5 and section 2.6 concludes.

2.2 Overview of food marketing in Malawi

Malawi's transport system comprises waterborne, rail, road, and air systems. The railway line covers about 942 km, which only serves the Central and Southern regions (Kaunda, 2016; Zant, 2018), but it is in poor condition and is usually not maintained (Robinson & Wakeford, 2013). Therefore, the road network remains the main mode of transport for both domestic and international trade, which uses about 56.4% of the imported liquid fuels (Kaunda, 2016; Lall et al., 2009; Robinson & Wakeford, 2013). About 15, 451 km of the road network connects cities, district capitals, and markets across the country. However, only about 26% of the road network is paved (Robinson & Wakeford, 2013). Over the period we investigate all the roads connecting local large markets across the country are paved.

Private traders play a critical role in the marketing of food in Malawi where agriculture is mostly dominated by smallholder farmers. Private traders purchase food commodities from smallholder farmers in the rural areas or periodic markets, aggregate, store and sell on to local food markets. It is difficult to organise most food commodities that smallholder farmers produce such as maize grain, legumes, roots and tubers because most of them produce smaller market surpluses and are geographically dispersed in the rural areas (Burke et al., 2020). Unlike most storable grains, rice production is mostly grown along the shores of Malawi's lakes; namely, Lake Malawi located in the Western Rift Valley, Lake Chilwa in Zomba district, Lake Chiuta in Machinga district, and Lake Malombe in Mangochi district. It is relatively easier to organise rice grain from small-scale rice millers who strategically operate along the main roads to access the grid power in the process acting as marketing points for rice grain (African Institute of Corporate Citizenship, 2016). In addition, the private traders supply storable grains to the Agricultural Development and Marketing Corporation (ADMARC), a parastatal enterprise, which sells the grains at subsidized prices across the country during the lean period. However, most private traders have limited access to private transport as a result they rely on the least-expensive means of available local transport such as bicycles, handcarts, and oxcarts in the rural areas to carry their products to the nearest paved road where they can easily access motorized transport (*i.e.*, open trucks). Along the paved roads, the traders target empty backhauls or open trucks locally known as *matola* to transport their products to designated markets (Fafchamps & Gabre-Madhin, 2006; Famine Early Warning Systems Network, 2018; Kachule & Franzel, 2009). The use of empty backhauls is popular among traders because the transport price can be negotiated. Fafchamps & Gabre-Madhin (2006) find that traders travel an average distance of 53 km with a maximum of 200 km to source out storable grains across the country. The distance private traders travel increases during the lean period. For instance, Famine Early Warning Systems Network (2018) reports that traders travel over 350 km to source out and supply storable grains to markets in the Southern region of Malawi during the lean period.

Although open trucks are not suitable to transport highly perishable food items such as fruits and vegetables because they are susceptible to physical damage and harsh conditions such as heat and rain, Kachule & Franzel (2009) report that traders use them to transport fruits across markets in Malawi. Fruits and vegetables have higher value to weight ratios, which would lower transport costs if traded over longer distances. However, these food items require cooled transport over longer distances to preserve their freshness. According to Kachule & Franzel (2009), Malawi has limited cooled transport, which only serve commercial farmers. As a result, long-distance trade of highly perishable food items particularly vegetables is limited compared to least perishable food items (Zant, 2018). This means that most traders source vegetables around the villages closer to the markets from smallholder farmers because of the need to get the vegetables quickly to the market to preserve their freshness. Conversely, fruits such as mangoes, tomatoes, and oranges are sourced from long distant locations when they are in season across the country. However, over 90 percent (20 000 Mt) of bananas are imported from Tanzania and Mozambique each year following infestation of the banana bunchy top virus in the country (Pondani, 2022; Tobacco Reporter, 2023). The private traders pack these fruits in 50/90 kg sacks or traditional baskets and transport them in open trucks across markets, which leads to high losses during transportation (Kachule & Franzel, 2009). Potatoes, sweet potatoes, and cassava are transported in a similar fashion like fruits across markets.

Trade of meat across markets does not involve slaughtered carcasses since refrigerated transport is limited in Malawi. Instead, live animals such as cattle and goats are transported over open trucks across markets, which is not suitable to transport them. Most markets have a slaughterhouse or designated place for slaughtering live animals. About 90 percent of the beef is supplied by smallholder farmers who are concentrated mainly in the lower Shire River Valley in the Southern region of the country (*i.e.,* Chikwawa and Nsanje districts) (Nyama World, 2017; Schmidt, 1969). Workman et al. (1998) observe that most smallholder farmers sell their cattle to local traders, butchers, and abattoirs during the lean period (*i.e*., before crop harvest) when livestock prices are lower. Similarly, supply of goats to local traders and butchers is dominated by smallholder farmers across the country (Banda & Dzanja, 2006; Chigwa, 2012). Turning to fish, about 85 percent of the fish that is consumed across the country is supplied by small-scale and subsistence fishers from Malawi's lakes (Allison & Mvula, 2002; Tran et al., 2022). Similarly, fresh or dried fish is transported over open trucks across markets. Production of eggs is largely dominated by commercial and semi-commercial poultry producers who operate in large urban centres mainly Blantyre, Lilongwe, Mzuzu and Zomba. Private market traders sell about 80 percent of the eggs from commercial poultry producers across markets mainly in urban centres (Commercial Agriculture for Smallholders and Agribusiness, 2020). We expect trade over longer distances to be limited for live animals and eggs due to difficulties in transporting them than dried fish across markets.

Processing of foods such as bread, buns, and ultra-pasteurized milk takes places in large urban centres. Similarly, repackaging of imported powdered milk takes place in large urban centres. Usually, private traders transport these food items in vans across the country, which allows transportation of larger volumes than open trucks.

2.3 Related literature

This section reviews various measure of price dispersion commonly used in empirical studies. Further, the section reviews previous studies that have used proxy measures for transport costs when examining the extent to which transport costs affect price dispersion of food commodities across markets or locations in developing countries.

2.3.1 Measures of price dispersion

There are several measures of price dispersion across markets that have been used in empirical studies such as the coefficient of variation (CV) in prices, difference between maximum and minimum prices, and absolute value of price difference (Aker, 2010a; Aker &

Fafchamps, 2015; Eckard, 2004; Jensen, 2007; Roehner, 1996; Zant, 2018). The CV measure is the ratio of the standard deviation of prices across markets within a location to the location's mean prices (Eckard, 2004). The CV closer to one indicates that price dispersion is greater around the mean, while the CV closer to zero indicates price dispersion is lower around the mean. Further, the CV equals to zero indicates that the average prices are equal across markets. This means that larger CV signifies price divergence, while smaller CV signifies price convergence. Turning to the difference between maximum and minimum prices, this measure captures the difference between the highest and the lowest average prices across markets in each location (Jensen, 2007). The larger the difference the larger the level of price dispersion, while the difference equals to zero indicates that the average prices are equal across markets. Similarly, this means that larger differences signify price divergence, while smaller differences signify price convergence. Finally, the absolute value of price difference captures the difference in prices across markets as a measure of price dispersion (Aker, 2010a). The larger the absolute difference the larger the level of price dispersion, while the absolute difference equals to zero indicates that the average prices are equal across markets. Similarly, this means that larger absolute differences signify price divergence, while smaller absolute differences signify price convergence.

2.3.2 Measures of transport costs across markets

Availability of data on transport rates or charges across various routes is limited in most developing countries. As a result, most studies use direct proxies for transport costs such as fuel costs, transport rates, proximity to a railway line, and removal of a fuel subsidies when examining the influence of transport costs on price dispersion (Aker, 2010a; Aker et al., 2014; Andersson et al., 2017; Fuje, 2019, 2020; Jensen, 2007; Roehner, 1996; Zant, 2018). Fixed effects estimator is the most used methodology in this literature.

2.3.2.1 Fuel costs

Jensen (2007) is among the first to investigate the influence of transport costs in fish marketing using fuel costs as a proxy measure of transport costs in the state of Kerala in India. The author finds that the increase in fuel prices, which increases transport costs, was correlated with higher price dispersion of fish across markets using the difference-in-differences setting.

However, the use of fuel costs alone as a proxy for transport costs does not capture the effects fuel costs have on distance that traders travel across markets.

2.3.2.2 Transport rates

There are other studies that use transport rates between market pairs to examine how transport costs influence price dispersion of storable grains. Aker (2010b) uses the cost of transporting millet per kilometre that was obtained from the Syndicat des Transporteurs Routiers (*i.e.*, a transporting company) as a measure of transport costs to show that the cost of transporting millet per kilometre increases price dispersion of millet grain across markets in Niger, while Aker et al. (2014) use estimated transport costs based on the distance between market pairs separated by the border between Niger and Nigeria and show that the increase in transport costs increases price dispersion of millet and cowpeas across the border. Similarly, Salazar et al. (2019) obtained transport costs data from the 'Sistema De Informação De Mercados Agrícolas De Moçambique' and predicted transport cost data for market pairs that had missing data using diesel price, distance, and floods occurrence indicator to show that transport costs increase price dispersion of maize grain across markets in Mozambique. Although this approach captures transport costs that traders incur, data on transport rates across various routes may not be available in most developing countries making it difficult to monitor the influence of transport costs on market integration of food commodities across space and time.

2.3.2.3 Proximity to a railway line

It is widely recognised that railway transport is least-expensive relative to road transport. For example, Zant (2018) show that proximity to railway transport services reduces price dispersion of maize, rice, groundnuts, and beans across markets in Malawi. Thus, railway transport system promotes domestic trade for storable grains and improves market efficiency. However, the use of railway transport is limited among traders in most developing countries. As a result, road transport remains the main means of moving food commodities across markets in most developing countries.

2.3.2.4 Removal of fuel subsidies

Other studies use the change in fuel policy to examine the effect of higher fuel prices on price dispersion of storable grains. Fuje (2019) show that the removal of the fuel subsidy on 4th October in 2008 in Ethiopia increased price dispersion of "teff", wheat, maize, sorghum, and barley between the capital city (*i.e.*, Addis Ababa) and other districts in the short run using a regression-discontinuity design, while Fuje (2020) show that the fuel subsidy reform increased price dispersion of "teff", wheat, maize, sorghum, and barley between the capital city (*i.e.*, Addis Ababa) and other districts in Ethiopia in the long run using a difference-in-differences specification with distance as a continuous treatment variable. While the use of the change in the fuel policy as a proxy measure of transport costs captures the change in transport costs between two regimes (*i.e.,* the fuel price shock), it does not capture the time-to-time changes in transport costs arising from the changes in fuel prices that traders incur when engaging in distance trade across markets.

2.3.3.5 Persistence in price differences

The previous studies have also found that the price differences in the previous period play a smaller role in predicting the price differences in the current period for most storable grains. For instance, Aker, (2010a, 2010b) finds a smaller estimate on the lagged price difference of millet (*i.e*., between 0.18 and 0.36) in Niger, Zant (2018) finds a smaller estimate of less than 0.25 on the lagged price difference for maize, rice, groundnuts, and beans in Malawi, while Salazar et al. (2019) finds a smaller estimate of 0.5 on the lagged price difference for maize in Mozambique. Thus, there is low persistence in price differences for food commodities across markets in developing countries.

In conclusion, these previous studies have shown that the increase in transport costs increases price dispersion of food commodities across markets. This chapter advances this literature by using a novel measure of transport costs as in Storeygard (2016) to systematically examine the implication of fuel price changes on price dispersion of various foods across markets in a developing country setting where fuel prices are endogenously determined. Further, these previous studies have focused on the effects of transport costs on price dispersion of storable staple grains that are less perishable such as millet, maize, rice, and sorghum. A further contribution we make to the literature is to examine the influence of transport costs on price dispersion of processed foods, more perishable foods, and nutrient-dense foods such as fruits and vegetables across markets in Malawi.

2.4 Methods

2.4.1 Conceptual framework

There are three main channels through which fuel prices would affect market prices of food (Abbott et al., 2008; Dillon & Barrett, 2016). According to Dillon & Barrett (2016), fuel price shocks would affect food prices either indirectly by increasing (i) the cost of production (*e.g.*, farm inputs such as inorganic fertilisers, and cost of operating farm machines such as tractors and pumps on fuel) and (ii) the demand to use corn to produce biofuel, which increases global maize prices and later transmit to maize prices in local markets or (iii) directly by increasing the transport costs. Dillon & Barrett (2016) examine the extent to which oil price shocks transmit to local maize prices across markets in Ethiopia, Kenya, Tanzania, and Uganda. The authors find no evidence of the indirect effect of fuel prices on local maize prices because of limited use of fuel-powered farm machinery in production of maize, and local maize prices did not respond to the changes in inorganic fertiliser prices. Conversely, the authors find that local maize prices responded more rapidly to global oil prices than global maize prices in the study countries. Thus, the main channel through which fuel price shocks affect food prices is via the direct effect on transport costs, which affect trade over longer distance in developing countries. According to Teravaninthorn & Raballand (2009), transport costs comprise the direct and indirect costs of operating a vehicle such as fuel, tires, insurance, and toll and roadblock payments.[3]

We use Enke-Samuelson-Takayama-Judge equilibrium model for spatially separated markets to examine whether variation in transport costs is associated with price dispersion of various foods across markets (Enke, 1951; Samuelson, 1952; Takayama & Judge, 1971). Assume two spatially separated markets (x, y) that are involved in direct trade for a homogenous food item (*e.g.*, maize grain) and trade is bidirectional. Further, assume the price of maize grain is P_{xt} in market x at time t and P_{yt} in market y at time t, and the transfer or transaction cost of

[3] Transport costs also include labour, time, and externality costs such as accidents, pollution, and congestion.

spatial arbitrage between the two markets is $\pi_{xy,t}$ at time t. The spatial market equilibrium model stipulates that the two markets, x and y, are in a long run competitive equilibrium when there is zero marginal profit to spatial arbitrage. Thus, the spatial arbitrage condition for direct trade between two markets with perfect integration (*i.e.*, well-functioning markets) are specified as follows:

$$P_{xt} - P_{yt} = \pi_{xy,t} \qquad (2.1)$$

This equilibrium condition means that traders would relocate the food commodity from the market in the surplus location to the market in deficit location at time t if the price difference covers the transport costs between them. Conversely, there is no incentive to trade if the price difference is less than transfer costs between the two markets (Aker, 2010b; Barrett & Li, 2002; Baulch, 1997; Fackler & Goodwin, 2001). Thus, any deviation from the spatial arbitrage equilibrium condition should be temporary in nature.

Equation [2.1] allows us to examine the influence of the changes in the transfer costs due to either policy change or shocks on the marginal profit of spatial arbitrage. The transfer costs, π_{xy}, may include search, transport, and insurance costs (Aker, 2010b; Andersson et al., 2017; Barrett & Li, 2002; Fafchamps & Gabre-Madhin, 2006; Zant, 2013). Since it has been estimated that transport costs comprises a larger share of transfer costs contributing about 39 percent in Malawi (Fafchamps & Gabre-Madhin, 2006) and that the larger share of the transport costs is fuel cost contributing about 40 percent in Africa (Teravaninthorn & Raballand, 2009), we anticipate the shock to transport costs due to the changes in fuel prices to have a positive association with spatial equilibrium price dispersion of various foods across markets (Aker, 2010b). This means that changes in transport costs directly affect the operations of market traders. Therefore, as transport costs increase, market traders will either increase the price of food in deficit market to cover transport costs or stop supplying certain food items, which they cannot recover their transport costs. As a result, consumers will face higher food prices or may be unable to access certain food items at all in deficit locations, while in surplus locations food will remain local at lower prices, thereby increasing price differentials of various foods across markets. Thus, consumers in surplus areas will benefit from lower food prices, while consumers in deficit areas will experience high food prices.

2.4.2 Empirical strategy

We use Aker's estimation procedure to examine the association between transport costs and price dispersion of various foods across markets in the short run (Aker, 2010b, 2010a; Aker et al., 2014). In our context, short-run effects are important given that traders immediately incorporate transport costs in setting up retail prices for food items sourced from various locations (Storeygard, 2016). This has an immediate impact on the price and availability of various foods in local markets, given that traders will not relocate food items from surplus production areas where the prices are lower to deficit locations where prices are higher if price differences do not cover transport costs. We specify our conceptual model in equation [2.1] for market-pair, x and y, at time t as follows:

$$P_{xy,t} = \beta_0 + \beta_1(dist_{xy})fuel_t + \sum_{j=1}^{3} \sigma_j \left(dist_{xy}\right)t^j + \omega_1 X_{xy,t} + \lambda_m + \tau_t + \theta_{xy} + \varepsilon_{xy,t} \quad (2.2)$$

where $P_{xy,t}$ is the absolute value of the price difference $|P_{xt} - P_{yt}|$, our measure of price dispersion, between market x and y at time t. $fuel$ represents diesel fuel pump price in thousands of Malawian Kwacha (MWK), and $dist$ represents the absolute value of the route distance over paved road between the two markets in hundreds of kilometres (Km), $|dist|$. An interaction between fuel price and distance between market pairs is our measure of transport costs. Thus, β_1 is our parameter of interest in equation [2.2]. The underlying assumption for identification of a causal impact of a fuel price shock on price differentials is that our measure of transport costs should not be correlated with the error term. Thus, there should not be other factors that affect the distance between market pairs and the change in price of fuel that affect price differentials. A positive and significant coefficient estimate on β_1 shows that the increase in fuel price is positively associated with the increase in price differentials, while a negative and significant coefficient estimate on β_1 shows that the increase in fuel price is negatively associated with the decrease in price differentials, on average. Further, we include distance-specific cubic time trends, $\sum_{j=1}^{3}\left(dist_{xy}\right)t^j$, to pick up differential effects by distance of potential omitted variables that change slowly over time such as road quality changes. Thus, the variation of the fuel price around the smooth trends allows β_1 to identified.

X is a set of market-pair time varying controls such as population density, crop or livestock production, and occurrence of floods or drought in the locations the markets operate that affect price dispersion between market x and y at time t (see table A.1 in the appendix for a description of the explanatory variables). λ_m is a set of monthly dummies, which capture seasonality in food availability between market x and y at time t. Further, we include time-specific fixed effects, τ_t, to capture time varying shocks common to all markets that are either observed or unobserved such as fuel price changes and devaluation of the local currency, and market-pair-specific fixed effects, θ_{xy}, to control for market level time invariant characteristics that are either observed or unobserved in every month such as the distance between market pairs, physical structures at the market, connection to the national power grid, and frequency of operation. Thus, covariates that are constant across market pairs or covariates that only vary across time will be wiped out by the fixed effects (Baltagi, 2021; Wooldridge, 2021). This offsets potential sources of omitted variable bias from variables that are collinear with both the market-pair and time-specific fixed effects, whether observed or unobserved.[4] β_0 represents the constant whereas σ_j and ω_1 are all unknown parameters to estimate. $\varepsilon_{xy,t}$ is the market-pair random error term.

Equation [2.2] is a dyadic regression. According to Fafchamps & Gubert (2007), estimation of a dyadic regression poses challenges relating to how the covariates enter the regression, number of links of each observation, and how to obtain correct standard errors. The dyadic relationship contains two sets of information or covariates (i) relating to the relationship between x and y, for example, our measure of distance between market pairs, and (ii) specific to each market x and y, for example, population density in each market. The first set of covariates connecting x and y enters equation [2.2] as they are. Conversely, to preserve the effect of $X_{x,y}$ on P_{xy} and the effect of $X_{y,x}$ on P_{xy}, the second set of covariates must be added to equation [2.2] in a symmetric manner depending on whether the relationship is directional or bidirectional (see, Fafchamps & Gubert, 2007). In our context symmetry is achieved when $\omega_1 X_{x,y} = \omega_1 X_{y,x}$ given that the relationship between the market pairs is assumed to be bidirectional (*i.e.*, $P_{xy} = P_{yx}$). Thus, each covariate of the second set should enter equation [2.2] as two covariates of the form $\omega_1|X_{1x} - X_{1y}|$ and $\omega_2|X_{1x} + X_{1y}|$ where ω_1 captures the

[4] However, the limitation of our methodology is that we are not able to control for time-varying market level characteristics that are either observed or unobserved, which may bias our estimates.

effect of differences in X_{1x} and X_{1y} on P_{xy} while ω_2 captures the effect of their combination on P_{xy}. For instance, population density should enter equation [2.2] as two covariates, namely the absolute difference in population density and the sum in population density between market x and y. However, Fafchamps & Gubert (2007) indicate that ω_2 is not identified if each observation has the same number of links due to dependence in dyadic relationships. Given that each market has the same number of links, we estimate equation [2.2] with covariates only in their absolute differences.[5]

We estimate equation [2.2] separately for each individual food item. Further, we obtain cluster-robust standard errors by clustering on market to account properly for all relevant dyadic error correlations or dependences in the data, which may lead to underestimation of the standard errors and larger t-statistics (*i.e.*, inaccurate inference) (Aronow et al., 2015; Cameron & Miller, 2014; Fafchamps & Gubert, 2007).[6] To determine how sensitive our estimates are to various distance levels between the market pairs, we re-estimate equation [2.2] with discretized distance.[7]

Population density

Fafchamps & Gabre-Madhin (2006) indicate that population density is related to transport and search costs that market traders incur. The authors indicate that high population density allows market traders to quickly organise larger loads, which reduces waiting time for transporters and lead to lower transport and search costs. Thus, population density increases trade or transaction frequency and efficiency. Therefore, we expect the increase in the difference in population density between markets to be associated with the decline in price

---

[5] Thus, we estimate equation [2.2] as

$$P_{xy,t} = \beta_0 + \beta_1(dist_{xy})fuel_t + \sum_{j=1}^{3} \sigma_j\,(dist_{xy})t^j + \omega_1|X_{x,t} - X_{y,t}| + \lambda_m + \tau_t + \theta_{xy} + \varepsilon_{xy,t}$$

[6] Clustering on one of the markets controls for spatial dependence between markets, while clustering on market pair controls for spatial dependence over time for each market pair (Aker, 2010a, 2010b; Aker & Fafchamps, 2015; Aronow et al., 2015; Cameron & Miller, 2014). Thus, clustering standard errors by market deals with general forms of serial correlation.

[7] We categorised the distance between the market pairs into 10 quantiles.

dispersion because higher population density is associated with higher food prices creating opportunities for spatial arbitrage (Ricker-Gilbert et al., 2014).

Floods or drought occurrence

Floods or drought occurrence may increase or reduce agricultural production, which may affect food supply in local markets and trade flows (Aker, 2010b; Salazar et al., 2019). Aker (2010b) finds that occurrence of droughts reduces price dispersion between markets for millet in Niger. Similarly, Salazar et al. (2019) find that droughts reduce price dispersion of maize, while floods increase price dispersion across markets in Mozambique. Thus, market integration is higher during occurrence of droughts than floods. Therefore, we expect price dispersion between markets to decline if one of the markets has a drought shock because food items are more likely to move from the unaffected area where prices may be lower to the affected area where food prices may be higher due to the reduction in agricultural production. Conversely, we expect price dispersion between markets to increase if one of the markets has a flood shock because rainfall above the mean leads to better agricultural production (Björkman-Nyqvist, 2013; Zant, 2018; Zimmermann, 2020). As a result, food items are less likely to move between areas that produce surplus agricultural production leading to the increase in price dispersion (*i.e.*, trade between surplus markets is limited).

Local production

Usually, high local production increases the supply of food, which relatively lowers food prices, while low local production reduces the supply of food, which relatively increases food prices (Minten & Kyle, 1999; Zant, 2018). Therefore, we expect the increase in the difference in local production between markets to be associated with a decline in price dispersion, which creates opportunities for spatial arbitrage from the surplus area to the deficit area (Aker, 2010b).

Potential endogeneity of local diesel price

Figure 2.1 presents the variation in local diesel price compared to global oil prices over time. The figure shows that local diesel and global oil prices followed a similar pattern (*i.e.*, local diesel price fluctuated in parallel with global oil prices) over the period we investigate (see figure A.9 in the appendix for the variation of uniform diesel prices across the country over time).[8] The Pearson correlation coefficient is 0.72 ($p<0.000$). Thus, local diesel prices are strongly positively correlated with global oil prices. While Malawi is a price taker (*i.e.,* the country cannot influence the price of fuel substantially), the government may decide whether and when to pass on changes in the global price to local diesel prices to avoid a spike in food prices thereby introducing a potential reverse causality. This is evident during the period when the government implemented the fuel subsidy (*i.e.*, between 2007 and 2012) where the price of local diesel fuel is less variable than the price of global oil compared to the period the fuel subsidy was removed (*i.e*., 2012 and 2021).[9] This means that local fuel prices may not adjust simultaneously to the changes in global oil price, indicating that the variation in local diesel price is potentially endogenous in Malawi.

[8] The variation of fuel price in local currency exhibits a step function with prices being flat between 2007 and June 2012, a sudden jump between June 2012 and January 2014, and then flatten again between January 2014 to July 2021.

[9] The Pearson correlation coefficient during the fuel subsidy period is 0.17 ($p<0.1950$), while the Pearson correlation coefficient during the period the fuel subsidy was removed is 0.85 ($p<0.000$). Thus, local diesel prices are strongly positively correlated with global oil prices after the removal of the fuel subsidy.

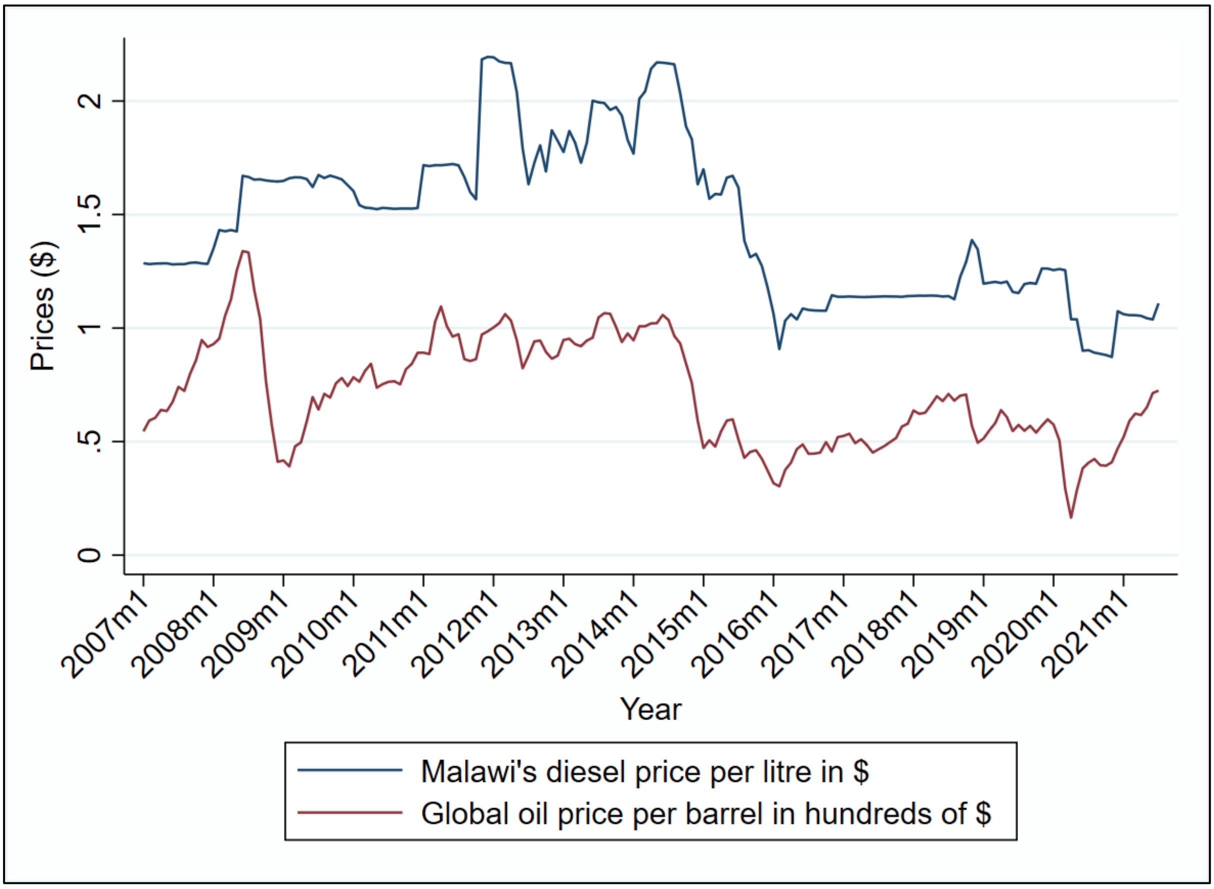

Figure 2.1: Variation in local diesel prices vs global oil prices over time
Source: Local diesel prices are from MERA, while global oil prices are from Thomson Reuters (2022) and data is available at http://www.eia.gov/dnav/pet/pet_pri_spt_s1_d.htm

Functional form

The next step is to determine the functional form and find an estimator for equation [2.2]. Our price differences across markets are piled at zero for each food item. This means that the prices do not differ at all in some months across the market pairs. Thus, the zeros in our data are observed values. A corner solution model or a model for count data is more plausible than a Heckman selection model. The Heckman selection model would be appropriate to deal with incidental truncation if the zeros in our data were missing or unobserved values. Therefore, equation [2.2] can be estimated appropriately using either the tobit estimator (Tobin, 1958) and the more flexible double hurdle model (Cragg, 1971) for the corner solution model or Poisson regression estimator for the count data. In equation [2.2] we allow the unobserved market-pair effect, θ_{xy}, to correlate arbitrarily with some of our covariates (*i.e.*, θ_{xy} as a parameter to estimate). While equation [2.2] can be estimated using Correlated random effects framework (CRE) (Chamberlain, 1982; Mundlak, 1978; Wooldridge, 2010, 2019), we use the fixed effects Poisson estimator because of its attractive features. Just like with ordinary least squares (OLS), the fixed effects Poisson estimator simply requires the conditional mean function of the dependent variable to be correctly specified for consistency, allows for arbitrarily dependence between unobserved effect and covariates, and the dependent variable does not need to be a count variable (Correia et al., 2019; Gourieroux et al., 1984; Hausman et al., 1984; Wooldridge, 1999, 2010). Thus, the dependent variable can be continuous or a corner solution (*i.e.*, specification of the distribution assumption of the dependent variable is not required or is unrestricted). When the conditional mean function of the dependent variable is correctly specified, Poisson regression comes to be the Poisson pseudo maximum likelihood (PPML) regression. Further, observations of the dependent variable that are equal to zero are naturally dealt with and do not cause the sample selection problem (Correia et al., 2019; Santos Silva & Tenreyro, 2006; Wooldridge, 1999, 2010). We implement the fixed effects Poisson estimator (*i.e.*, PPML) using Stata's command for estimating (pseudo) Poisson regression models with multiple high-dimensional fixed effects (*i.e.*, PPMLHDFE) because it converges much faster in the presence of fixed effects and estimates are consistent in the presence of heteroskedasticity (Correia et al., 2019; Santos Silva & Tenreyro, 2006). However, we are not able to control for potential endogeneity of the local diesel price using oil price which is set on a global market as an instrument for diesel local price because PPML suffers from the incidental parameter problem when the model has time-specific fixed effects (Cameron & Trivedi, 2013; Storeygard, 2016).

Robustness checks

We also estimate alternative specifications to our main specification, equation [2.2], as robustness checks. To test whether there is saturation effect (*i.e.*, non-linear effect) of the association between transport costs and price differences, we estimate equation [2.2] with squared distance interacted with fuel price as an additional regressor, $(dist_{xy})^2 fuel_t$. Over time, fuel prices are autocorrelated such that the current fuel price can be used as a proxy for last month's fuel price (Storeygard, 2016). We re-estimate equation [2.2] with distance interacted with the current fuel prices (*i.e.*, contemporaneous impact) along with distance interacted with lagged fuel price (*i.e.*, enduring, or persistent impact). Further, the Enke-Samuelson-Takayama-Judge spatial equilibrium model is a spatial autoregressive model where the current price differences depend on the previous price differences (Aker, 2010a; Roehner, 1996). To control for the price differences in the previous period, $P_{xy,t-1}$, we respecify equation [2.2] as follows:

$$P_{xy,t} = \delta_0 + \delta_1 P_{xy,t-1} + \delta_2 (dist_{xy}) fuel_t + \sum_{j=1}^{3} \gamma_j \, (dist_{xy}) t^j + \varphi_1 X_{xy,t} + \lambda_m + +\theta_{xy} + \mu_{xy,t} \quad (2.3)$$

where $P_{xy,t-1}$ is the lag of the price difference, which takes on both zero or positive values depending on whether the previous price difference was a zero or positive value. δ_1 is parameter of interest, which measures whether there is state dependence in the price differences (*i.e.,* whether the price differences in the previous period help to predict the price differences in the current period). We expect δ_1 to be between 0 and 1. The more the price difference persists the closer the coefficient would be to 1. Further, δ_1 can be interpreted as the speed of adjustment to the long-run equilibrium (Aker, 2010a).[10] δ_0 represents the constant whereas δ_1, δ_1, δ_2, γ_j and φ_1 are all unknown parameters to estimate. $\mu_{xy,t}$ is the market-pair random error term. The rest of the variables are the same as those in equation [2.2].

Estimating equation [2.3] using the fixed effects Poisson estimator will be inconsistent because the lag of the price difference and market-pair fixed effects are correlated (Cameron & Trivedi, 2013; Wooldridge, 2005). There are several models that have been established based on how the lagged dependent variable is incorporated into the conditional mean of the exponential function mean in the presence of individual-specific effects. However, the preferred model for count data has not been established in the literature for both panel and time series data compared to the linear case (see, Cameron & Trivedi, 2013). An exponential feedback model (EFM) allows the lagged dependent variable to enter the conditional mean in levels, which may be explosive for $\delta_1 > 0$ because the dependent variable is nonnegative and may poorly fit the data due to possible sharp discontinuities (Cameron & Trivedi, 2013). A solution to this

[10] In equation [2.3], the association between transport costs and price differences in the long run is measured as: $\frac{\delta_1}{(1-\delta_1)}$.

problem is to allow the lagged dependent variable to enter the conditional mean in logs and adjust for the values equal to zero with a constant that takes the values between zero and one. Blundell et al. (2002) establish a linear feedback model (LFM) that allows the lagged dependent variable to enter the conditional mean linearly in levels, which prevent sharp discontinuities. Another approach is to use GMM estimator or nonlinear instrumental variables approach to identify the parameter on the lagged dependent variable where the additional lags of the dependent variable are used as instruments for the first lag of the dependent variable (Cameron & Trivedi, 2013). However, a major limitation with this estimation procedure is that estimation suffers from the incidental parameter problem when the model has time-specific fixed effects, especially for large panels (Cameron & Trivedi, 2013).

Another approach is to use a random effect framework where the unobserved heterogeneity may be allowed to correlate with the initial condition (*i.e.*, initial observations of the dependent variable) (Cameron & Trivedi, 2013; Trivedi, 2010; Wooldridge, 2005). However, the distribution of the dependent variable or unobserved heterogeneity is required for the parameter on the lagged dependent variable to be identified (see, Wooldridge, 2005). Wooldridge (2005) integrates the unobserved heterogeneity out of the density distribution of the dependent variable given the covariates, averages of time-variant covariates, and initial observations to obtain the density distribution of the unobserved heterogeneity, which has a random effects Poisson form with the Gamma distribution. Given that the initial observations are not random, the usual random effects Poisson estimator (*i.e.*, random effects maximum likelihood approach) can be used to identify the parameter on the lagged dependent variable (Cameron & Trivedi, 2013; Trivedi, 2010; Wooldridge, 2005). However, this procedure is limited to balanced panel data (Trivedi, 2010; Wooldridge, 2005).

Pooling the data may be adequate where individual-specific effects may be ignored and the correlation between dependent variable and the lagged dependent variable (*i.e.*, serial correlation) may be controlled by including sufficient lags of the dependent variable. Further, time series methods may be used. Zeger-Qaqish model is one of the time series methods, which allows the lagged dependent variable to be incorporated into the conditional mean in logs and the lags of the dependent variable determine the conditional mean (Cameron & Trivedi, 2013). Similarly, the values of the dependent variable equal to zero are rescaled to the constant. Estimation with the fixed effects Poisson estimator is consistent since asymptotics are achieved as $T \to \infty$ (Blundell et al., 2002; Cameron & Trivedi, 2013; Hill et al., 1998). According to Cameron & Trivedi (2013), consistency of the Poisson QMLE is achieved once sufficient lags of the dependent variable are used, the functional form is correctly specified (*i.e.,* $\varepsilon_{t|t-1}$), and heteroskedastic-robust standard errors are used for inference. However, major limitations relate to an ad hoc choice of the value of the constant and pooling the data into a single long time series sample (Cameron & Trivedi, 2013). Since we treat the market-pairs as clustered samples with T possible observations and we are interested in the individual-specific fixed effects, θ_{xy}, we use nonlinear instrumental variables approach, instrumental variable Poisson (ivpoisson) estimator, with $P_{xy,t-2}$, $P_{xy,t-3}$, and $P_{xy,t-4}$ as instruments for $P_{xy,t-1}$ without time-specific fixed effects. The ivpoisson estimator implements a two-step GMM estimation procedure with additive or multiplicative errors (Windmeijer & Santos Silva, 1997). We allow time varying covariates to capture time variation.[11]

Construction of market pairs and potential selection bias

For N number of markets, we can construct a set of W market-pairs as follows:

$$W_{xy} = \frac{1}{2}N(N-1), \qquad x = 1, \dots, N-1; y = x+1, \dots, N \qquad (2.4)$$

W is also known as dyads in network formation literature.[12] W comprises undirected market pairs (*i.e.,* W_{xy} = W_{yx}) whereas self-connected market pairs (*i.e.,* W_{xx} or W_{yy} leading to $x = y$) are removed (Aronow et al., 2015; Cameron & Miller, 2014; Graham, 2017; Roehner, 1996) Further, we remove duplicates where $x > y$ in undirected market pairs given that W_{xy} = W_{yx} (Cameron & Miller, 2014).[13] With T (*i.e.,* $t = 1, \dots, T$) time periods, there can be as many as WT possible observations. Assume $S = 1$ if we observe prices in each market pair and zero otherwise. For each market pair W, we draw a subsample of market pairs G whenever $S = 1$. Thus, our subsample G is restricted to market pairs that have price observations that allows

[11] Given that we do not explicitly control for the market-pair and time fixed effects, our coefficient of interest, δ_2, might also be picking up the effects of other factors that are correlated with the fuel price such as devaluation of the local currency.

[12] This means that our primary equation [2.2] has a canonical form where the price differences and the covariates are a series of $\frac{1}{2}N(N-1)$ matrices (Fafchamps & Gubert, 2007).

[13] Undirected relationship means that W will contain 2 observations per market pair in the direction from x to y and another one in the direction from y to x.

us to compute price differences between the market pairs. Selection of the market pairs into G based on the outcome (*i.e.*, $S = 1$) may introduce a sample selection bias problem, which would lead to inconsistent estimates in equation [2.2]. Thus, selection bias may arise whenever we exclude market pairs with missing prices.

There are several mechanisms that may lead to missing prices in food markets. According to Andree (2021), prices may be missing in some local markets when food items are available due to inadequate resources to collect data, impassable roads connecting the markets, and sudden local conflicts and crises that interfere with data collection. Further, food prices may tend to be missing in some markets at certain times of the year despite efforts to collect the data (Gilbert et al., 2017). If the prices are missing completely at random, then estimation of equation [2.2] with our subsample will be consistent. This may be the case, for instance, when the prices for the market pairs always tend to be missing for the whole year or at the same time during each year. Conversely, if the missingness in the prices is not at random, then estimation of equation [2.2] with our subsample will be inconsistent due to sample selection bias problem. This may be the case, for instance, when the prices for the market pairs exist in some months and not in other months without a pattern or not systematically.

To determine how the missingness may affects our main results, we proceed as follows: (i) we set the missing price differences to their maximum price differences observed in each year between the market pairs and re-estimate equation [2.2] to determine the size of the bias. Although this ignores demand, it reflects the idea that the food item is missing because it is too expensive to supply for market traders. (ii) We estimate equation [2.2] with market-month fixed effects to control for market-specific seasonality (Dietrich et al., 2022; Zant, 2018). Controlling for market-pair specific seasonality and market-pair fixed effects could also help to deal with the sample selection bias when missingness in prices is related to seasonality.

2.4.3 Data

This study uses consumer price monitoring data that Malawi's National Statistical Office (NSO) collects to compute the monthly Consumer Price Index (CPI). This dataset was made available to the Changing Access to Nutritious Diets in Africa and South Asia (CANDASA) project (Kaiyatsa et al., 2019). It contains monthly retail prices for 26 various foods that were consistently collected from January 2007 through to July 2017 across 29 rural markets (See

Kaiyatsa et al. (2019) for a description on how the data are collected). Further, we collected additional price data from NSO from August 2017 through to July 2021 across the 29 rural markets, and additional price data for 3 urban markets from January 2007 through to July 2021. Thus, we have price data for 32 markets across the country spanning from January 2007 through to July 2021 (see figure A.1 for the spatial distribution of food markets across the country). In accordance with FAO and FHI 360 (2016), we classified the food items into animal source foods (7 items), vegetables (5 items), fruits (2 items), legumes and nuts (3 items), and staples including roots and tubers (9 items). Then, we classified the food items into processed and unprocessed food items to identify foods that can easily be organised into larger volumes. About 22 food items are unprocessed, while only 4 food items are processed (*i.e.,* ultra-pasteurized milk, powdered milk, white bread standard loaf, and white buns). We also grouped the food items into more perishable and less perishables foods. About 20 food items are more perishable, while 6 food items are less perishable. Finally, we classified the food items into more nutritious (animal source foods, vegetables, and fruits) and less nutritious (legumes and nuts, and staples including roots and tubers) foods. About 14 food items are more nutritious, while 12 food items are less nutritious.

Then, we compiled secondary data from various sources. The data on monthly average diesel pump prices were obtained from the Malawi Energy Regulatory Authority (MERA), annual population density (both current and projections) in the districts the markets operate were obtained from the Malawi's Population and Housing Census, which are collected and made available by NSO. Further, we obtained annual agricultural, livestock, and fish production data from the Ministry of Agriculture. The climate data (*i.e.*, daily precipitation, minimum and maximum temperature, and elevation) for computation of the Standardised Precipitation-Evapotranspiration Index (SPEI) for each location the markets operate were obtained from the National Aeronautics and Space Administration (NASA) Langley Research Center (LaRC) Prediction of Worldwide Energy Resource (POWER) Project funded through the NASA Earth Science/Applied Science Program.[14] The daily climate data are available from 1981 through to 2024. Given our data, we calculate SPEI values using Hargreaves approach for each location in

[14] https://power.larc.nasa.gov/data-access-viewer/

each month using a time scale of three months (Beguería & Vicente-Serrano, 2017).[15] SPEI takes on both negative and positive values, where positive values represent wet events and negative values represent dry events (Letta et al., 2021). Thus, larger positive SPEI values signal floods while smaller negative SPEI values signal drought. We use the larger positive SPEI values (>=90th percentile) to construct a dummy variable for floods, and the smaller negative SPEI values (<=20th percentile) to construct the dummy variable for drought conditions. Finally, we obtained the route distances over paved roads between the market pairs from the Google Maps on Malawi's paved road network.

2.4.4 Descriptive statistics

Table 2.1 presents the average monthly price differences of each food item across market pairs from Jan 2007 through to July 2021.[16] With 32 markets, we would have 496 market pairs with 86, 800 possible observations per food item over 14 years and 7 months (*i.e.*, 175 months). However, not all foods are available at each market in each month due to seasonality and stockouts.[17,18] Table 2.1 shows that the average price differences of each food item in MWK/Kg varies widely across the market pairs. Powdered milk, utaka (Lake Malawi cichlid), and usipa (Lake Malawi sardine) are the only food items with price differences of greater than MWK500/kg across market pairs. Most food items have price differences of less than MWK200/kg across markets.

[15] SPEI is calculated as the difference between precipitation (w) and potential evapotranspiration (z): $y_i = w_i - z_i$. y_i is climatic water balance. The log-logistic probability distribution is used to standardize y_i to allow comparison over space and time, and at various time scales (Beguería et al., 2014; Vicente-Serrano et al., 2010).
[16] Figure A.2 for the proportion of the price differences for each food item equal to zero. The results indicate that maize grain (ADMARC), goat meat, beef, and white buns have greater than 10% of price differences equal to zero. The proportion of price differences equal to zero is smaller for most food items.
[17] Figure A.3 shows the proportions of the most available food item over the period under investigation. The results indicate that tomatoes, maize grain (private), onions, beef, brown beans, white beans, goat meat, usipa, potatoes, rape leaves, maize grain (ADMARC), rice grain, eggs, cabbage, fresh milk, groundnuts, white buns, and bananas have greater than 80% of observed prices across the market pairs while cassava has less than 60% of observed prices across the market pairs.
[18] Figures A.4 – A.8 present the average month market-pair availability of each food item over the period under investigation. The results indicate that average month market-pair availability each year varies across the food items.

Table 2.1: Summary statistics for monthly food price differences in MWK/Kg

Food classification	Value addition	Food item	Obs.	Mean	Std. Dev	Min	Max
Staples including roots and tubers (*Less nutrient-dense*)	Unprocessed/more perishable	Fresh cassava	44023	47.79	55.89	0.00	542.04
	Unprocessed/more	Maize flour *Woyera*	57195	142.72	139.21	0.00	1030.30
	Unprocessed/less perishable	Maize grain (private)	79971	38.72	40.95	0.00	338.90
	Unprocessed/less perishable	Maize grain ADMARC	70099	19.60	37.14	0.00	190.00
	Unprocessed/more perishable	Potatoes	77243	74.59	76.78	0.00	649.47
	Unprocessed/less perishable	Rice grain	69666	94.04	90.98	0.00	682.34
	Unprocessed/more perishable	Sweet potatoes	59931	36.68	39.57	0.00	392.18
	Processed/more perishable	White bread standard loaf	76513	30.44	32.48	0.00	317.88
	Processed/more perishable	White buns	58832	181.57	176.12	0.00	1024.31
Legumes and nuts (*Less nutrient-dense*)	Unprocessed/less perishable	Brown beans dried	79826	161.25	168.98	0.00	1686.99
	Unprocessed/less perishable	Shelled groundnuts	67104	176.60	177.22	0.00	1879.23
	Unprocessed/less perishable	White beans dried	60546	138.08	140.09	0.00	1584.01
Animal source foods (*More* nutrient-*dense*)	Unprocessed/more perishable	Beef	74,127	197.02	196.83	0.00	1355.26
	Unprocessed/ more perishable	Eggs	71621	146.19	132.56	0.00	1552.87
	Processed/ more perishable	Ultra-pasteurized milk	62329	60.99	64.10	0.00	463.94
	Unprocessed/ more perishable	Goat meat	74588	164.14	179.22	0.00	1518.35
	Processed/ more perishable	Powdered milk	40868	1152.30	1178.31	0.00	6356.50
	Unprocessed more perishable	Usipa sun dried (Lake Malawi sardine)	78279	840.58	879.52	0.00	7938.16
	Unprocessed/ more perishable	Utaka dried (Lake Malawi cichlid)	56715	954.83	998.71	0.00	9029.67
Vegetables (*More nutrient-dense*)	Unprocessed/more perishable	Cabbage	72133	35.18	34.31	0.00	401.14
	Unprocessed/more perishable	Fresh okra	53956	151.71	161.74	0.00	1285.56
	Unprocessed/more perishable	Fresh onions	80090	162.75	147.99	0.00	1434.18
	Unprocessed/more perishable	Fresh pumpkin leaves (*Nkhwani*)	64589	101.92	104.22	0.00	827.87
	Unprocessed/more perishable	Rape leaves *(Tanapusi)*	74680	66.58	75.32	0.00	735.61
Fruits (*More* nutrient-*dense*)	Unprocessed/ more perishable	Bananas	66053	64.77	68.93	0.00	691.67
	Unprocessed/ more perishable	Fresh tomatoes	80558	116.72	122.43	0.00	975.34

Notes: Animal source foods, vegetables, and fruits are more nutrient-dense than legumes and nuts, and staples including roots and tubers. Overall, most of our more nutrient-dense foods under consideration are more perishable and unprocessed.

Table 2.2 shows the descriptive statistics for the variables used in the analysis. The average diesel price is MWK570.07 with 274.15 standard deviation. Our measure of climatic shocks shows that about 10 percent of both market pairs experienced floods while about 20 percent of the market pairs had one market that experienced floods over the period under investigation. Similarly, about 10 percent of both market pairs experienced droughts while about 21 percent of the market pairs had one market that experienced droughts. This may suggest that the locations in which the markets operate have been drier over the period we investigate, on average. On average, the population density is about 171 persons per square Km in the district the markets operate. The average distance between market pairs is 380 Km with standard deviation of 238 Km.

Table 2.2: Summary statistics for variables used in the analysis

Variable	Mean	Std. dev.	Min	Max
Diesel price	570.07	274.15	178.7	990.4
Floods in both market pair	0.092	0.289	0	1
Floods in one market pair	0.196	0.397	0	1
Drought in both market pair	0.094	0.292	0	1
Drought in one market pair	0.209	0.406	0	1
Population density in district market operates	171.41	99.15	34.05	488.49
Distance between market pairs	379.62	237.59	33	1098

2.5 Empirical results

2.5.1 The association between transport costs and price differences in the short run

2.5.1.1 The association between transport costs and overall price differences

We first estimate the influence of changes in transport costs on overall price differentials before examining the association between transport costs and price differentials for each food item (*i.e.,* food-item specific effects). Table 2.3 presents the results of the association between changes in transport costs and overall price differentials in the short run. Contrary to our expectation, table 2.3 shows that the relationship between transport costs and overall price differentials is negative and significant at 95% confidence interval (column 1, row 1). Thus, the increase in fuel price is associated with the reduction in overall price differentials for various foods under investigation in the short run by -4.8% (= 100[exp(-0.0493)-1]), on average. While this finding contrasts with our expectation, this finding suggests that market traders supply various foods efficiently as transport costs between markets increase, *ceteris paribus*. Thus, the increase in transport costs is associated with overall price convergence or better market integration for various food under investigation in the short run. However, this finding may be due to improvement in trucking competition, which lowers transport costs despite the increase in fuel prices over time (Competition and Fair Trading Commission, 2016; Kunaka et al., 2018; Lall et al., 2009).

Table 2.3: Heterogenous effects of transport costs on price differentials by type of food

Dependent variable: $\lvert P_{xt} - P_{yt} \rvert$	**All foods**	**Processed foods**	**Perishable foods**	**Nutrient-dense foods**
	(1)	**(2)**	**(3)**	**(4)**
Diesel price x distance	-0.0493***			
	(0.0130)			
Diesel price x distance x processed		-0.0776**		
		(0.0338)		
Diesel price x distance x perishable			0.0319***	
			(0.00938)	
Diesel price x distance x nutrient - dense				0.0355***
				(0.00931)
Distance-specific linear time trends	0.00818	0.0160	0.0491**	0.0498**
	(0.0213)	(0.0254)	(0.0235)	(0.0221)
Distance-specific quadratic time trends	-0.00000807	-0.0000215	-0.0000742**	-0.0000751**
	(0.0000328)	(0.0000392)	(0.0000365)	(0.0000342)
Distance-specific cubic time trends	2.20e-09	9.66e-09	3.73e-08**	3.77e-08**
	(1.68e-08)	(2.00e-08)	(1.87e-08)	(1.75e-08)
Difference in population density	-0.0214	-0.0196	-0.0215	-0.0215
	(0.0521)	(0.0519)	(0.0522)	(0.0522)
=1 if one of the markets experienced flood shocks	-0.00333	-0.00291	-0.00174	-0.00171
	(0.00449)	(0.00458)	(0.00458)	(0.00461)
=1 if one of the markets experienced drought shocks	-0.00164	-0.00134	-0.000588	-0.000690
	(0.00652)	(0.00672)	(0.00682)	(0.00684)
Difference in local production	-0.0325	-0.0269	-0.0329	-0.0333
	(0.0605)	(0.0621)	(0.0595)	(0.0595)
Month FE	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
Food-market-pair FE	Yes	Yes	Yes	Yes
N	1751528	1751528	1751528	1751528

Note: Observations in each column are from the full sample. Poisson pseudo-likelihood regression with multiple levels of fixed effects (ppmlhdfe) estimator results for each food items are presented in each column. The dependent variable is absolute value of price difference. Route distance over paved road between the market pairs is measured in hundreds of kilometres, and diesel fuel price in thousands of Malawian Kwacha. Dyadic clustered standard errors at the market in parentheses * p<0.10, ** p<0.05, *** p<0.010

Then, we investigate the source of the counterintuitive influence of the changes in transport costs on overall price differentials in the short run by estimating whether processed foods, perishable foods, or nutrient-dense foods (*i.e.*, foods with high concentration of essential vitamins and minerals) modify the influence of transport costs on overall price dispersion of various foods across markets in the short run. We interact our measure of transport costs with a dummy variable that takes on a value of one when the food item is processed (perishable or nutrient-dense in other specifications). Starting with processed versus unprocessed foods, we find that the association between transport costs and overall price differences is negative and significant at 95% confidence interval (column 2, row 2). Thus, the increase in fuel price is associated with a reduction in overall price differentials for processed foods compared to unprocessed foods in the short run by -7.5% (= 100[exp(-0.0776)-1]), on average. This finding suggests that market traders supply processed foods more efficiently than unprocessed foods as transport costs between markets increase, indicating that processed foods are more integrated than unprocessed foods. This makes sense given that processed foods are easier to aggregate (*i.e.*, lower search costs) and transport in large volumes over longer distances compared to unprocessed foods.

Turning to perishability (column 3, row 3), we find that the increase in fuel prices is associated with the increase in price differentials for more perishable foods compared to less perishable foods by 3.2% (= 100[exp(0.0319)-1]) at 95% confidence interval. Although more perishable foods have higher value to weight ratios, we attribute this finding to the difficulty to aggregate (*i.e.*, higher search costs) and transport them in large volumes over longer distances due to unavailability of cooled transport system in Malawi. This finding suggests that there is poor price convergence or market integration for more perishable foods than for less perishable foods as transport costs between markets increase.

Moving on to nutrient-dense versus less nutrient dense foods (column 4, row 4), we find that the increase in fuel prices is associated with an increase in price differentials for more nutrient-dense foods compared to less nutrient-dense foods across markets by 3.6% (= 100[exp(0.0355)-1]). Thus, market traders do not supply more nutrient-dense foods efficiently as transport costs between markets increase, *ceteris paribus*. This finding suggests that there is poor price convergence or market integration for more nutrient-dense foods than for less nutrient-dense foods as transport costs between markets increase, which has nutritional implications.

In summary, this sub-section has shown that the changes in transport costs are associated with a decline in overall price differences across markets in the short run. This counterintuitive influence is driven by processed foods but not by perishable foods and nutrient-dense foods for which the changes in transport costs are associated with an increase in overall price differences across markets in the short run. The next two sub-sections examine (i) the association between changes in transport costs and price differentials for each food item, and (ii) whether the magnitude of the association between transport costs and price differences for each food item are smaller for market pairs that are closer to each other than market pairs that are far from each other in the short run.

2.5.1.2 The association between transport costs and price differences for each food item

This sub-section estimates the association between changes in transport costs and price differentials for each food item to determine the source of the counterintuitive influence of the changes in transport costs on price differentials in the short run. The results are presented by food groups (*i.e*., staples including roots and tubers, animal source foods, legumes and nuts, vegetables, and fruits).

2.5.1.2.1 Staples including roots and tubers

Figure 2.2 presents the point estimates along with their confidence intervals of the association between transport costs and price dispersion of staples including roots and tubers across markets. In figure 2.2, the marker represents estimated coefficient of our measure of transport costs for each food item, the spike represents the confidence interval at 95% level, and the vertical line along the x-axis at zero represents a reference line that indicates whether estimated coefficient is significantly different from zero or not (Jann, 2014). The estimated coefficient is not significantly different from zero if the spike (*i.e.*, confidence interval) touches or crosses the reference line. The direction of the association between transport costs and price differences remains negative for cassava, rice grain, sweet potatoes, and white bread in the short run. However, the relationship is significant only for rice grain at 95% confidence interval. Thus, the increase in fuel price is associated with the decrease in price differentials for rice grain across markets by -9.7% (= 100[exp(-0.102)-1]) (note that the size of the influence in brackets is consistent with table A.2 in the appendix).[19,20] To illustrate, an increase

[19] Figure A.9 in the appendix presents the time dummies after fixed effects Poisson estimator. Overall, the results indicate that the price differences have been increasing over time for staples.
[20] We find that the direction of the association between transport costs and price differences for staples remain the same using linear regression fixed effect estimator (see table A.7 in the appendix).

in distance by 50 Km and fuel price by MWK100 is associated with the decline in price differences for rice grain by MWK-0.46/kg (=-0.097x94.04x0.5x0.1 where 94.04 is average price difference for rice grain). We attribute this finding to high value to weight ratio that provide an incentive for trade and low search costs through rice millers along the main roads in high rice production areas that allows traders to organise large volumes (African Institute of Corporate Citizenship, 2016). Conversely, we find that the direction of the association between transport costs and price differences becomes positive for maize flour dehulled, maize grain (private), and maize grain (ADMARC), potatoes, and white buns in the short run. However, these relationships are significant only for maize flour dehulled, maize grain (private), and maize grain (ADMARC) at 95% confidence interval. Thus, the increase in fuel price is associated with the increase in price differentials for these food items across markets, *ceteris paribus.* To illustrate, an increase in distance by 50 Km and fuel price by MWK100 is associated with the increase in price differences for maize flour dehulled, maize grain (private), and maize grain (ADMARC) by MWK0.56/kg (=0.078x142.72x0.5x0.1 where 142.72 is average price difference of maize flour dehulled), MWK0.17/kg (=0.079x43x0.5x0.1 where 43 is average price difference of maize grain) and MWK0.57/kg (=0.406x28x0.5x0.1 where 28 is average price difference of maize grain (ADMARC)) across markets, respectively. Thus, market traders do not supply more of maize flour dehulled, maize grain (private), and maize grain (ADMARC) produced outside the districts the markets operate as transport costs increase, which is associated with the increase in price differences for these foods across markets consistent with the previous literature (Aker, 2010a, 2010b; Aker et al., 2014; Fuje, 2019, 2020; Salazar et al., 2019). This is a sign that markets for staple maize grain are poorly integrated to allow traders to transport larger volumes that lower transport costs per unit across markets. We attribute these findings to higher search costs given that small-scale farmers are geographically dispersed and produce a limited surplus, which makes it harder for traders to aggregate large volumes (Burke et al., 2020; Ochieng et al., 2019).

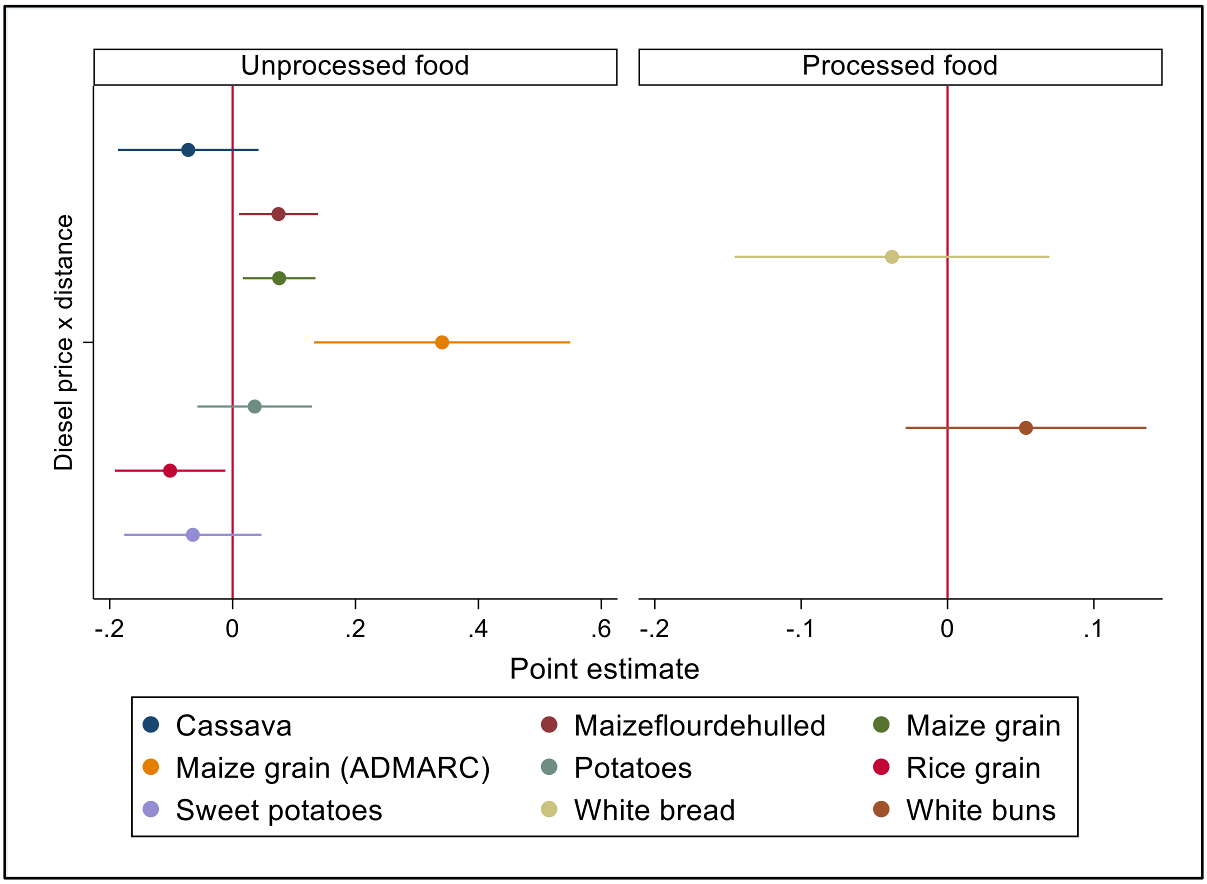

Figure 2.2: The association between transport costs and market price dispersion of staples including roots and tubers

The association between transport costs and price differences for maize flour dehulled, maize grain (private), and maize grain (ADMARC) remains positive and significant when we control for market-specific seasonality at 95% confidence interval (see table A.12 in the appendix). When the missing price differences are mapped to their maximum price differences in each year, we find similar results of the association between transport costs and price differences for maize flour dehulled, maize grain (private) and maize grain (ADMARC) at 95% confidence interval (see table A.17 in the appendix). However, the association between transport costs and price differences for cassava becomes significant at 95% confidence interval. Controlling for non-linear effects in distance, we find that the direction of the association between transport costs and price differences remains the same for most staple foods. However, the direction of the relationship flips signs for potatoes, white bread, and white buns. The association between transport costs and price differences diminishes for cassava and white buns at 95% confidence interval (see table A.22 in the appendix). To illustrate, this finding means that the first kilometre is associated with the decrease in price differences for cassava by MWK-7.8/kg (=100[exp(-0.178)-1]x0.01x48 where -0.178 is consistent with table A.22 and 48 is average price differences), and the second kilometre is associated with the decrease in price differences by MWK-6.9/kg (= (100[exp(-0.178)-1]x0.01x48) + 2(100[exp(0.009)-1]x0.01x48)), and so on. The same applies to white buns.

Accounting for the interaction of distance with lagged fuel price as an additional covariate, we find that the direction of the contemporaneous association between transport costs and price differences remains the same for cassava, maize flour dehulled, maize grain (ADMARC), potatoes, white bread, and white buns (see table A.27 in the appendix). However, the contemporaneous association between transport costs and price differences is significant only for maize grain (private) at 95% confidence interval. This finding means that the association between transport costs and price differences for maize grain (private) is much stronger in the month before the fuel price change. Further, we find that the direction of the lagged term of transport costs is positive for most staple foods, except for maize flour dehulled, rice grain, sweet potatoes, and white bread.

2.5.1.2.2 Legumes and nuts

Figure 2.3 presents the point estimates along with their confidence intervals of the association between transport costs and the price dispersion of legumes and nuts across markets (see table A.3 in the appendix for the magnitude of the estimated coefficients).[21,22] We find that the direction of the association between transport costs and price differences remains negative for groundnuts but becomes positive for brown and white beans. However, the relationship is significant only for brown beans at 95% confidence interval. Similarly, the increase in fuel price is associated with an increase in price differences across markets for brown beans by 9.2% (= 100[exp(0.0884)-1]) (note that the size of the influence in brackets is consistent with table A.3 in the appendix, *ceteris paribus*. Thus, market traders do not supply more of brown beans produced outside the districts the markets operate as transport costs increase, which is associated with the increase in price differences for these foods across markets. We attribute this finding to high search costs of aggregating the food item from small-

[21] Figure A.10 in the appendix presents the time dummies after fixed effects Poisson estimator. Overall, the results indicate that the price differences have been increasing over time for legumes and nuts.
[22] Overall, we find similar results using linear regression fixed effects estimator (see table A.8 in the appendix).

scale farmers. Similarly, this means that the markets for legumes and nuts are not largely well integrated.

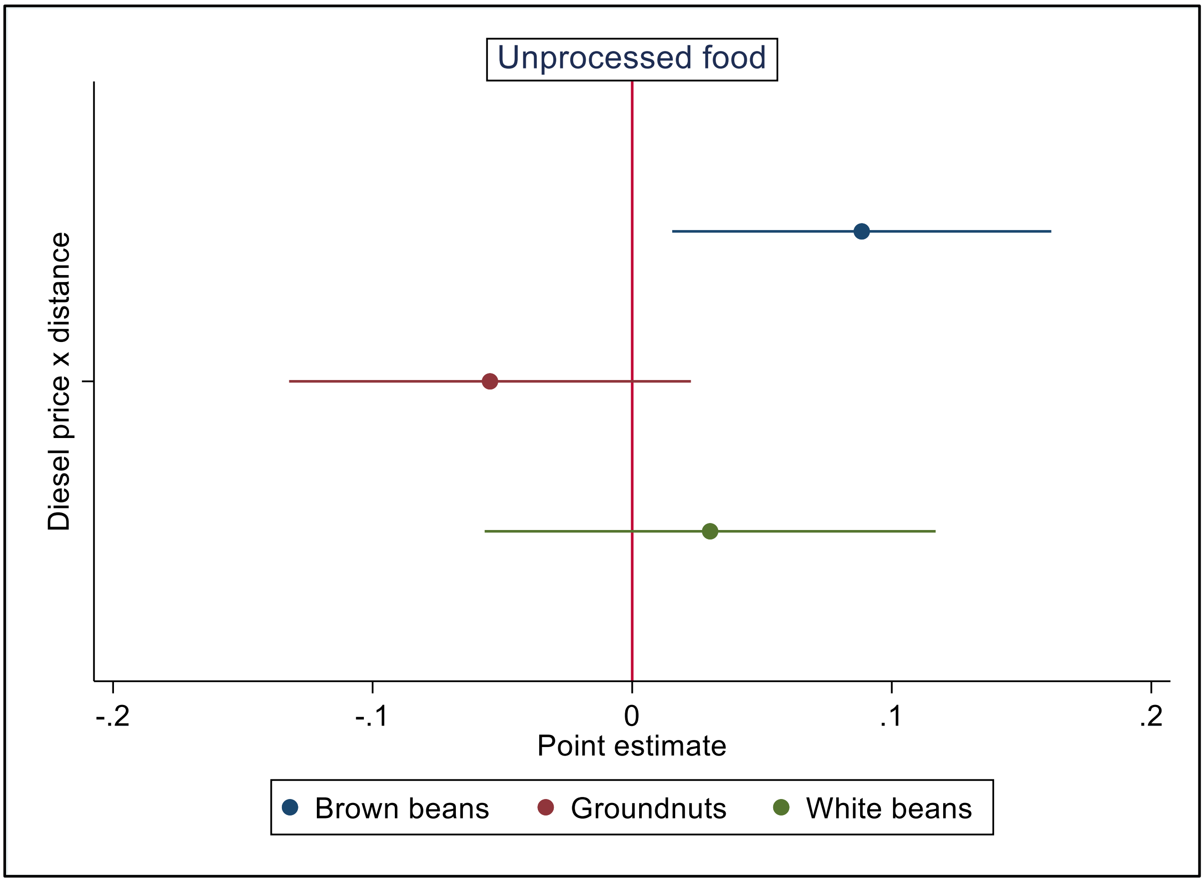

Figure 2.3: The association between transport costs and market price dispersion of legumes and nuts

The association between transport costs and price differences for legumes and nuts remains the same when we control for market-specific seasonality at 95% confidence interval (see table A.13 in the appendix). Further, the direction of the association between transport costs and price differences for brown beans and groundnuts remain the same, while for white beans becomes negative when missing price differences are mapped to their maximum price differences in each year (see table A.18 in the appendix). The direction of the association between transport costs and price differences for legumes and nuts remains the same while controlling for non-linear effects in distance (see table A.23 in the appendix). However, we find that association between transport costs and price differences does do not diminish for legumes and nuts at 95% confidence interval.

The direction of the contemporaneous association between transport costs and price differences for legumes and nuts remains positive when we account for the interaction of distance with lagged fuel price as the additional covariate (see table A.28 in the appendix). However, the contemporaneous association is significant for brown and white beans at 95% confidence interval. The lagged term is negative and significant at 95% confidence interval for white beans, but the magnitude of the association is larger in the contemporaneous term than in the lagged term. This finding suggests that the association between transport costs and price differences for white beans is much stronger in the month of fuel price change.

2.5.1.2.3 Animal source foods

Figure 2.4 presents the point estimates along with their confidence intervals of the association between transport costs and the price dispersion of animal source foods across markets (see table A.4 in the appendix for the magnitude of the estimated coefficients).[23,24] Starting with unprocessed food, figure 2.4 shows that the direction of the association between transport costs and price differences remains negative for beef, goat meat, and utaka but becomes positive for eggs and usipa. However, the relationships are significant only for beef, eggs, and goat meat at 95% confidence interval. Thus, the increase in transport costs is associated with the reduction in price differentials for beef by -8.9% (=100[exp(-0.0937)-1]) and goat meat by -10.6% (= 100[exp(0.112)-1]) in the short run at 95% confidence interval (note that the size of the effects in brackets are consistent with table A.4 in the appendix). Conversely, the increase in transport costs is associated with the increase in price differentials for eggs by 12.4% (= 100[exp(0.117)-1]) in the short run at 95% confidence interval. Since live animals and eggs are less traded over longer distances due to various difficulties to transport them, there may be other transactions costs that may not be reflected in transport costs that restrict trade for eggs such as perishability and promote trade for beef and goat meat such as lower search costs.

[23] Figure A.11 in the appendix presents the time dummies after fixed effects Poisson estimator. Overall, the results indicate that the price differences have been increasing over time for animal source foods.
[24] We find similar results of the direction of the association between transport costs and price differences for animal source foods using linear fixed effects estimator (see table A.9 in the appendix).

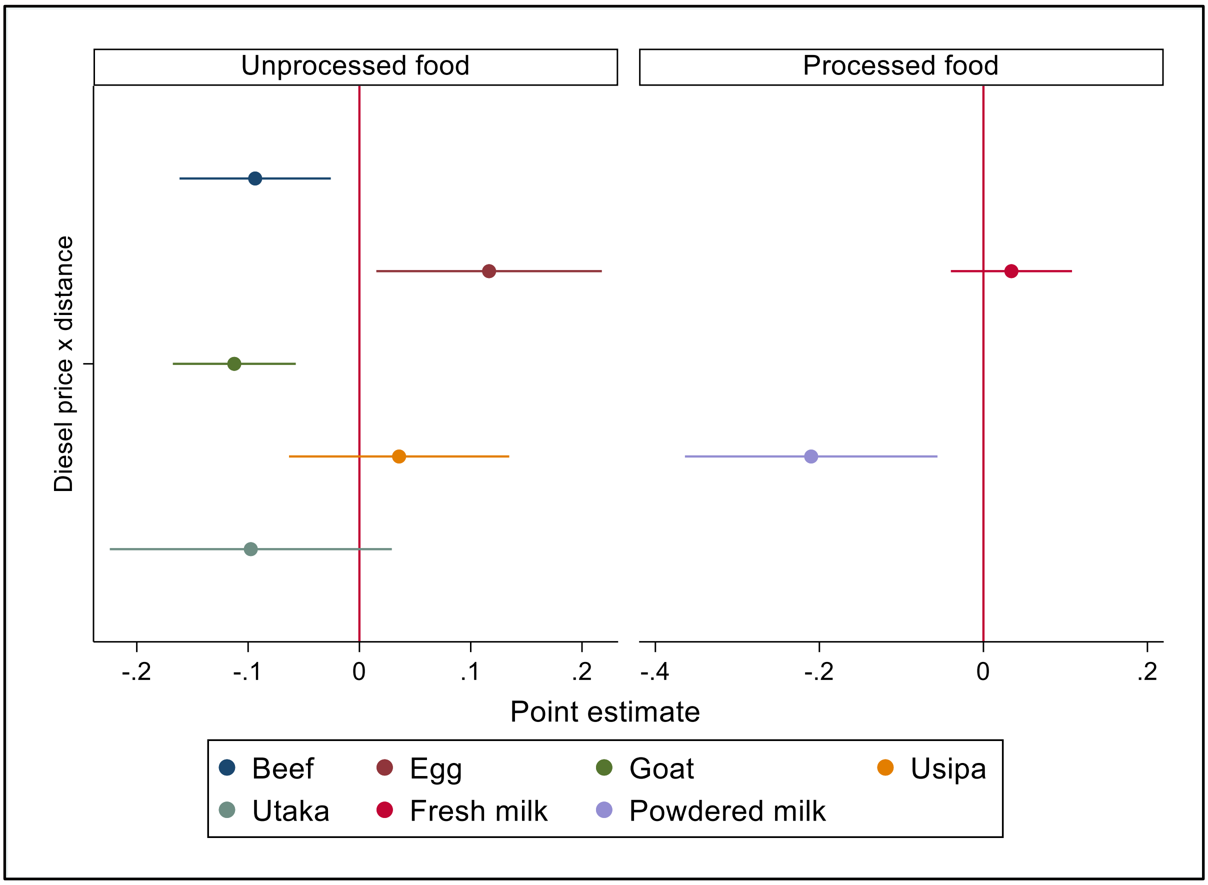

Figure 2.4: The association between transport costs and market price dispersion of animal source foods

Turning to processed animal source foods, we find that the association between transport costs and price differences remains negative for powdered milk but becomes positive for Ultra- pasteurized milk. However, the relationship is significant only for powdered milk at 95% confidence interval. Thus, the increase in transport costs is associated with the decline in price differentials for powdered milk by -18.9% (=100[exp(-0.21)-1] where -0.21 is consistent with table A.4 in the appendix) at 95% confidence interval, on average. This means that the increase in fuel price with distance is not associated with trade restriction for powdered milk across markets, *ceteris paribus*. We attribute this negative association between transport costs and price differences for powdered milk to their product nature (*i.e.*, being processed and packaged), which allows traders to organise (*i.e.*, lower search costs) and transport larger volumes between markets thereby lowering transport costs per unit volume.

The association between transport costs and price differences for animal source foods remains the same when we control for market-specific seasonality at 95% confidence interval (see table A.14 in the appendix). Further, when the missing price differences are mapped to their maximum price differences in each year, the association between transport costs and price differences for animal source foods remains the same at 95% confidence interval (see table A.19 in the appendix). Controlling for non-linear effects in distance, we find that the direction of the association between transport costs and price differences for animal source foods remain the same (see table A.24 in the appendix). We find significant marginal diminishing relationship between transport costs and price differences for beef, eggs, and goat meat at 95% confidence interval.

The direction of the association between transport costs and price differences for most animal source foods remain the same, except for ultra-pasteurized milk when we account for the interaction of distance with lagged fuel price as the additional covariate (see table A.29 in the appendix). The negative contemporaneous association between transport costs and price differences for goat meat remains significant at 95% confidence interval. The lagged term is negative and significant for beef and powdered milk at 95% confidence interval. This finding suggests that the association between transport costs and price differences for beef and powdered milk is much larger in the month before the change in fuel price.

2.5.1.2.4 Vegetables

Figure 2.5 presents the point estimates along with their confidence intervals of the association between transport costs and the price dispersion of vegetables across markets (see table A.5 in the appendix for the magnitude of the estimated coefficients).[25,26] We find that the direction of the association between transport costs and price differences remains negative for cabbage and rape leaves but becomes positive for okra, onions, and pumpkin leaves. However, the associations are insignificant at 95% confidence interval for vegetable foods under

[25] Figure A.12 in the appendix presents the time dummies after fixed effects Poisson estimator. Overall, the results indicate that the price differences have been increasing over time for vegetables.
[26] We find that the direction of the association between transport costs and price differentials for vegetables are similar using the linear regression fixed effects estimator, except that the sign flips for rape leaves (see table A.10 in the appendix).

consideration since these foods are highly perishable; hence, less traded over longer distances across markets in the absence of cooled transportation system.

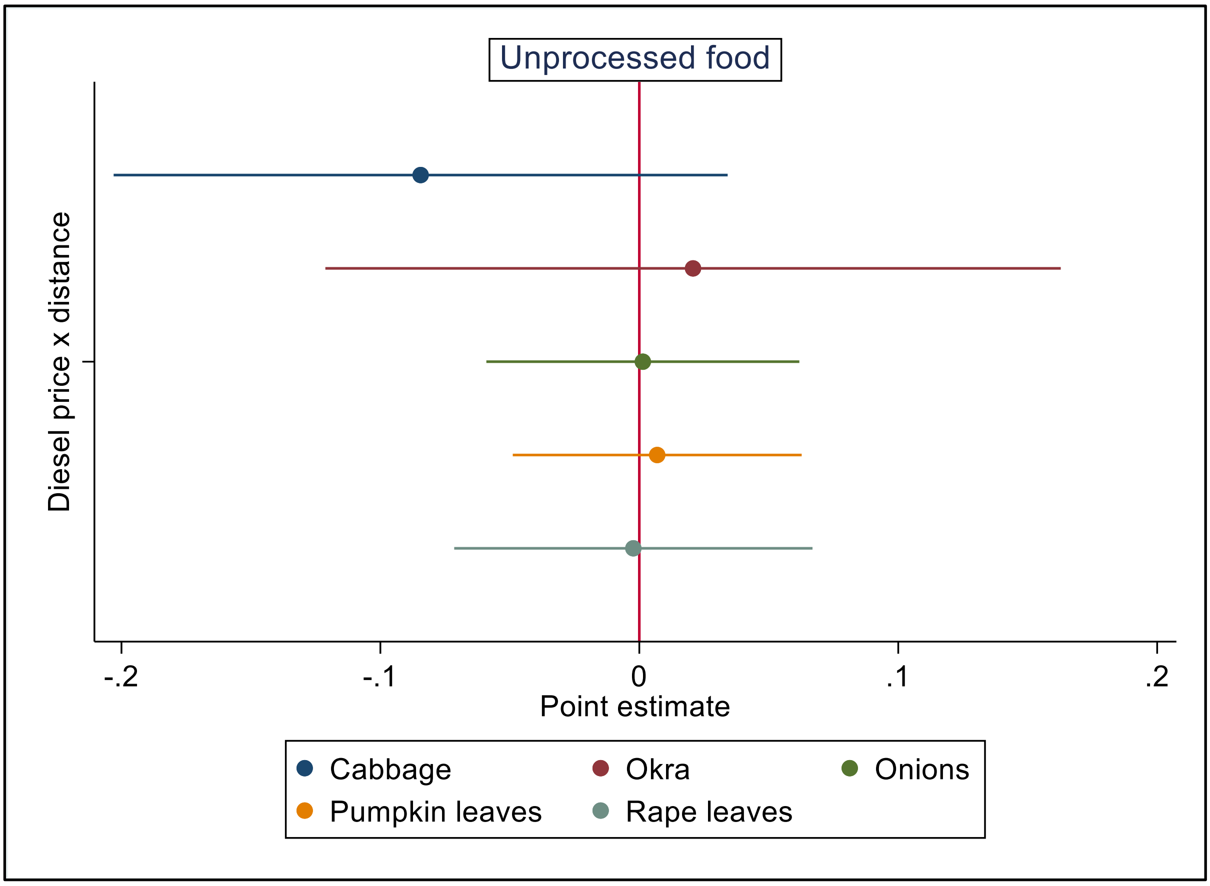

Figure 2.5: The association between transport costs and market price dispersion of vegetables

The association between transport costs and price differences remains the same for vegetables when we control for market-specific seasonality at 95% confidence interval, except that the direction of the relationship flips sign for rape leaves (see table A.15 in the appendix). We find similar results of the association between transport costs and price differences for rape leaves when the missing price differences are mapped to their maximum price differences in each year, except that the direction of the relationship flips sign and becomes significant for pumpkin leaves (see table A.20 in the appendix). Controlling for non-linear effects in distance, the direction of the association between transport costs and price differences for vegetables flips sign for okra, onions, and rape leaves (see table A.25 in the appendix). However, we find that there is significant diminishing association between transport costs and price differences for okra at 95% level. Accounting for the interaction of distance with lagged fuel price as the additional covariate, we find that both the contemporaneous and lagged terms are insignificant for vegetables at 95% confidence interval (see table A.30 in the appendix).

2.5.1.2.5 Fruits

Figure 2.6 presents the point estimates along with their confidence intervals of the association between transport costs and the price dispersion of fruits across markets (see table A.6 in the appendix for the magnitude of the estimated coefficients).[27,28] We find that the direction of the association between transport costs and price differences remains negative for both bananas and tomatoes in the short run. However, the relationship is significant only for bananas at 95% confidence interval. This means that increase in transport costs is associated with the reduction in price differentials for bananas by -7.9% (=100[exp(-0.082)-1] where -0.082 is consistent with table A.6 in the appendix), on average. Thus, increase in fuel price with distance is not associated with trade restriction for bananas since traders use empty backhauls or open trucks to transport them across markets, *ceteris paribus*. We attribute this finding to lower search costs given that over 90 percent of bananas traded across markets are imported.

[27] Figure A.13 in the appendix presents the time dummies after fixed effects Poisson estimator. Overall, the results indicate that the price differences have been increasing over time for fruits.
[28] We find similar results of the direction of the association between transport costs and price differentials using linear regression fixed effects estimator (see table A.11 in the appendix).

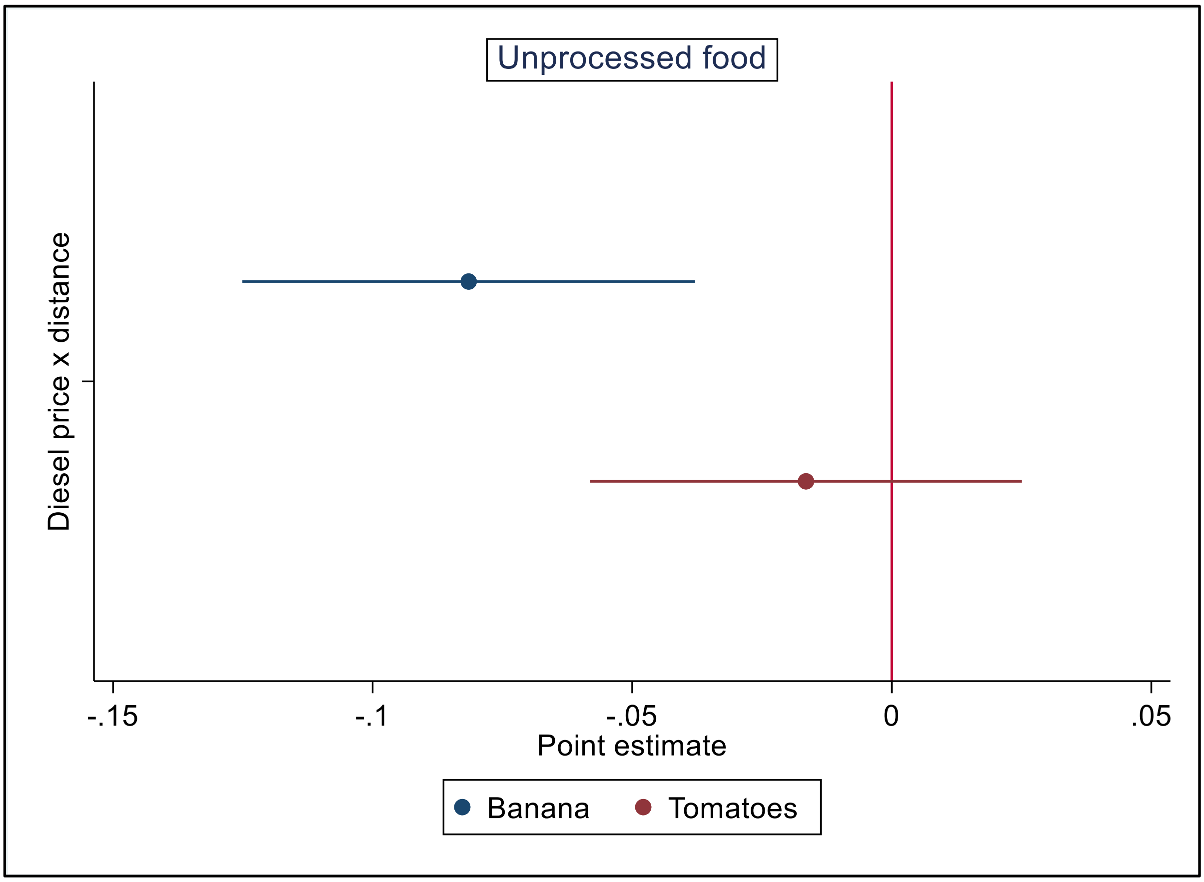

Figure 2.6: The association between transport costs and market price dispersion of fruits

The association between transport costs and price differences for bananas and tomatoes remains the same when we control for market-specific seasonality at 95% confidence interval (see table A.16 in the appendix). We find similar results of the association between transport costs and price differences for bananas and tomatoes when the missing price differences are mapped to their maximum price differences in each year (see table A.21 in the appendix). Further, we find significant diminishing effect of the association between transport costs and price differences for bananas while controlling for non-linear effects in distance (see table A.26 in the appendix). We find similar results of the contemporaneous association between transport costs and price differences for bananas and tomatoes at 95% confidence interval when we account for the interaction of distance with lagged fuel price as an additional covariate (see table A.31 in the appendix). However, the lagged term is positive and significant for bananas at 95% level, suggesting that the magnitude of the association between transport costs and price differentials is larger in the month before the fuel price change.

In summary, this sub-section has shown that the association between the changes in transport costs and price differences remains negative and significant for some foods such as rice grain, beef, goat meat, powdered milk, and bananas but it becomes positive and significant for other foods such as maize flour dehulled, maize grain (private), maize grain (ADMARC), brown beans, and eggs in the short run. These findings, further, suggest that the counterintuitive influence of the changes in transport costs on overall price differentials is driven by food items that have lower search costs and are easy to aggregate. This may suggest that the increase in transport costs is associated with the shift in trade from other foods to those foods that have lower search costs and are easier to organise across markets.

2.5.1.3 Spatial heterogeneity in the short run

This sub-section examines whether the magnitude of the association between transport costs and price differences for each food item is smaller for market pairs that are closer to each other than market pairs that are far from each other (*i.e.*, spatial heterogeneity of the association between transport costs and price differences) in the short run.

2.5.1.3.1 Staples including roots and tubers

Figure 2.7 presents the point estimates along with their confidence intervals of the association between transport costs and the price dispersion of staples including roots and tubers with varying distance levels across markets (see table A.32 in the appendix for the magnitude of the estimated coefficients). Figure 2.7 shows that the association between transport costs and price differences remains negative and significant for some distance levels for rice grain (152 – 200 km, 268 – 331 km, 394 – 479 km to 735 – 1098 km), while for white bread it remains negative but becomes significant for some distance levels (152 – 200 km, 201 – 267 km, and 601 – 734 km) compared to the base category of market pairs that are less than 102 km apart at 95% confidence interval. These findings suggest that the negative association between transport costs and price differences for rice grain and white bread is smaller for market pairs that are closer, *ceteris paribus*. Conversely, we find that the association between transport costs and price differences remains positive and significant for some distance levels for maize flour dehulled (152 – 200 km and 268 – 331 km) and maize grain (ADMARC, 735 – 1098 km) compared to the base category of market pairs that are less than 102 km apart at 95%

confidence interval. Thus, the positive association between transport costs and price differences for maize flour dehulled and maize grain (ADMARC) is smaller for market pairs that are closer, *ceteris paribus*. However, the association between transport costs and price differences for maize grain (private) becomes negative and significant for distance between 102 – 151 km compared to the base category of market pairs that are less than 102 km apart at 95% confidence interval. This finding suggests that private traders trade maize grain efficiently between markets that are closer.

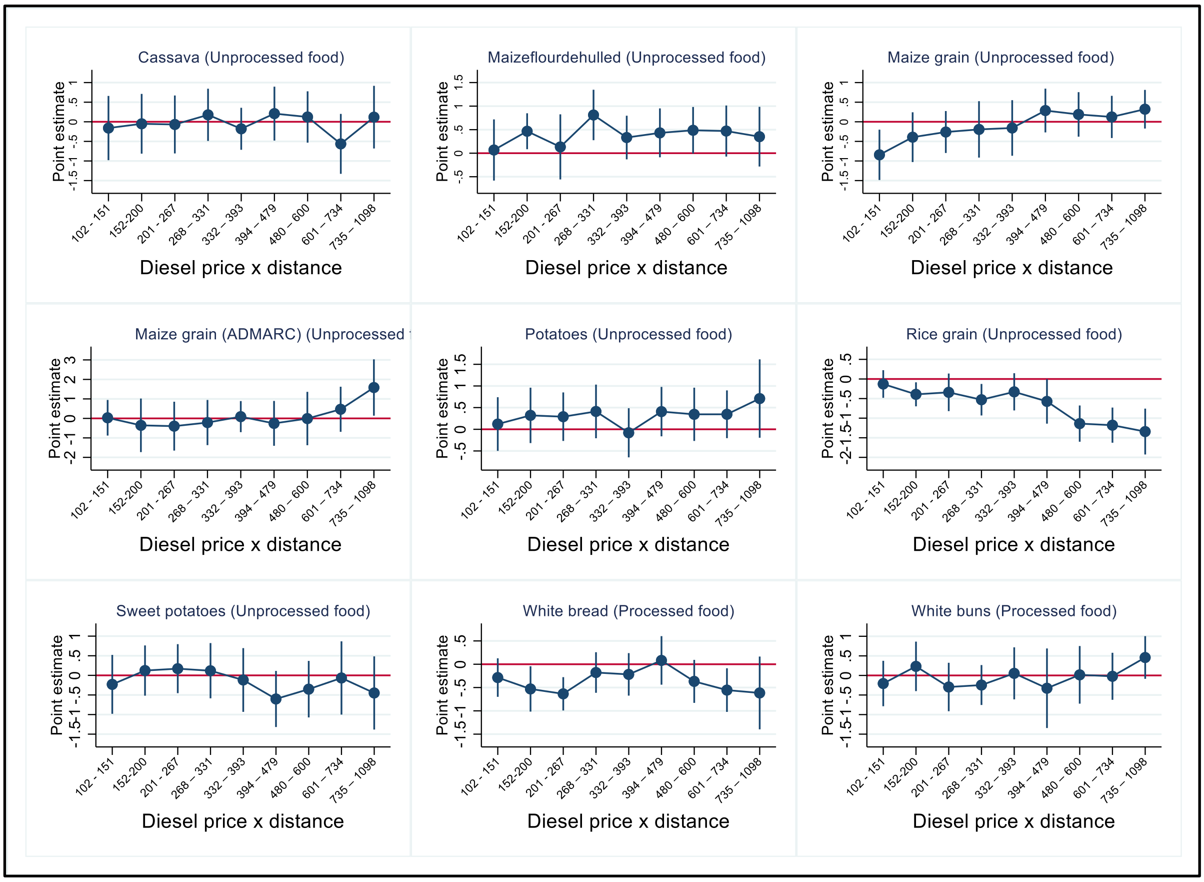

Figure 2.7: The association between transport costs and market price dispersion of staples including roots and tubers at various distance levels

2.5.1.3.2 Legumes and nuts

Figure 2.8 presents the point estimates along with their confidence intervals of the association between transport costs and the price dispersion of legumes and nuts with varying distance levels across markets (see table A.33 in the appendix for the magnitude of the estimated coefficients). Figure 2.8 shows that the association between transport costs and price differences remains positive and significant for some distance levels for brown beans (from 268 – 331 km), while for white beans it remains positive but becomes significant for some distance levels (from 102 – 151 km, 152 – 200 km, 268 – 331 km, 480 – 600 km, and 735 – 1098 km) compared to the base category of market pairs that are less than 102 km apart at 95% confidence interval. This suggests that the positive association between transport costs and price differences for brown and white beans is smaller for market pairs that are closer, *ceteris paribus*.

Figure 2.8: The association between transport costs and market price dispersion of legumes and nuts at various distance levels

2.5.1.3.3 Animal source foods

Figure 2.9 presents the point estimates along with their confidence intervals of the association between transport costs and the price dispersion of animal source foods with varying distance levels across markets (see table A.34 in the appendix for the magnitude of the estimated coefficients). Figure 2.9 shows that the association between transport costs and price differences remains positive and significant for some distance levels for eggs (152 – 200 km, 201 – 267 km, 394 – 479 km to 735 -1098 km), while for usipa it becomes significant for some distance levels (from 102 – 151 km to 268 – 331 km) compared to the base category of market pairs that are less than 102 km apart at 95% confidence interval. This suggests that the positive association between transport costs and price differences for eggs and usipa is smaller for market pairs that are closer, *ceteris paribus*. Conversely, we find that the association between transport costs and price differences remains negative and significant for some distance levels for beef (480 – 600 to 735 – 1098 km), goat meat (201 – 267 km, and 332 – 393 km to 735 – 1098 km), and powdered milk (102 – 151 km to 201 – 267 km, 394 - 479 km to 735 – 1098 km) compared to the base category of market pairs that are less than 102 km apart at 95% confidence interval. However, the association between transport costs and price differences for utaka becomes significant for distances between 735 and 1098 km compared to the base category of market pairs that are less than 102 km apart at 95% confidence interval. Similarly, these findings suggest that the negative association between transport costs and price differences for beef, goat meat and powdered milk is smaller for market pairs that are closer, while the finding for utaka may mean that the price difference across market pairs for utaka at such larger distance are simply random once saturation point or effect is reached (Roehner, 1996).

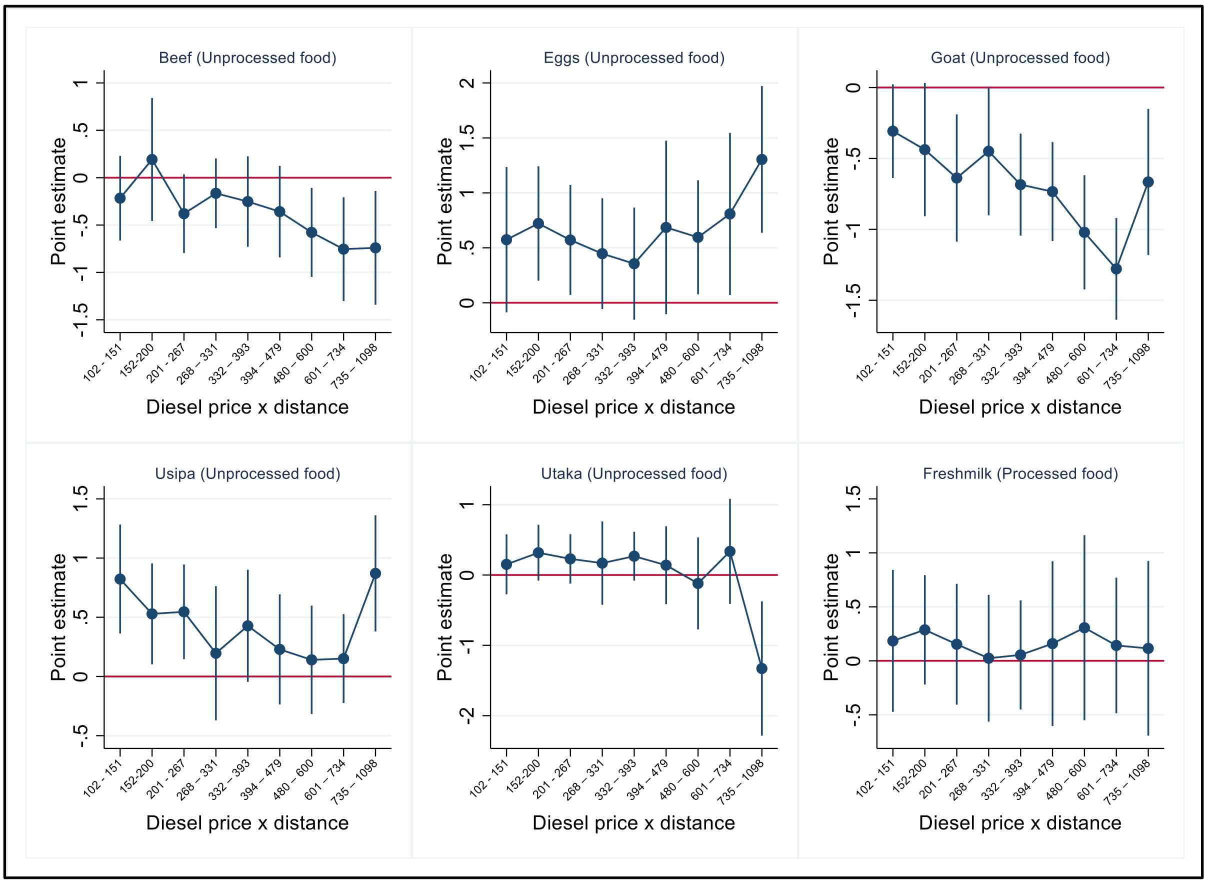

Figure 2.9: The association between transport costs and market price dispersion of animal source foods at various distance levels

2.5.1.3.4 Vegetables

Figure 2.10 presents the point estimates along with their confidence intervals of the association between transport costs and the price dispersion of vegetables with varying distance levels across markets (see table A.35 in the appendix for the magnitude of the estimated coefficients). We find that the association between transport costs and price differences for vegetables remain insignificant, except for cabbage (735 – 1098 km) compared to the base category of market pairs that are less than 102 km apart at 95% confidence interval. Similarly, this finding may mean that the price differences across market pairs for cabbage at such larger distance are simply random once saturation point or effect is reached (Roehner, 1996).

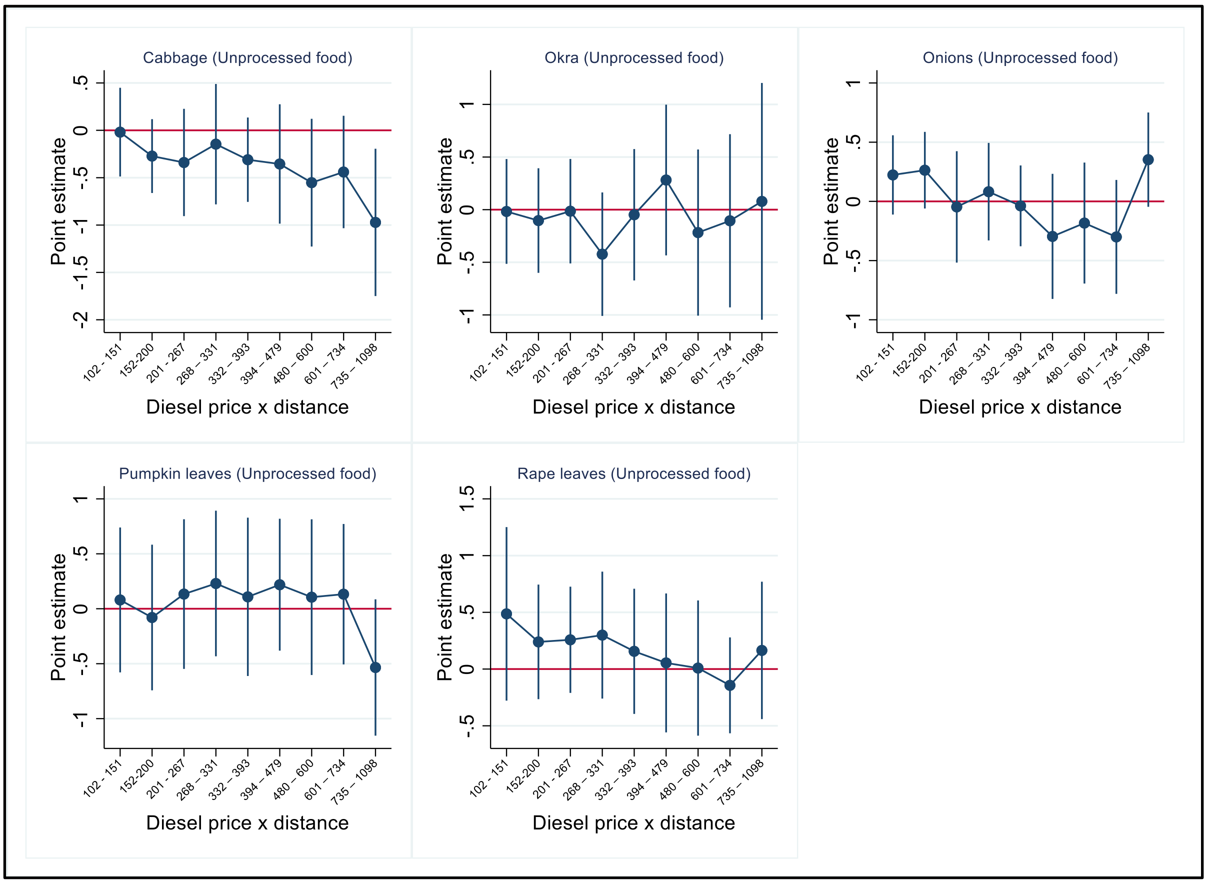

Figure 2.10: The association between transport costs and market price dispersion of vegetables at various distance levels

2.5.1.3.5 Fruits

Figure 2.11 presents the point estimates along with their confidence intervals of the association between transport costs and the price dispersion of fruits with varying distance levels across markets (see table A.36 in the appendix for the magnitude of the estimated coefficients). We find that the association between transport costs and price differences remains negative and significant for some distance levels for bananas at 601 – 734 km but it becomes significant for tomatoes at the distance of between 394 and 479 km compared to the base category of market pairs that are less than 102 km apart at 95% confidence interval. Thus, the negative association between transport costs and price differences for fruits is smaller for market pairs that are closer.

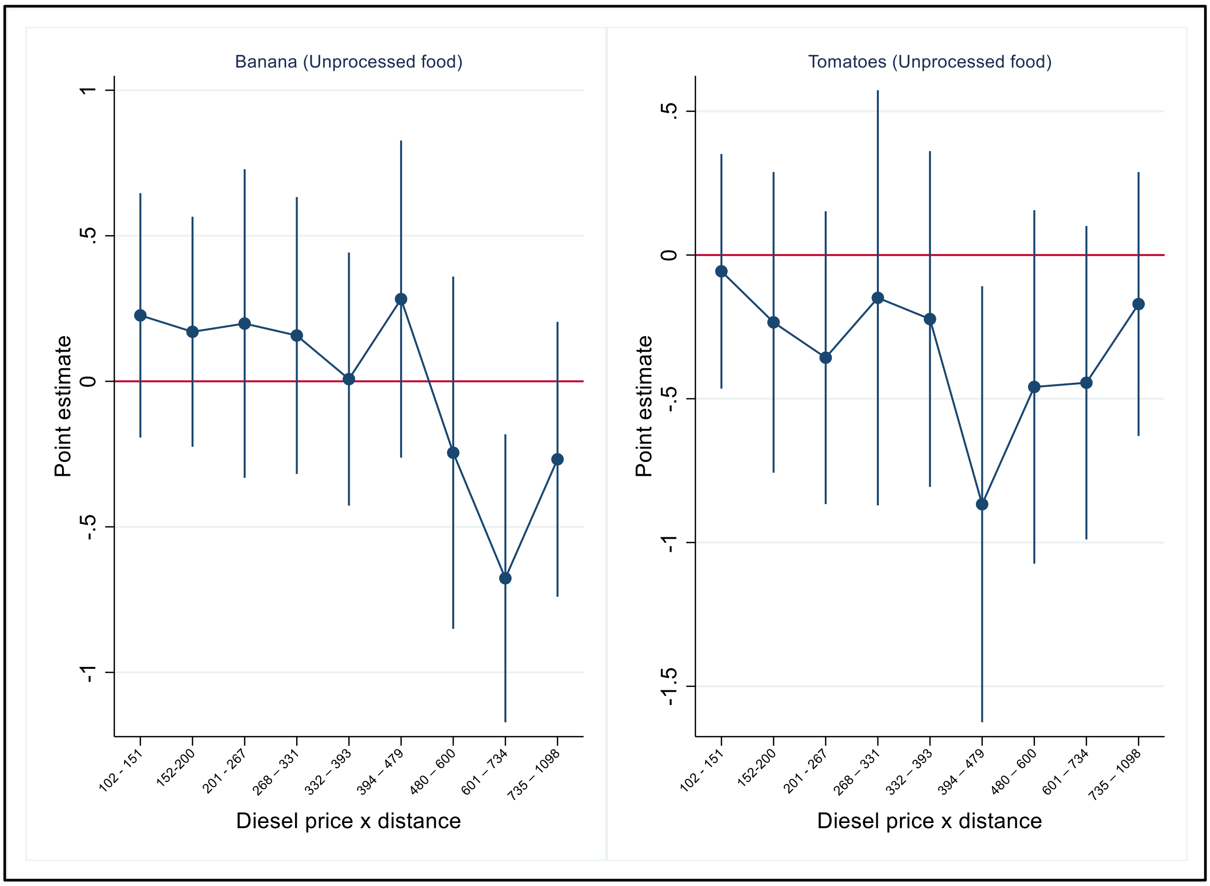

Figure 2.11: The association between transport costs and market price dispersion of fruits at various distance levels

In summary, this sub-section has shown that the negative association between the changes in transport costs and price differences remains significant at some distance levels for rice grain, beef, goat meat, powdered milk, and bananas but it becomes significant for some distance levels for white bread and tomatoes relative to the base category of market pairs that are less than 102 km apart. Conversely, the positive association between the changes in transport costs and price differences remains significant at some distance levels for maize flour dehulled, maize grain (ADMARC), brown beans, and eggs but it becomes significant for some distance levels for usipa and white beans relative to the base category of market pairs that are less than 102 km apart. However, the association between the changes in transport costs and price differences for maize grain (private) becomes negative and significant at smaller distance levels relative to the base category of market pairs that are less than 102 km apart, suggesting that private traders supply maize grain efficiently between markets that are closer. Overall, the magnitudes of the relationships are smaller for market pairs that are closer to each other for all foods under consideration.

2.5.2 The association between transport costs and price differences in the long run

This sub-section explores the association between changes in transport costs and price differentials for each food item in the long run.

2.5.2.1 Staples including roots and tubers

Figure 2.12 presents the point estimates along with their confidence intervals of the association between transport costs and the price dispersion of staples including roots and tubers while controlling for the lag of the price difference across markets (see table A.37 in the appendix for the magnitude of the estimated coefficients).[29] We find that the coefficient estimates for the lagged price differences are statistically significant at 95% confidence interval for both unprocessed and processed staple foods, but closer to zero in most cases. This means that the price differences in the previous period play a smaller role in predicting the price differences in the current period for most staple foods. For instance, the previous period price differences for cassava predicts about MWK0.32 (= [exp(0.0067)-1]x47.79, where 0.0067 is consistent with table A.38 in the appendix and 47.79 is average price difference) in the current period price differences. Thus, there is very low persistence in price differences over time consistent with the previous literature (Aker, 2010a; Salazar et al., 2019; Zant, 2018). Figure 2.12 shows that the association between transport costs and price differences remains positive and significant at 95% confidence interval for maize flour dehulled, maize grain (private), and maize grain (ADMARC) but it becomes significant for potatoes and white buns in the long run. Conversely, the association between transport costs and price differences becomes positive and significant for cassava, rice grain, sweet potatoes, and white bread at 95% confidence interval in the long run. These findings indicate that market traders to do not

[29] We find that the direction of the association between transport costs and price differences and lagged price differences remain the same for staple foods, except for sweet potatoes using Arellano–Bond estimator (see table A.42 in the appendix). However, the associations are not significant for maize grain (private), sweet potatoes, and white buns at 95% confidence interval. Further, the direction of the association between transport costs and price differences flips for some food such as cassava, rice grain, sweet potatoes, and white bread compared to results from the main specification. Since estimation with ivpoisson does not include time fixed effects, we attribute the change in the direction of the relationships to different functional forms.

supply these foods efficiently across markets as transport costs between markets increase. The long run positive associations of transport costs on price differences for staple foods are as follows: cassava (4.6% = [exp((0.042/1-0.0671)-1)*100]), maize flour dehulled (4.8%), maize grain (private, 2%), maize grain (ADMARC, 2.8%), potatoes (3.3%), rice grain (2%), sweet potatoes (2.3%), white bread (3.5%), and white buns (3.8%) (note that the size of the relationships in brackets are consistent with table A.38 in the appendix). Thus, the long-term relationship between transport costs and price differences for maize flour dehulled is large among staple foods.

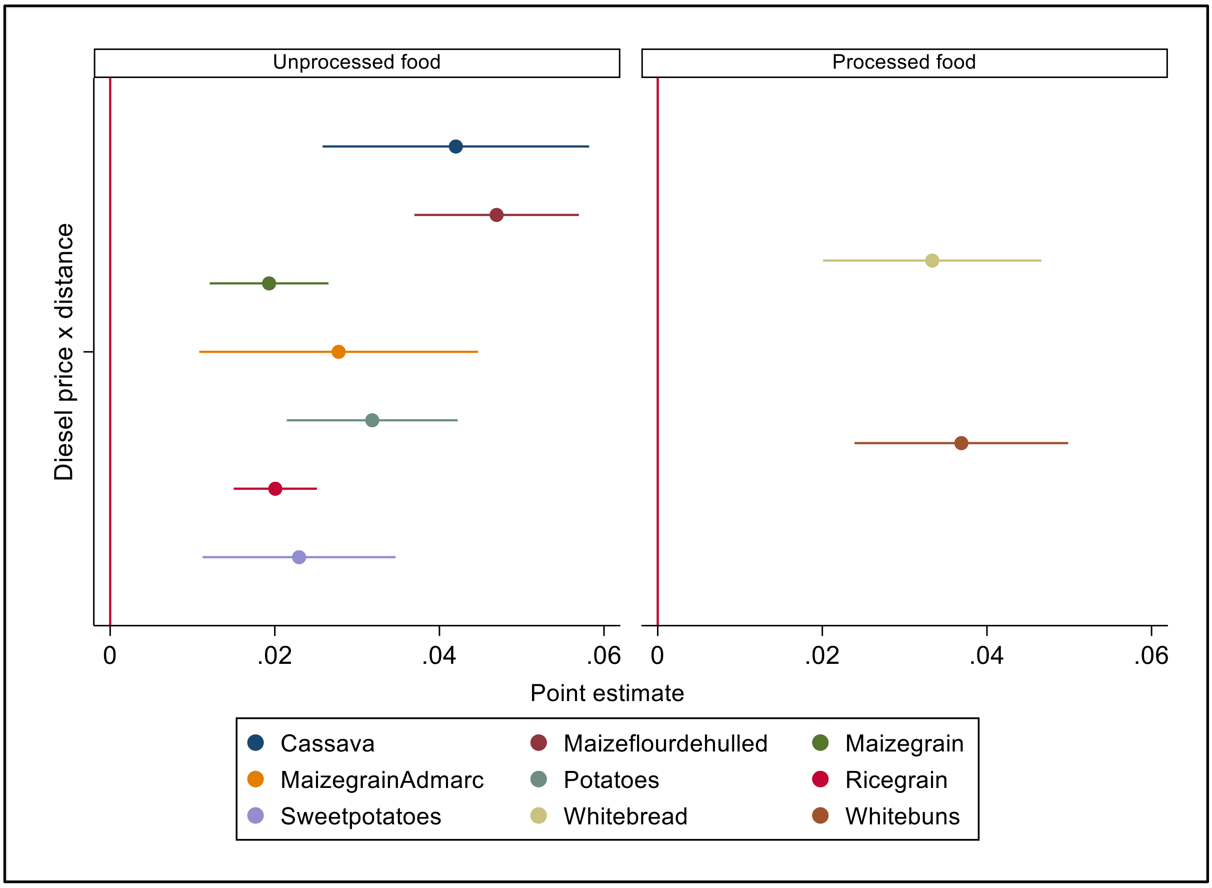

Figure 2.12: The association between transport costs and market price dispersion of staples including roots and tubers while accounting for price differences in the previous period

2.5.2.2 Legumes and nuts

Figure 2.13 presents the point estimates along with their confidence intervals of the association between transport costs and the price dispersion of legumes and nuts while controlling for the lag of the price difference across markets (see table A.38 in the appendix for the magnitude of the estimated coefficients).[30] Similarly, we find that the coefficient estimates for the lagged price differences for legumes and nuts are statistically significant, but closer to zero. Thus, the price differences in the previous period play a small role in predicting the price differences in the current period for legumes and nuts. Figure 2.13 shows that the association between transport costs and price differentials for white beans remain positive but becomes significant, while for groundnuts it becomes positive and significant at 95% confidence interval. However, the association between transport costs and price differentials for brown beans remain positive but becomes insignificant at 95% confidence interval. Thus, market traders do not supply white beans and groundnuts efficiently across markets as transport costs increase, *ceteris paribus.*

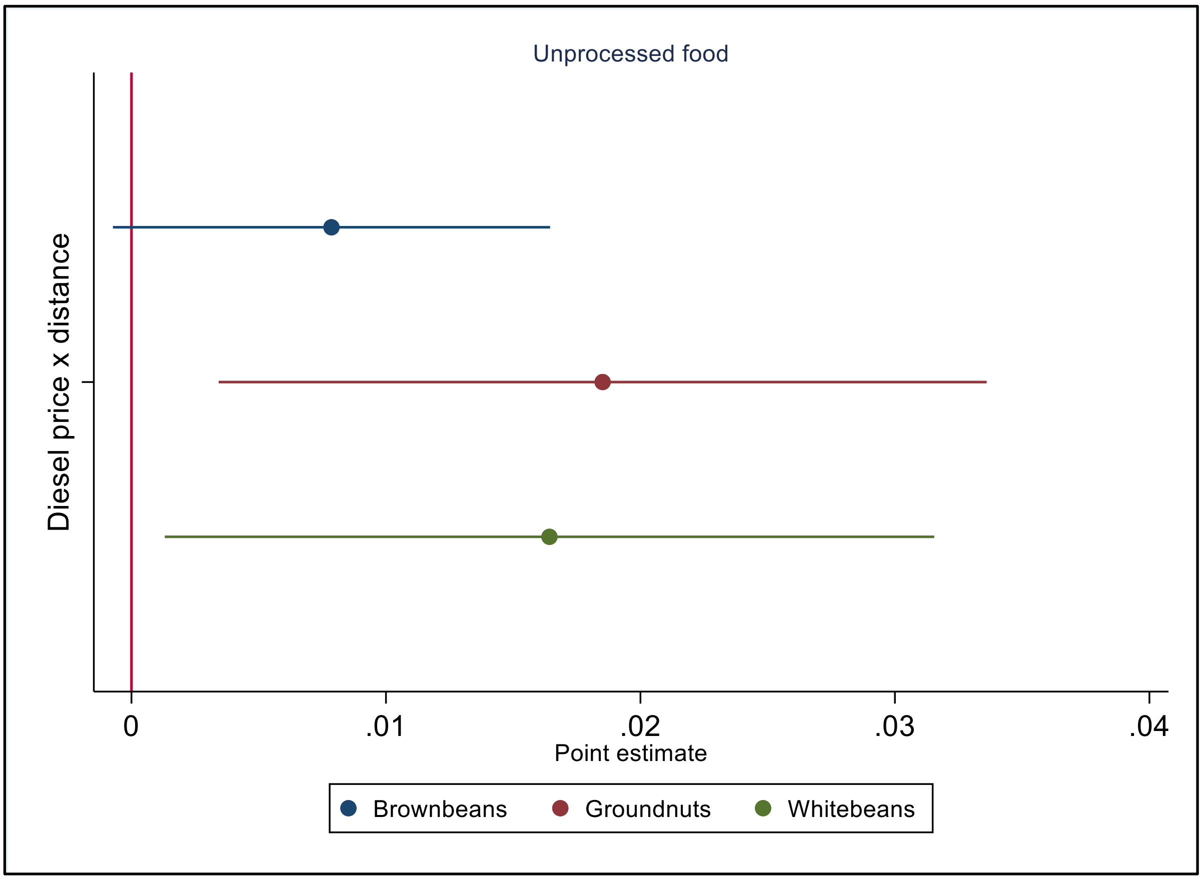

Figure 2.13: The association between transport costs and market price dispersion of legumes and nuts while accounting for price differences in the previous period

[30] We find that the direction of the association between transport costs and price differences and lagged price differences remain the same using Arellano–Bond estimator (see table A.43 in the appendix). Similarly, the direction of the association between transport costs and price differences flips for groundnuts compared to results from the main specification. Since estimation with ivpoisson does not include time fixed effects, we attribute the change in the direction of the impacts to different functional forms.

2.5.2.3 Animal source foods

Figure 2.14 presents the point estimates along with their confidence intervals of the association between transport costs and the price dispersion of animal source foods while controlling for the lag of the price difference across markets (see table A.39 in the appendix for the magnitude of the estimated coefficients).[31] We find that the coefficient estimates for the lagged price differences for both unprocessed and processed animal source foods are statistically significant, but closer to zero. Similarly, this means that the price differences in the previous period play a small role in predicting the price differences in the current period for animal source foods. Figure 2.14 shows that the association between transport costs and price differences remains positive and significant for eggs but it becomes significant for ultra-pasteurized milk at 95% confidence level. Conversely, the association between transport costs and price differences becomes positive and significant for beef, goat meat, and powdered milk at 95% confidence level. However, the direction of the association between transport costs and price differences for usipa remains positive and insignificant, while for utaka it remains negative and insignificant at 95% confidence level. Thus, transport costs have a positive long run relationship with price differences for both unprocessed and processed animal source foods under consideration, except for usipa and utaka.

[31] We find that the direction of the association between transport costs and price differences and lagged price differences remain the same, except for beef, ultra-pasteurized milk, and usipa using Arellano–Bond estimator (see table A.44 in the appendix). Similarly, the direction of the association between transport costs and price differences flips for beef, goat meat, powdered milk, and utaka compared to results from the main specification. Since estimation with ivpoisson does not include time fixed effects, we attribute the change in the direction of the impacts to different functional forms.

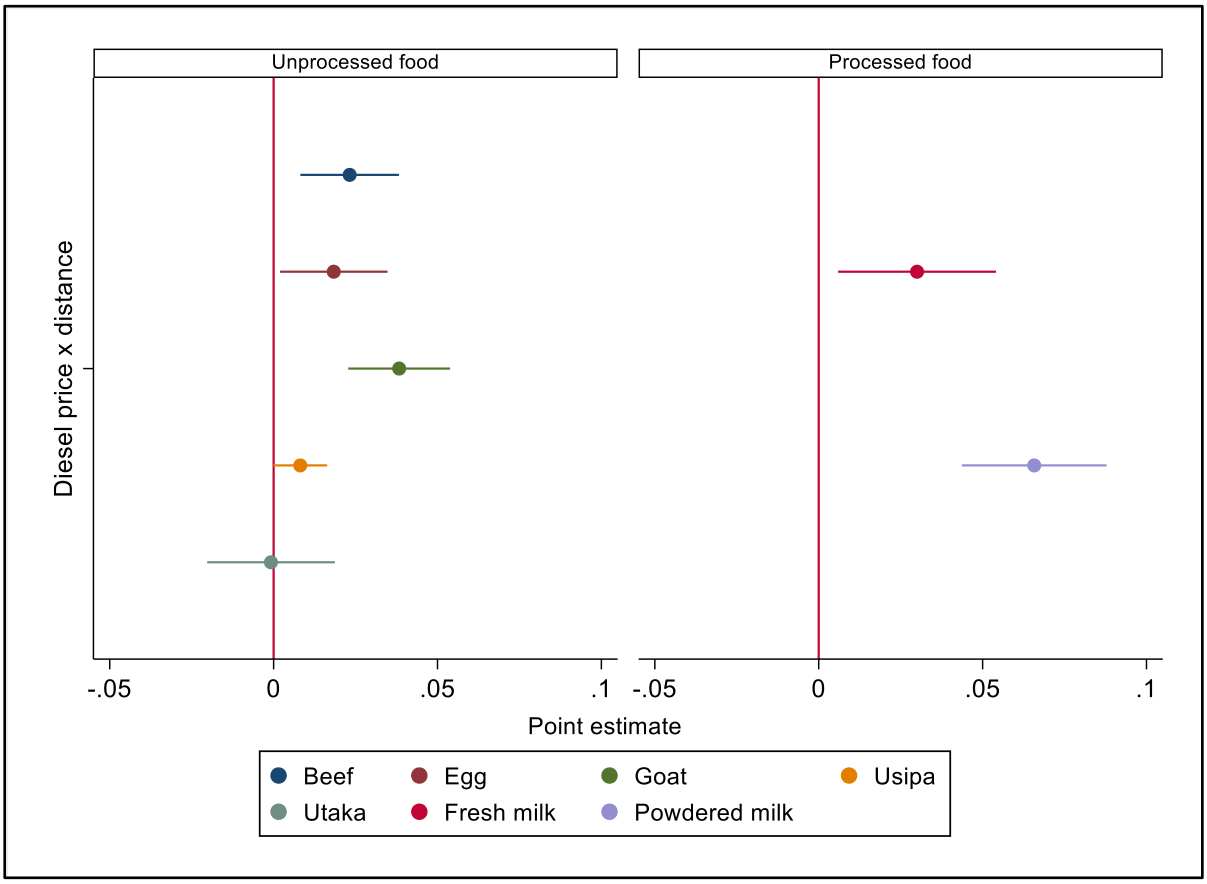

Figure 2.14: The association between transport costs and market price dispersion of animal source foods while accounting for price differences in the previous period

2.5.2.4 Vegetables

Figure 2.15 presents the point estimates along with their confidence intervals of the association between transport costs and the price dispersion of vegetables while controlling for the lag of the price difference across markets (see table A.40 in the appendix for the magnitude of the estimated coefficients).[32] Similarly, we find that the coefficient estimates for the lagged price differences for all vegetables are statistically significant, but closer to zero. Thus, the price differences in the previous period play a small role in predicting the price differences in the current period for vegetables. Figure 2.15 shows that the association between transport costs and price differences remains positive but becomes significant for okra, onions and pumpkin leaves, while for cabbage and rape leaves it becomes positive and

[32] We find that the direction of the association between transport costs and price differences and lagged price differences remain the same, except for okra using Arellano–Bond estimator (see table A.45 in the appendix). Similarly, the direction of the association between transport costs and price differences for cabbage and rape leaves compared to results from the main specification. Since estimation with ivpoisson does not include time fixed effects, we attribute the change in the direction of the impacts to different functional forms.

significant at 95% confidence level. These findings suggest that market traders do not supply vegetables efficiently across markets as transport costs increase in the long run, *ceteris paribus.*

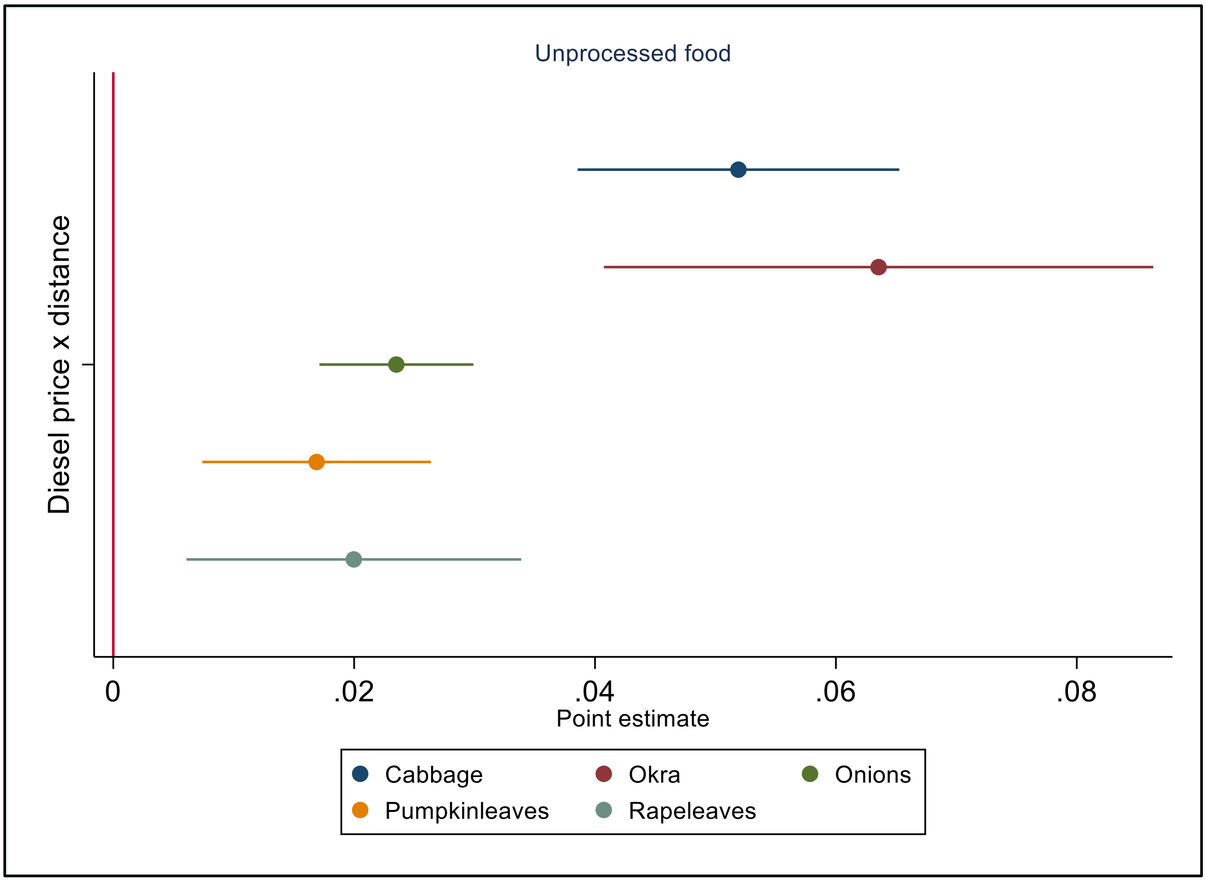

Figure 2.15: The association between transport costs and market price dispersion of vegetables while accounting for price differences in the previous period

2.5.2.5 Fruits

Figure 2.16 presents the point estimates along with their confidence intervals of the association between transport costs and the price dispersion of fruits while controlling for the lag of the price difference across markets (see table A.41 in the appendix for the magnitude of the estimated coefficients).[33] Similarly, we find that the coefficient estimates for the lagged price differences for all fruits are statistically significant, but closer to zero. Thus, the price differences in the previous period play a small role in predicting the price differences in the

[33] We find that the direction of the association between transport costs and price differences remains the same for bananas while for tomatoes it becomes negative using Arellano–Bond estimator (see table A.46 in the appendix). Similarly, the direction of the association between transport costs and price differences flips for both bananas and tomatoes compared to results from the main specification. Since estimation with ivpoisson does not include time fixed effects, we attribute the change in the direction of the impacts to different functional forms.

current period for fruits. Figure 2.16 shows that the association between transport costs and price differences becomes positive for both bananas and tomatoes. However, the relationship is significant only for bananas at 95% confidence interval. Thus, market traders do not supply bananas efficiently across markets as transport costs increase in the long run, *ceteris paribus.*

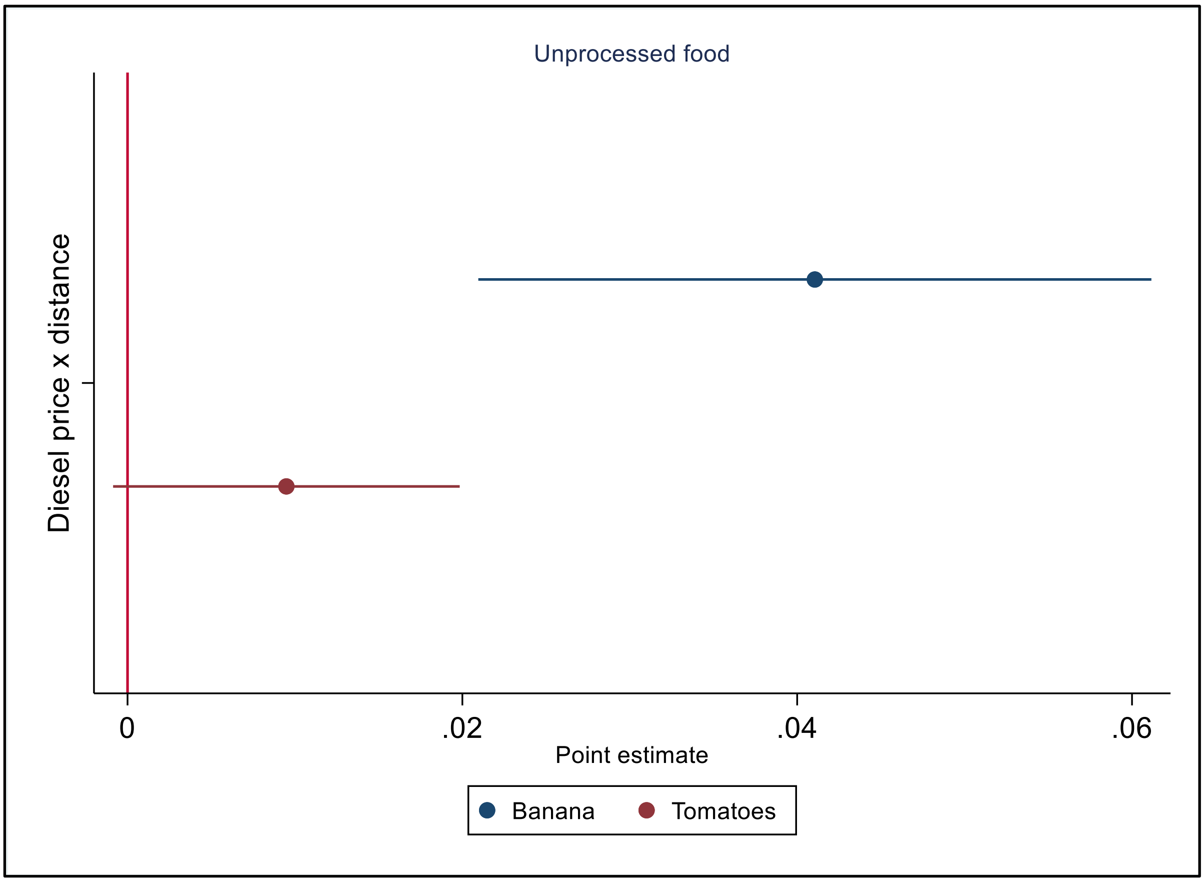

Figure 2.16: The association between transport costs and market price dispersion of fruits while accounting for price differences in the previous period

In summary, this section has shown that the price differences in the previous period play a small role in predicting the price differences in the current period in the long run. Further, the association between transport costs and price differences is positive and significant at 95% confidence level for most foods under consideration, except for brown beans, usipa, utaka, and tomatoes. Thus, market traders do not supply most foods efficiently across markets as transport costs between markets increase in the long run.

2.6 Conclusions and policy implications

This chapter estimates the influence of transport costs on price dispersion of various foods across markets in Malawi. Given that not all foods are produced within locations markets operate, market traders transport food from areas of high production (*i.e.,* surplus locations*)* to areas of low production (*i.e.,* deficit locations*)*, which involves transport costs. However, the larger share of marketing costs that market traders incur is transport costs (Fafchamps & Gabre-Madhin, 2006), which impedes market traders to transport food items from markets in surplus areas to markets in deficit areas. This increases price dispersion of various foods across markets, which increases food prices and reduces food affordability in markets located in deficit areas.

To better understand how transport costs are associated with price dispersion of various foods across markets, we systematically examine how the changes in fuel price with distance are associated with price dispersion of various foods across markets in Malawi. We use monthly consumer price monitoring panel data that Malawi NSO collects to compute the monthly CPI. Our dataset has monthly retail prices for 26 homogeneous food items that were consistently collected across 32 markets from January 2007 to July 2021. Further, we obtained monthly average diesel pump prices from the Malawi Energy Regulatory Authority and the route distance over paved roads between the market pairs from Google Maps. In combination with other data from various sources, we first estimate how the changes in transport costs are associated with overall price dispersion of various foods across markets using the panel non-linear dyadic regression model via the Poisson pseudo-likelihood regression estimator with multiple levels of fixed effects in the short run. Then, we investigate whether processed foods, perishable foods, or nutrient-dense foods modify the relationship between transport costs and overall price dispersion of various foods across markets in the short run. Finally, we estimate separately the panel non-linear dyadic regression model via the Poisson pseudo-likelihood regression estimator with multiple levels of fixed effects, and the instrumental variable Poisson (ivpoisson) estimator for each food item to examine the association between transport costs and price dispersion across markets in the short and long runs, respectively.

Overall, the results from our analysis reveal that the increase in fuel price with distance is associated with the reduction in overall price differentials for various foods under investigation across markets in the short run, on average. We find that this counterintuitive influence is driven by processed foods but not by perishable foods and nutrient-dense foods for which the changes in transport costs is associated with an increase in overall price differences across markets in the short run. Moving on to separate results for each food item, we find that the association between the changes in transport costs and price differences remains negative and significant for rice grain, beef, goat meat, powdered milk, and bananas but it becomes positive and significant for maize flour dehulled, maize grain (private), maize grain (ADMARC), brown beans, and eggs providing additional evidence that the counterintuitive influence of the changes in transport costs on overall price differentials is driven by food items that have lower search costs and easy to aggregate in the short run.

Turning to spatial heterogeneity of the association between transport costs and price differences for each food item, we find that the negative association between the changes in transport costs and price differences remains significant at some distance levels for rice grain, beef, goat meat, powdered milk, and bananas but it becomes significant for some distance levels for white bread and tomatoes relative to the base category of market pairs that are less than 102 km apart in the short run. Conversely, the positive association between the changes in transport costs and price differences remains significant at some distance levels for maize flour dehulled, maize grain (ADMARC), brown beans, and eggs but it becomes significant for some distance levels for usipa and white beans relative to the base category of market pairs that are less than 102 km apart. However, the association between the changes in transport costs and price differences for maize grain (private) becomes negative and significant at smaller distance levels relative to the base category of market pairs that are less than 102 km apart, suggesting that private traders supply maize grain efficiently between markets that are closer. Overall, the magnitudes of the relationships are smaller for market pairs that are closer to each other for all foods under consideration. Moving on to the influence of transport costs on price differences for each food item in the long run, we find that the price differences in the previous period play a small role in predicting the price differences in the current period, and the increase in transport costs is associated with the increase in price differentials for most foods under investigation, except for brown beans, usipa, utaka, and tomatoes.

Our study has demonstrated how the increase in transport costs, as impacted by the increase in fuel price, is associated with the price dispersion of various foods across markets in Malawi. Overall, transport costs shock is associated with the decrease in spatial inequality in overall food affordability across markets in the short run. However, spatial inequality in food affordability widens for maize flour dehulled maize grain, maize grain (private), maize grain (ADMARC), brown beans, and eggs in the short run. Given that these food items are important in a Malawian diet, these findings indicate that there are both food security and nutritional implications of changes in transport costs. Since the increase in transport costs will limit trade, increase consumer prices, and reduce food affordability across markets there is need to devise strategies that will lower search costs to allow market traders to easily organise larger loads that will minimise the effect of fuel costs on distance, which is associated with poor market integration across the country. Examining whether the increase in trucking competition improves market integration of various foods is an area for further research. According to Fafchamps & Gabre-Madhin (2006), either removing taxes on diesel fuel that large trucks use or removing toll road fees for vehicles carrying food items across the country would lower transport costs for market traders. Whether removing taxes on diesel fuel prices or toll road fees will reduce transport costs is an area for further research. Another potential area for further research is to examine general equilibrium effects of increases in fuel costs on the economy. What we do know is that increasing market integration of food over time will allow market traders to organise and transport larger loads that will lower transport costs per unit volume that will in turn reduce the effect of fuel costs on distance and promote trade from surplus locations to deficit locations. In the longer term, there is need to consider investment in least-costs transport alternatives to road transport such as rail transportation (Donaldson, 2018; Zant, 2018) to increase food affordability and improve nutrition across the country.

2.7 Appendix A

Table A.1: Description and measurement of variables used in the study

Variable	Description	Data type	Data source
Prices	Monthly market food prices in Malawian Kwacha	Time variant	National Statistical Office
Population density	Annual number of people per unit of area per district.	Time variant	Malawi's Population and Housing Census
SPEI	This is a new climatic drought intensity index based on precipitation, temperature, and evapotranspiration for monitoring drought in diverse system.	Time variant	NASA LaRC POWER Project
Diesel price	Monthly average diesel pump prices in Malawi Kwacha.	Time variant	Malawi Energy Regulatory Authority
Estimates	Annual crop, livestock, and fish production estimates in Mt	Time variant	Ministry of Agriculture
Distance between market pairs	Average route distance over paved roads between market pairs in Km	Time invariant	Google Maps

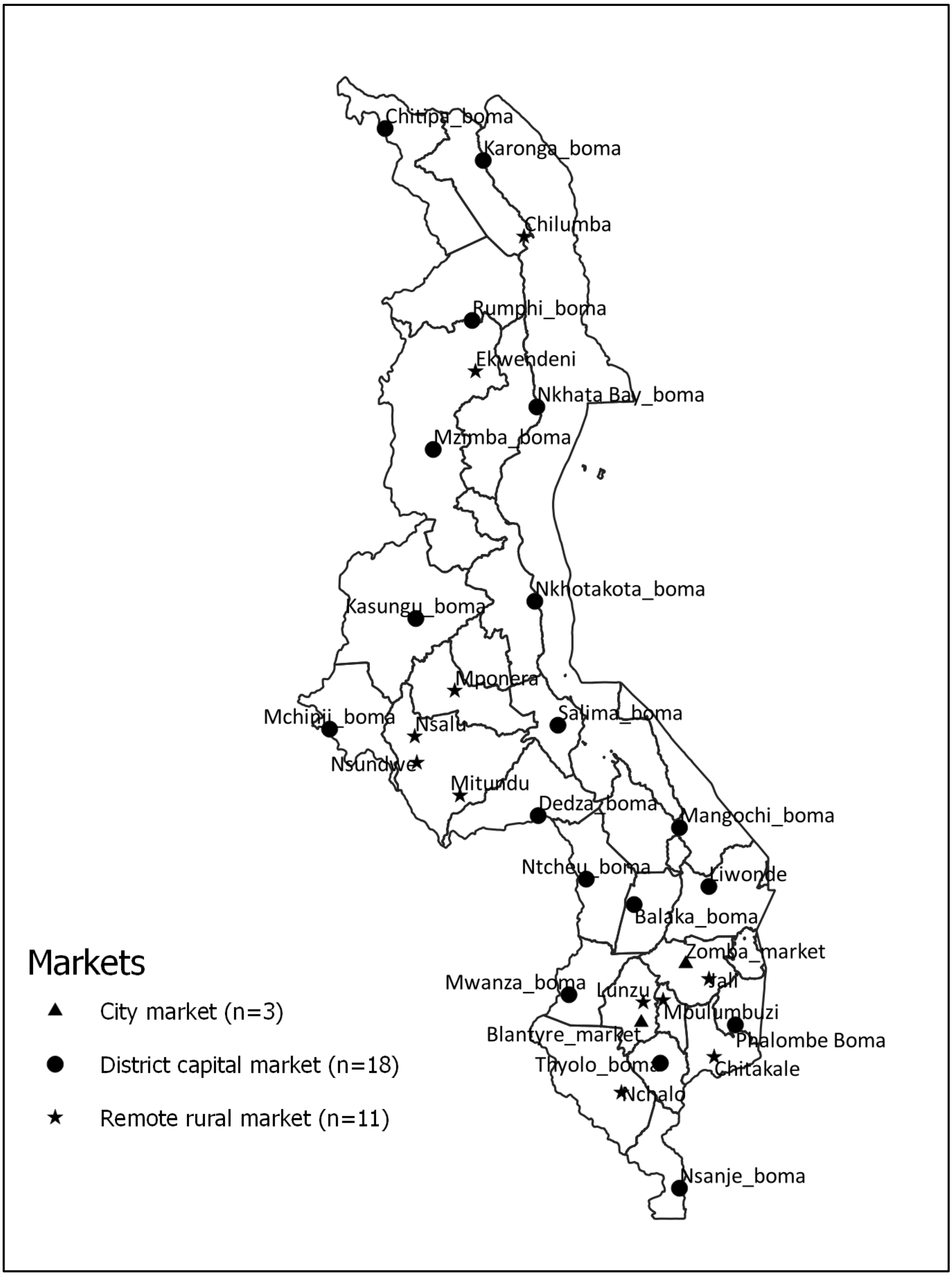

Figure A.1: Local large food markets across Malawi

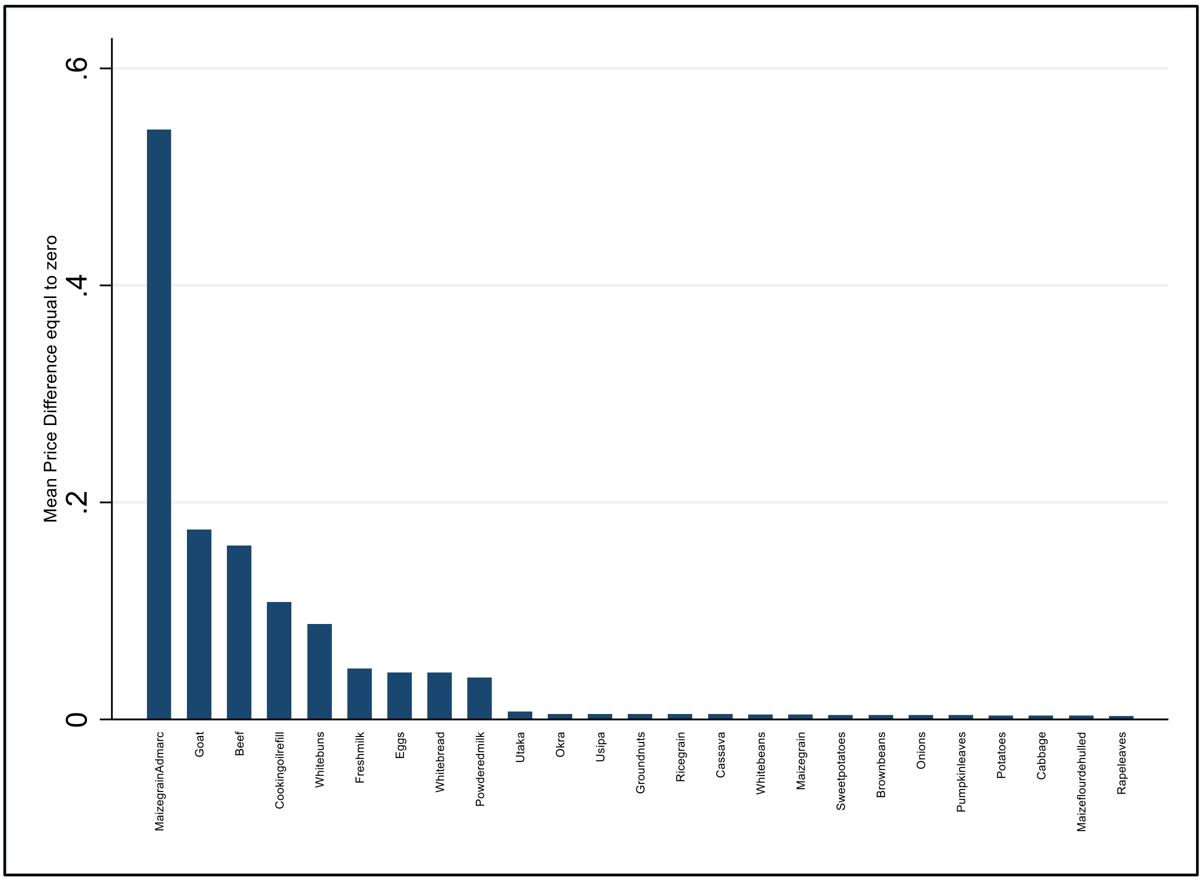

Figure A.2: Percent market-pair month of price differences equal to zero for each food item

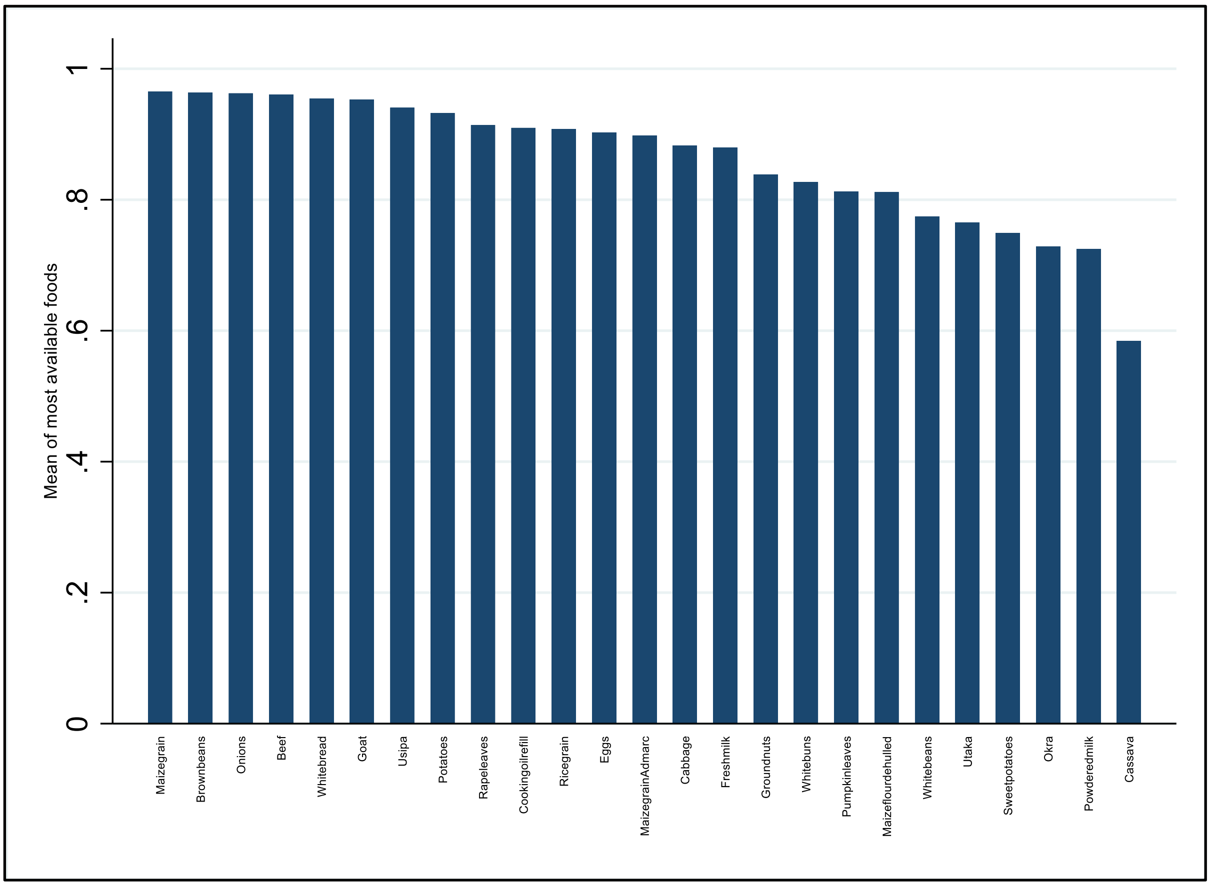

Figure A.3: Percent market-pair month of the most available foods

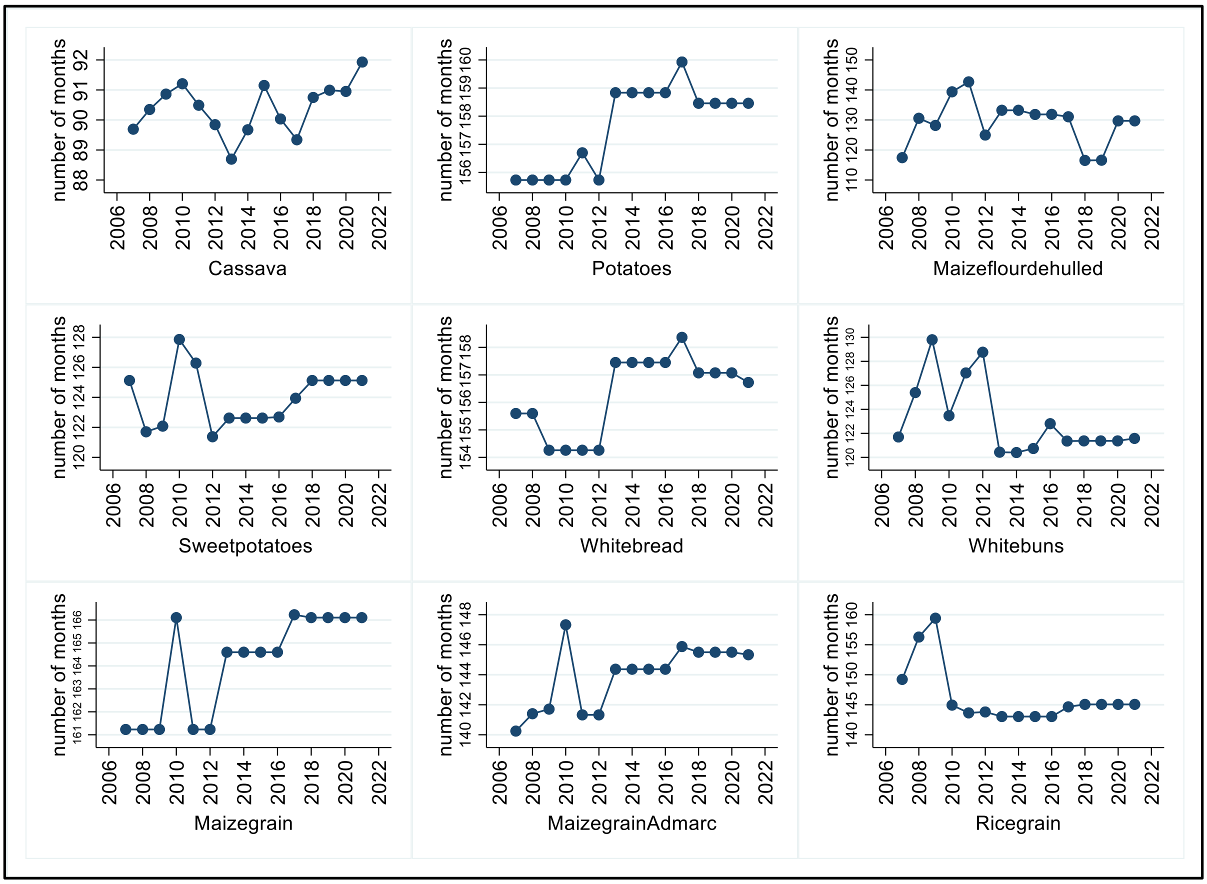

Figure A.4: Average month market-pair availability for staples including roots and tubers

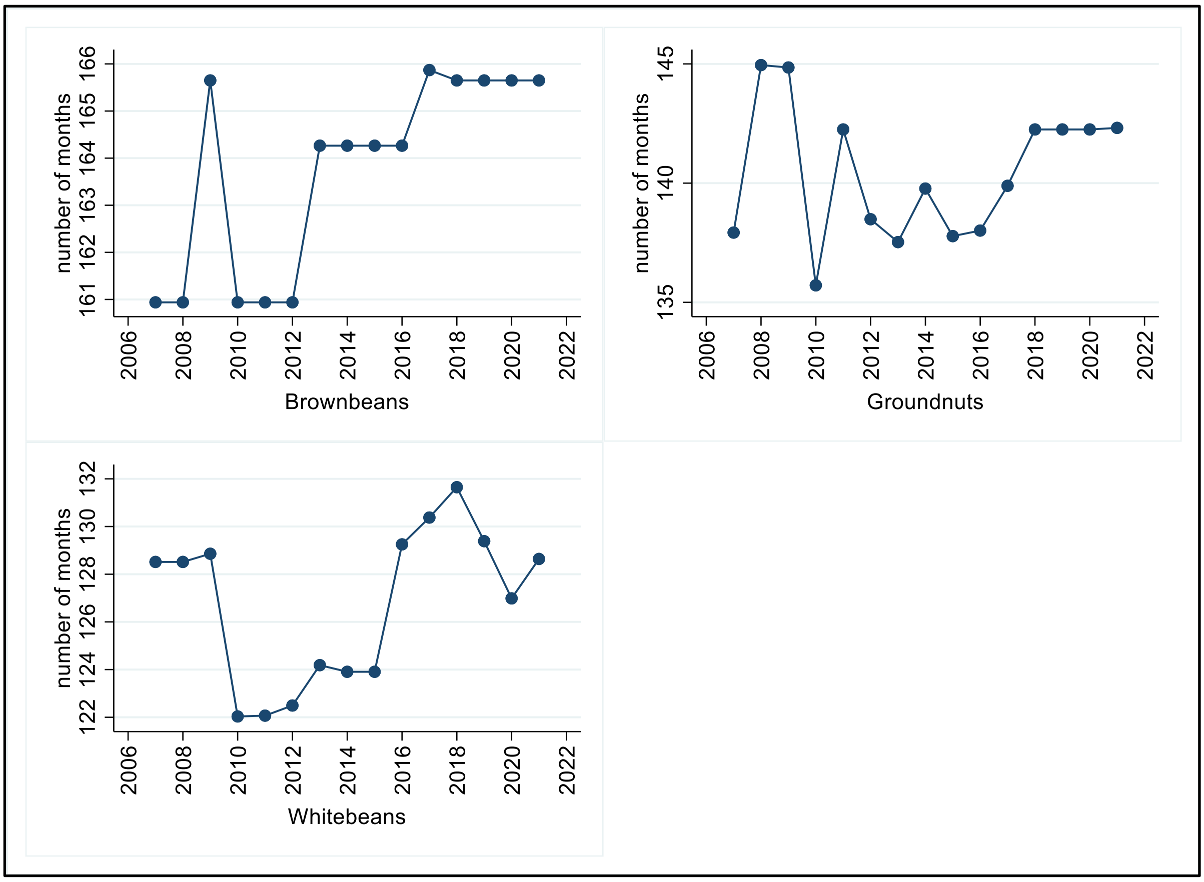

Figure A.5: Average month market-pair availability for legumes and nuts

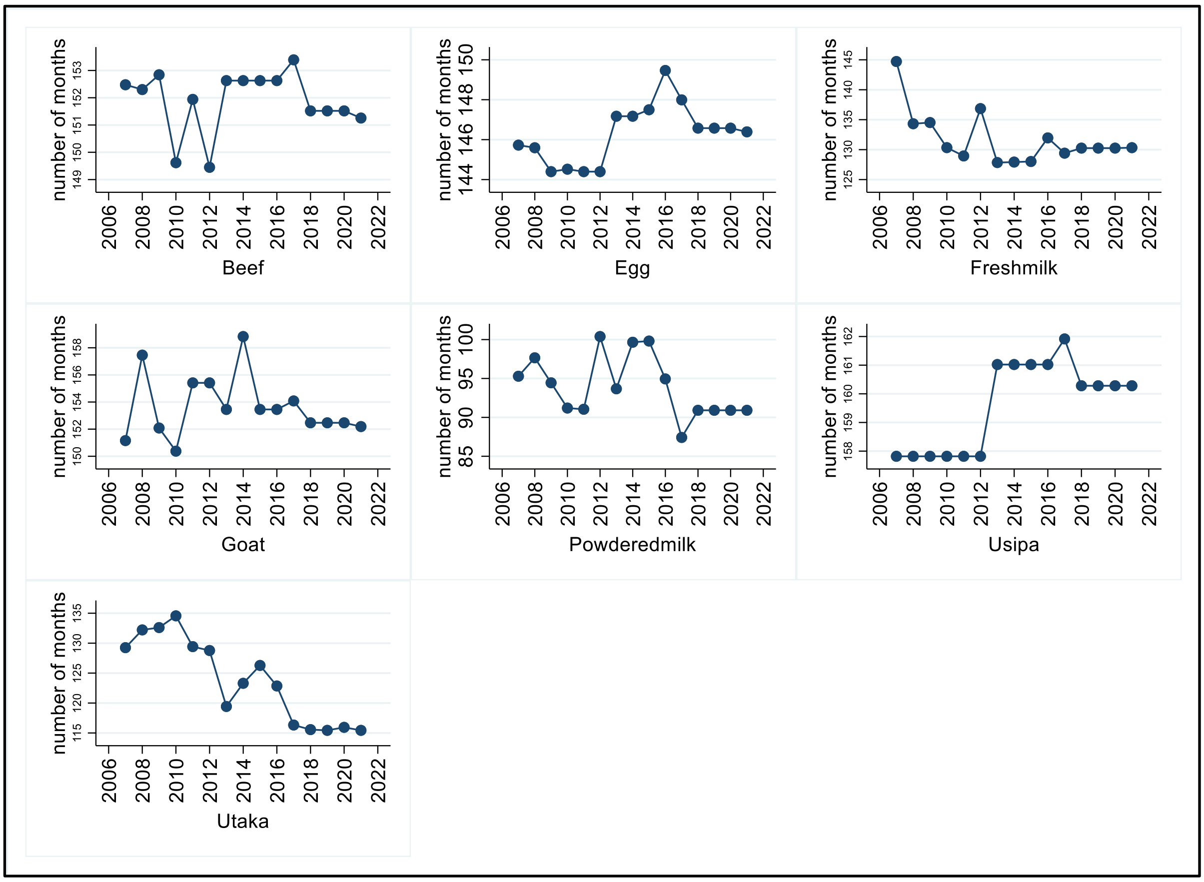

Figure A.6: Average month market-pair availability for animal source foods

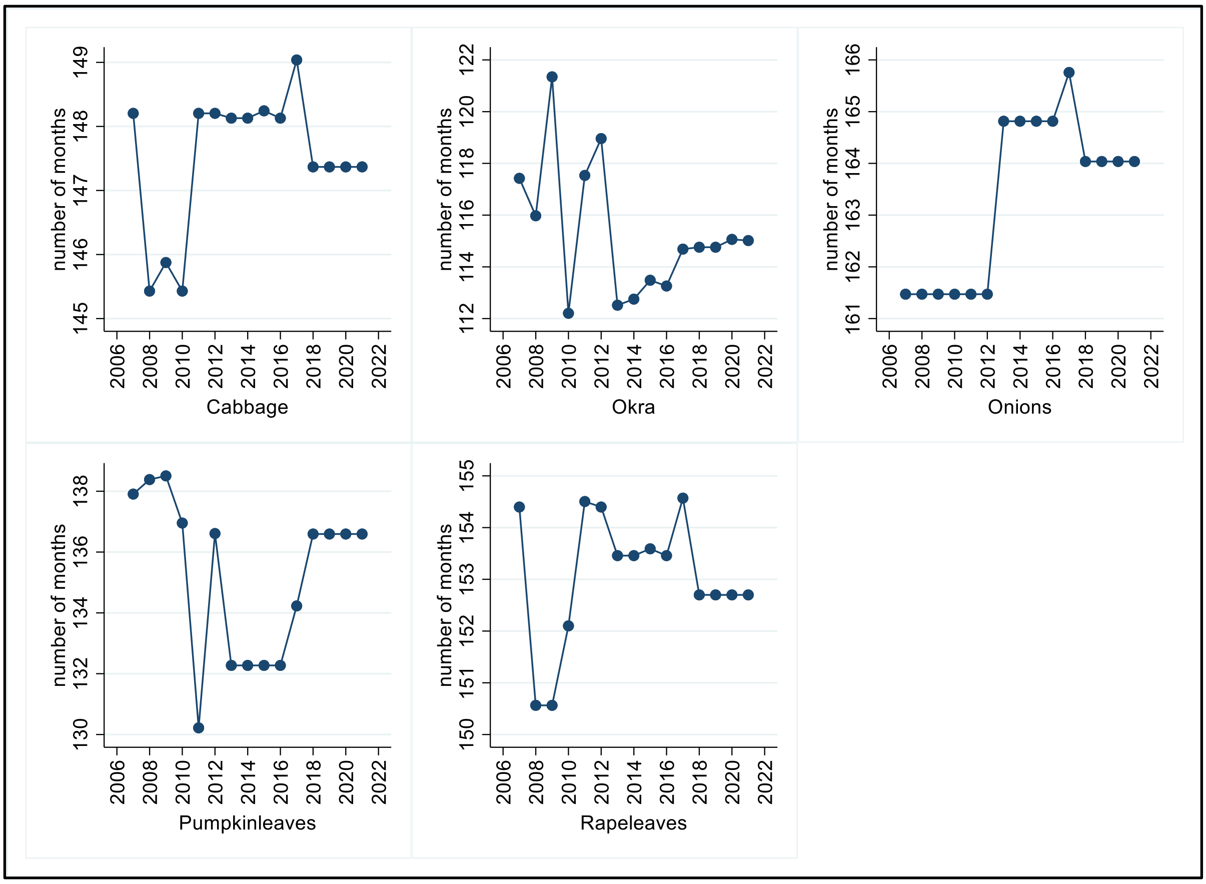

Figure A.7: Average month market-pair availability for vegetables

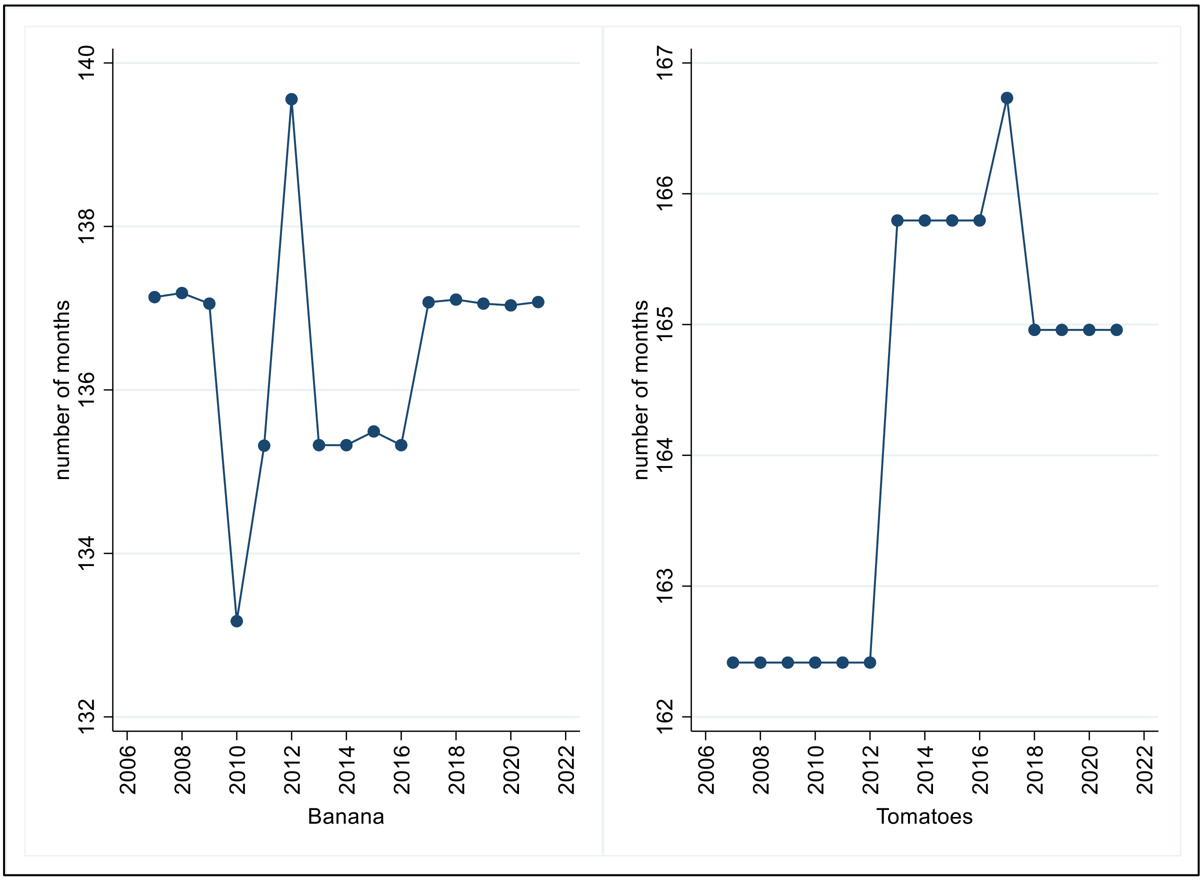

Figure A.8: Average month market-pair availability for fruits

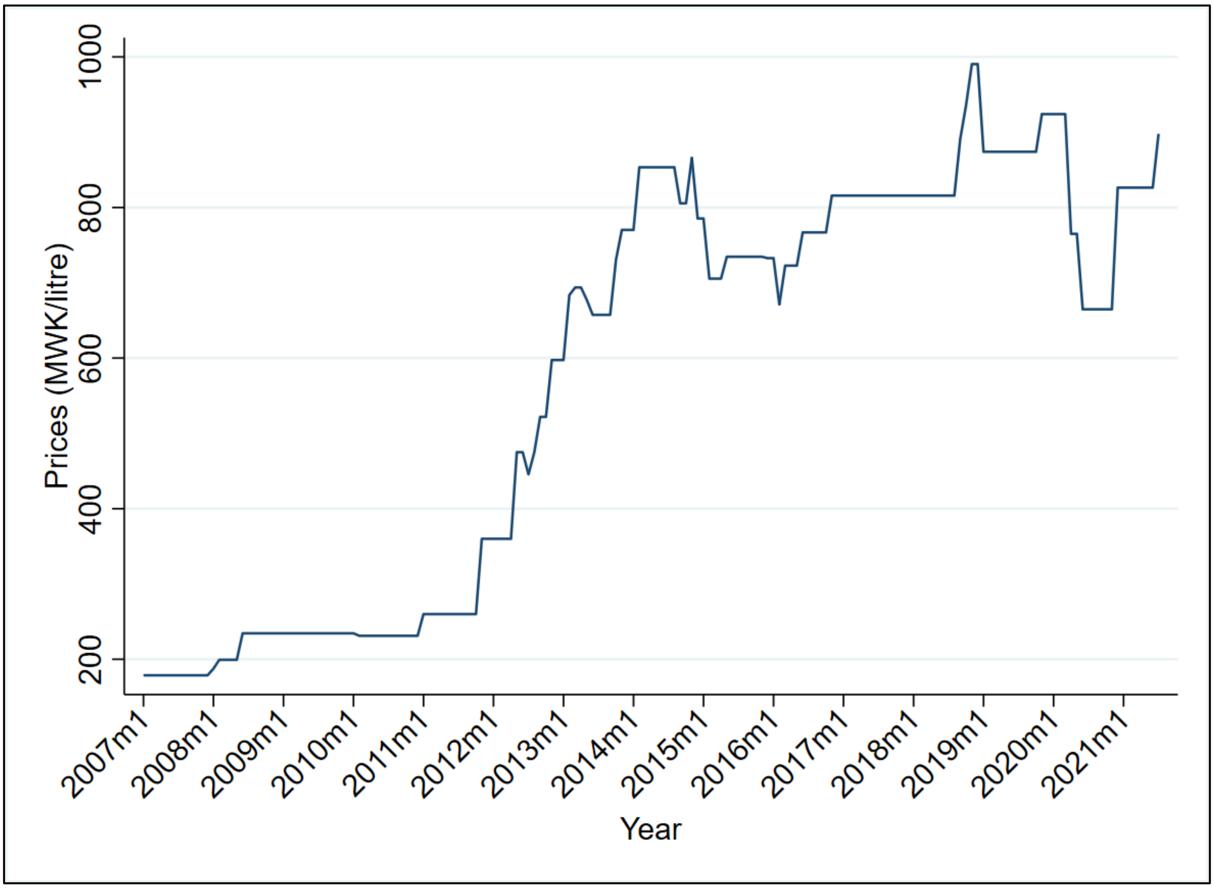

Figure A.9: Variation in local diesel prices over time

Table A.2: The association between transport costs and market price dispersion of staples including roots and tubers

Dependent variable: $\|P_{xt} - P_{yt}\|$	Full sample	Cassava	Maize flour dehulled	Maize grain (private)	Maize grain (ADMARC)	Potatoes	Rice grain	Sweet potatoes	White bread	White buns
Diesel price x distance	0.0267*	-0.0722	0.0747**	0.0758**	0.341***	0.0360	-0.102**	-0.0647	-0.0379	0.0536
	(0.0156)	(0.0584)	(0.0328)	(0.0301)	(0.106)	(0.0476)	(0.0459)	(0.0570)	(0.0549)	(0.0420)
Distance-specific linear time trends	0.0259	-0.278***	0.118	-0.0926	0.378**	-0.0872	0.223***	-0.0435	0.00715	0.0540
	(0.0438)	(0.0763)	(0.0829)	(0.0579)	(0.168)	(0.0699)	(0.0759)	(0.0702)	(0.112)	(0.114)
Distance-specific quadratic time trends	-0.0000353	0.000429***	-0.000173	0.000136	-0.000607**	0.000140	-0.000332***	0.0000756	-0.0000115	-0.0000805
	(0.0000671)	(0.000119)	(0.000128)	(0.0000880)	(0.000264)	(0.000107)	(0.000117)	(0.000109)	(0.000174)	(0.000172)
Distance-specific cubic time trends	1.56e-08	-0.000000220***	8.41e-08	-6.65e-08	0.000000321**	-7.42e-08	0.000000164***	-4.26e-08	6.39e-09	3.97e-08
	(3.41e-08)	(6.10e-08)	(6.59e-08)	(4.46e-08)	(0.000000138)	(5.45e-08)	(5.94e-08)	(5.64e-08)	(8.89e-08)	(8.66e-08)
Difference in population density	-0.0895**	-0.0226	-0.0469	-0.00780	-0.00242	-0.0544	0.0203	-0.0237	0.0692	-0.267***
	(0.0386)	(0.0670)	(0.0726)	(0.0445)	(0.0521)	(0.0581)	(0.0349)	(0.0286)	(0.0609)	(0.0747)
=1 if one of the markets experienced flood shocks	0.000808	0.0272*	0.00477	0.0241	0.00181	-0.00469	0.00186	0.0237	0.0114	-0.0180*
	(0.00424)	(0.0147)	(0.0118)	(0.0167)	(0.0266)	(0.0137)	(0.0156)	(0.0212)	(0.0141)	(0.0109)
=1 if one of the markets experienced drought shocks	-0.0172***	-0.0464**	-0.0267*	0.0142	-0.0326	0.0286***	-0.0572***	-0.0139	0.0318	-0.0185
	(0.00561)	(0.0184)	(0.0153)	(0.0147)	(0.0305)	(0.0101)	(0.0149)	(0.0139)	(0.0236)	(0.0220)
Difference in local production	-0.667***	-0.823	1.211	0.566	0.978	-1.343***	0.501	-0.868	-	-
	(0.253)	(1.005)	(0.901)	(0.534)	(1.262)	(0.456)	(0.331)	(1.050)	-	-
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Market-pair FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	593473	44023	57195	79971	64135	77243	69666	59931	76513	58832

Note: Poisson pseudo-likelihood regression with multiple levels of fixed effects (ppmlhdfe) estimator results for each food items are presented in each column. The dependent variable is absolute value of price difference. Route distance over paved road between the market pairs is measured in hundreds of kilometres, and diesel fuel price in thousands of Malawian Kwacha. Dyadic clustered standard errors at the market in parentheses * p<0.10, ** p<0.05, *** p<0.010

Table A.3: The association between transport costs and market price dispersion of legumes and nuts

Dependent variable: $\lvert P_{xt} - P_{yt} \rvert$	Full sample	Brown beans	Groundnuts	White beans
Diesel price x distance	0.0141	0.0884**	-0.0548	0.0300
	(0.0279)	(0.0373)	(0.0395)	(0.0443)
Distance-specific linear time trends	0.0195	0.121**	-0.0254	-0.0537
	(0.0360)	(0.0507)	(0.0438)	(0.0569)
Distance-specific quadratic time trends	-0.0000283	-0.000188**	0.0000437	0.0000847
	(0.0000558)	(0.0000784)	(0.0000678)	(0.0000881)
Distance-specific cubic time trends	1.34e-08	9.66e-08**	-2.47e-08	-4.46e-08
	(2.87e-08)	(4.03e-08)	(3.49e-08)	(4.52e-08)
Difference in population density	-0.0617	-0.0655	-0.0488	-0.0766
	(0.0598)	(0.0548)	(0.0568)	(0.0913)
=1 if one of the markets experienced flood shocks	0.00385	0.000444	0.00564	0.00244
	(0.00796)	(0.0107)	(0.0106)	(0.0108)
=1 if one of the markets experienced drought shocks	-0.0246**	-0.0322**	-0.0200	-0.0220
	(0.0103)	(0.0132)	(0.0129)	(0.0139)
Difference in local production	-0.168*	0.870*	-0.264***	1.285**
	(0.0904)	(0.503)	(0.0744)	(0.533)
Month FE	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
Market-pair FE	Yes	Yes	Yes	Yes
N	207476	79826	67104	60546

Note: Poisson pseudo-likelihood regression with multiple levels of fixed effects (ppmlhdfe) estimator results for each food items are presented in each column. The dependent variable is absolute value of price difference. Route distance over paved road between the market pairs is measured in hundreds of kilometres, and diesel fuel price in thousands of Malawian Kwacha. Dyadic clustered standard errors at the market in parentheses * p<0.10, ** p<0.05, *** p<0.010

Table A.4: The association between transport costs and market price dispersion of animal source foods

Dependent variable: $\lvert P_{xt} - P_{yt} \rvert$	Full sample	Beef	Eggs	Ultra-pasteurized milk	Goat meat	Powdered milk	Usipa	Utaka
Diesel price x distance	-0.0808***	-0.0937***	0.117**	0.0341	-0.112***	-0.210***	0.0357	-0.0976
	(0.0171)	(0.0347)	(0.0517)	(0.0377)	(0.0282)	(0.0786)	(0.0505)	(0.0646)
Distance-specific linear time trends	-0.0166	0.00446	0.00843	0.151	-0.0416	-0.00296	0.0518	-0.0709
	(0.0304)	(0.0353)	(0.0742)	(0.115)	(0.0602)	(0.117)	(0.0760)	(0.0582)
Distance-specific quadratic time trends	0.0000303	-0.0000182	-0.0000144	-0.000235	0.0000608	0.0000102	-0.0000785	0.000122
	(0.0000471)	(0.0000552)	(0.000115)	(0.000179)	(0.0000930)	(0.000183)	(0.000119)	(0.0000885)
Distance-specific cubic time trends	-1.73e-08	1.56e-08	7.71e-09	0.000000121	-2.93e-08	-6.32e-09	3.96e-08	-6.81e-08
	(2.42e-08)	(2.87e-08)	(5.96e-08)	(9.27e-08)	(4.76e-08)	(9.48e-08)	(6.18e-08)	(4.47e-08)
Difference in population density	-0.00401	-0.0229	-0.00577	0.0260	-0.00117	0.193**	-0.0238	-0.112
	(0.0604)	(0.0554)	(0.0290)	(0.0954)	(0.0235)	(0.0780)	(0.0828)	(0.0904)
=1 if one of the markets experienced flood shocks	-0.00655	-0.00414	-0.0190*	-0.0193*	0.00864	-0.0404***	0.0238***	-0.00523
	(0.00606)	(0.0135)	(0.0114)	(0.0116)	(0.0111)	(0.0128)	(0.00901)	(0.0131)
=1 if one of the markets experienced drought shocks	0.00130	0.00603	0.0306**	0.0354	-0.00808	0.00570	-0.0198	0.0141
	(0.0108)	(0.0166)	(0.0155)	(0.0221)	(0.0134)	(0.0263)	(0.0126)	(0.0180)
Difference in local production	0.0966	1.167	-1.800	-	2.384*	-	0.0554	0.166*
	(0.0924)	(4.672)	(2.657)	-	(1.286)	-	(0.246)	(0.0954)
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Market-pair FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	458524	74127	71621	62329	74588	40866	78279	56714

Note: Poisson pseudo-likelihood regression with multiple levels of fixed effects (ppmlhdfe) estimator results for each food items are presented in each column. The dependent variable is absolute value of price difference. Route distance over paved road between the market pairs is measured in hundreds of kilometres, and diesel fuel price in thousands of Malawian Kwacha. Dyadic clustered standard errors at the market in parentheses * p<0.10, ** p<0.05, *** p<0.010

Table A.5: The association between transport costs and market price dispersion of vegetables

Dependent variable: $\lvert P_{xt} - P_{yt} \rvert$	Full sample	Cabbage	Okra	Onions	Pumpkin leaves	Rape leaves
Diesel price x distance	0.00593 (0.0224)	-0.0844 (0.0605)	0.0208 (0.0725)	0.00136 (0.0308)	0.00691 (0.0285)	-0.00230 (0.0353)
Distance-specific linear time trends	0.0629* (0.0336)	0.0785 (0.0855)	0.0186 (0.0956)	0.0211 (0.0500)	0.151** (0.0692)	0.0700 (0.0550)
Distance-specific quadratic time trends	-0.0000955* (0.0000517)	-0.000109 (0.000131)	-0.0000287 (0.000147)	-0.0000291 (0.0000772)	-0.000238** (0.000109)	-0.000106 (0.0000846)
Distance-specific cubic time trends	4.81e-08* (2.64e-08)	4.95e-08 (6.69e-08)	1.44e-08 (7.46e-08)	1.31e-08 (3.95e-08)	0.000000124** (5.66e-08)	5.33e-08 (4.31e-08)
Difference in population density	-0.0165 (0.0440)	-0.0737 (0.0490)	-0.00974 (0.0368)	-0.0878* (0.0532)	0.0868* (0.0489)	0.0113 (0.0589)
=1 if one of the markets experienced flood shocks	0.00299 (0.00575)	0.0198 (0.0133)	0.00618 (0.0155)	-0.0128 (0.0101)	0.000551 (0.0107)	0.0313** (0.0138)
=1 if one of the markets experienced drought shocks	0.0263*** (0.00719)	0.0454*** (0.0155)	0.0294* (0.0158)	0.0527*** (0.0158)	-0.0251** (0.0127)	-0.00796 (0.0156)
Difference in local production	-1.271 (0.933)	0.298 (1.201)	1.963 (2.305)	-4.656** (2.199)	2.660 (2.066)	3.722** (1.607)
Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Market-pair FE	Yes	Yes	Yes	Yes	Yes	Yes
N	345444	72133	53953	80090	64588	74680

Note: Poisson pseudo-likelihood regression with multiple levels of fixed effects (ppmlhdfe) estimator results for each food items are presented in each column. The dependent variable is absolute value of price difference. Route distance over paved road between the market pairs is measured in hundreds of kilometres, and diesel fuel price in thousands of Malawian Kwacha. Dyadic clustered standard errors at the market in parentheses * p<0.10, ** p<0.05, *** p<0.010

Table A.6: The association between transport costs and market price dispersion of fruits

Dependent variable: $\lvert P_{xt} - P_{yt} \rvert$	Full sample	Bananas	Tomatoes
Diesel price x distance	-0.0364**	-0.0815***	-0.0165
	(0.0170)	(0.0223)	(0.0212)
Distance-specific linear time trends	0.0412	-0.00446	0.0508
	(0.0352)	(0.0838)	(0.0481)
Distance-specific quadratic time trends	-0.0000573	0.00000882	-0.0000704
	(0.0000549)	(0.000129)	(0.0000740)
Distance-specific cubic time trends	2.63e-08	-5.09e-09	3.21e-08
	(2.84e-08)	(6.60e-08)	(3.78e-08)
Difference in population density	-0.0167	-0.0277	-0.0112
	(0.0396)	(0.0436)	(0.0397)
=1 if one of the markets experienced flood shocks	0.000961	0.00651	-0.00162
	(0.0128)	(0.0143)	(0.0148)
=1 if one of the markets experienced drought shocks	0.00851	0.0220*	0.00180
	(0.0118)	(0.0113)	(0.0151)
Difference in local production	-4.015***	0.245	-4.222***
	(1.312)	(6.991)	(1.197)
Month FE	Yes	Yes	Yes
Time FE	Yes	Yes	Yes
Market-pair FE	Yes	Yes	Yes
N	146611	66053	80558

Note: Poisson pseudo-likelihood regression with multiple levels of fixed effects (ppmlhdfe) estimator results for each food items are presented in each column. The dependent variable is absolute value of price difference. Route distance over paved road between the market pairs is measured in hundreds of kilometres, and diesel fuel price in thousands of Malawian Kwacha. Dyadic clustered standard errors at the market in parentheses * p<0.10, ** p<0.05, *** p<0.010

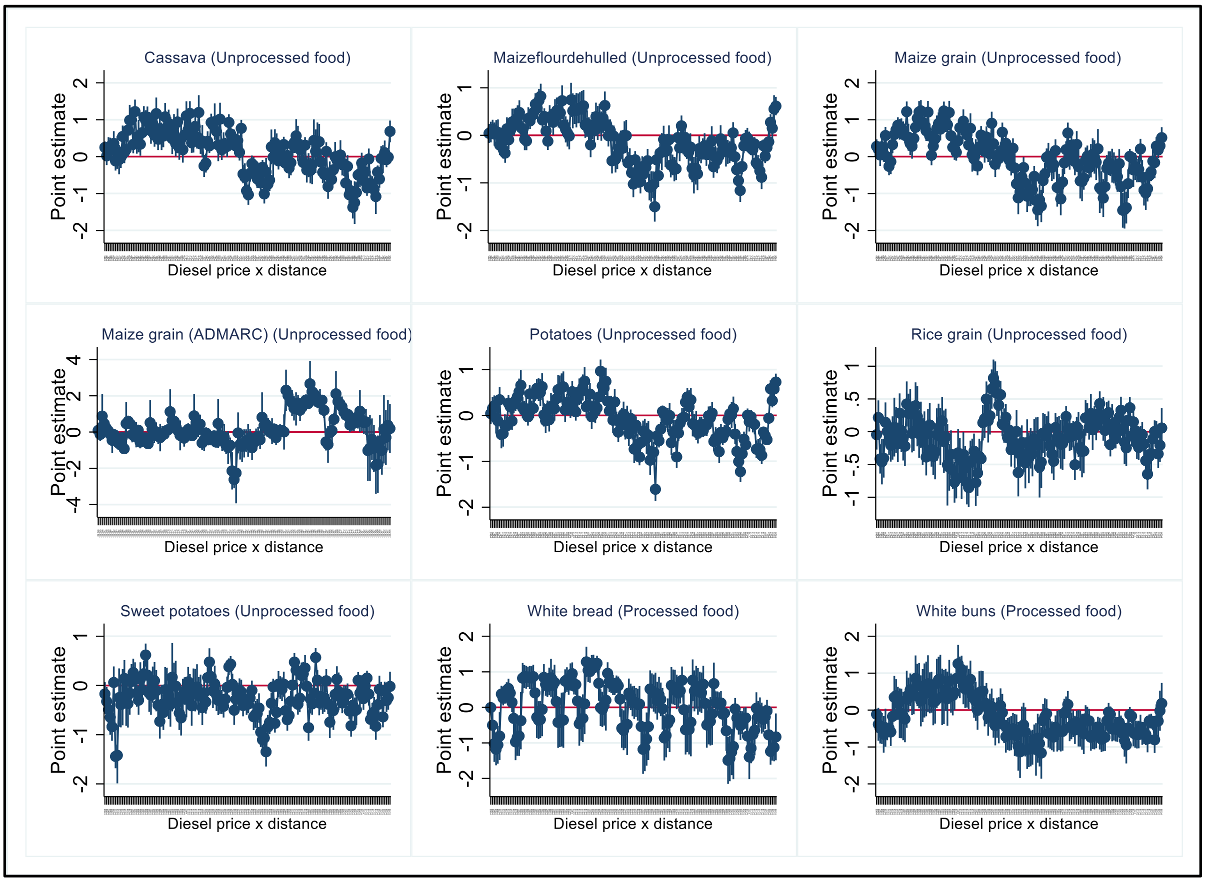

Figure A.10: Time dummies after fixed effects Poisson estimator for staples

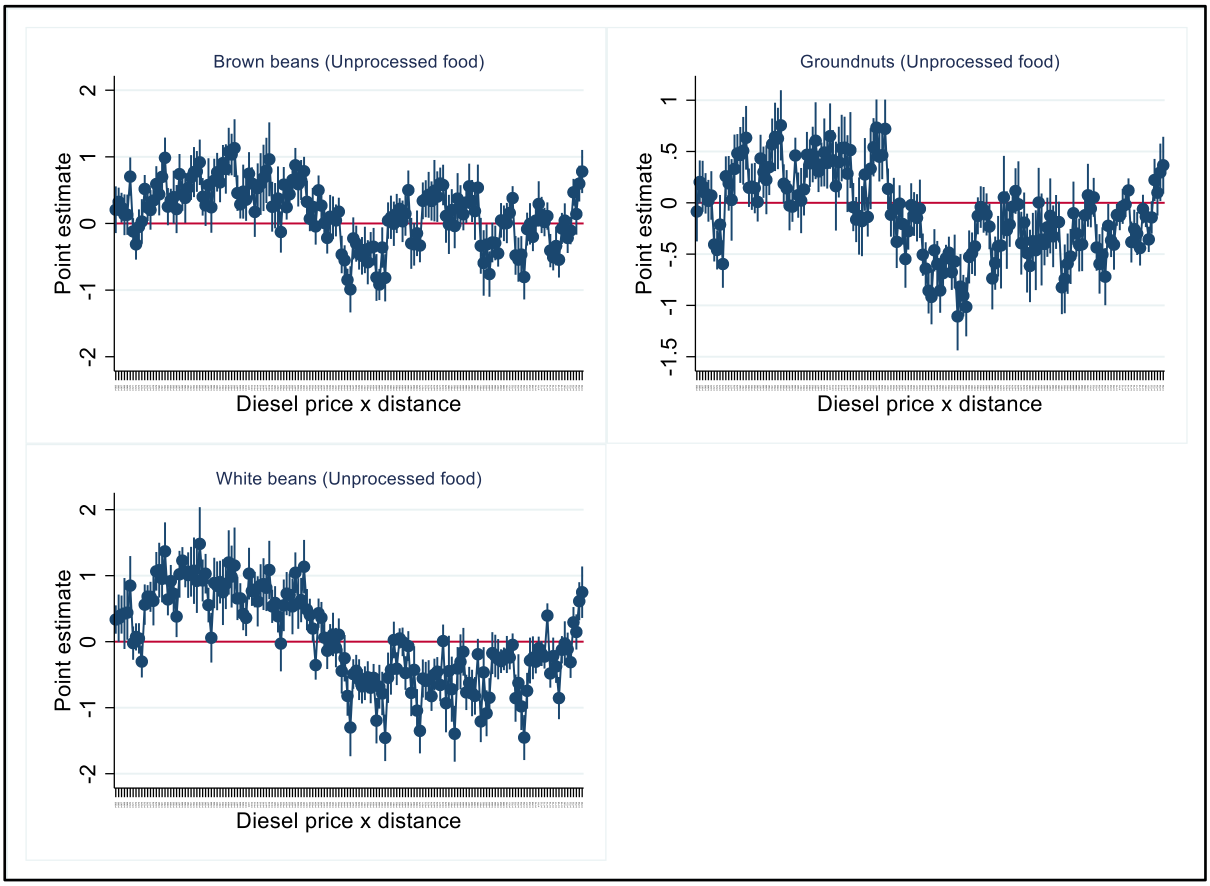

Figure A.11: Time dummies after fixed effects Poisson estimator for legumes and nuts

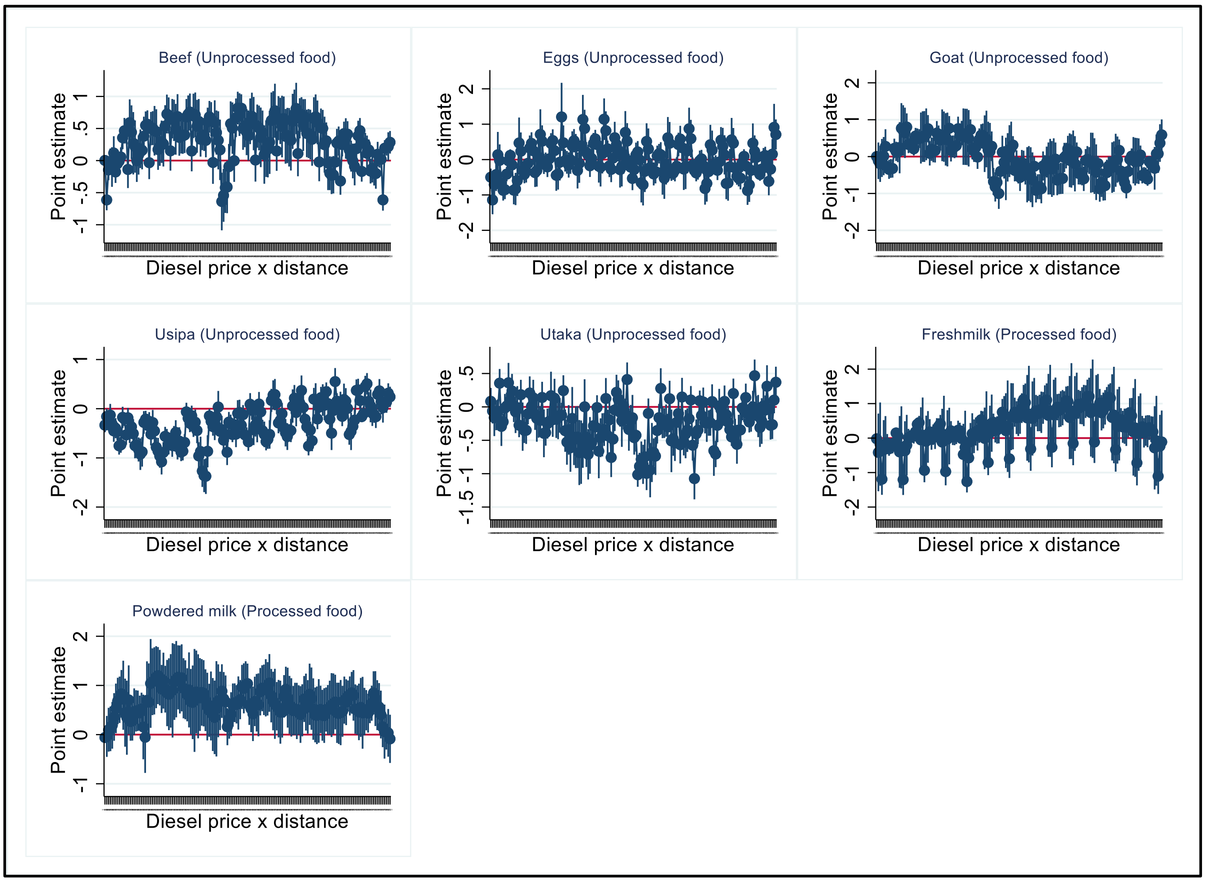

Figure A.12: Time dummies after fixed effects Poisson estimator for animal source foods

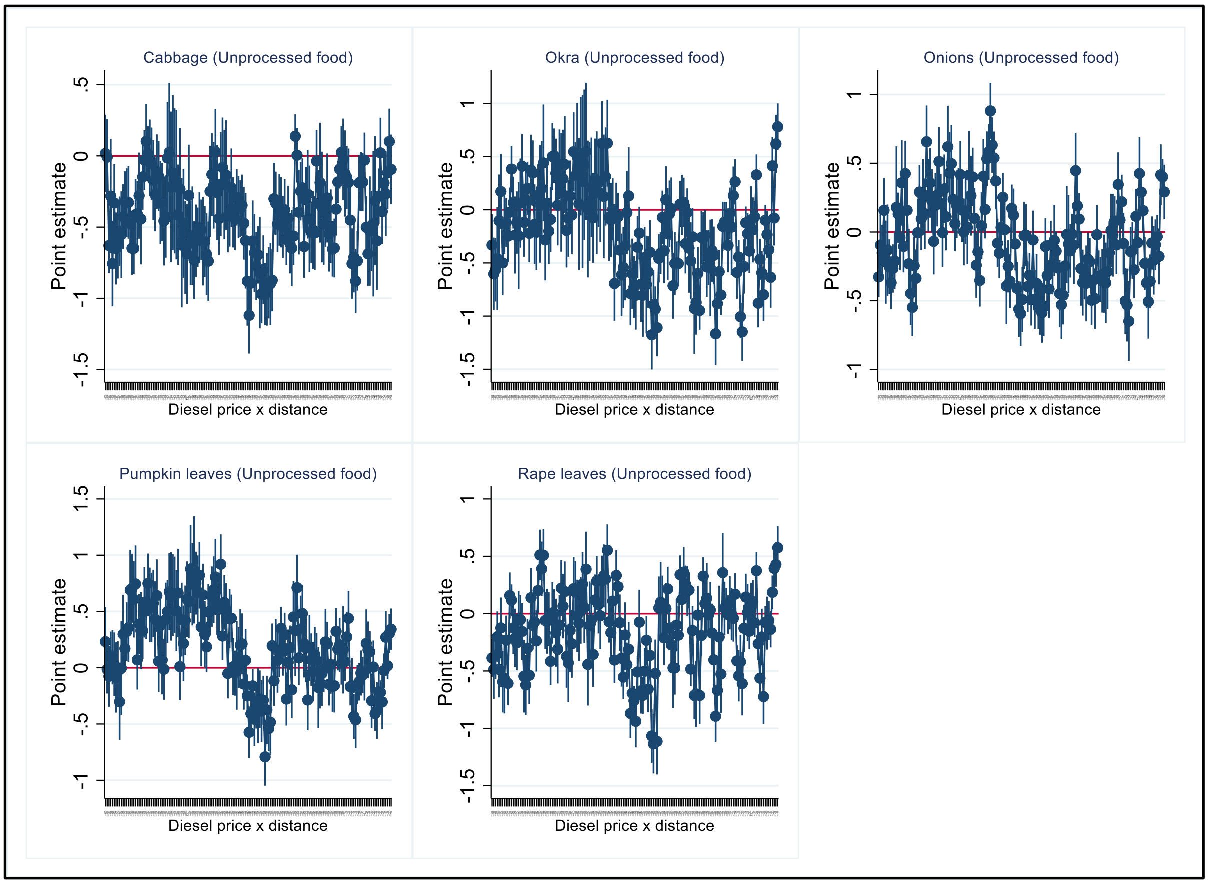

Figure A.13: Time dummies after fixed effects Poisson estimator for vegetables

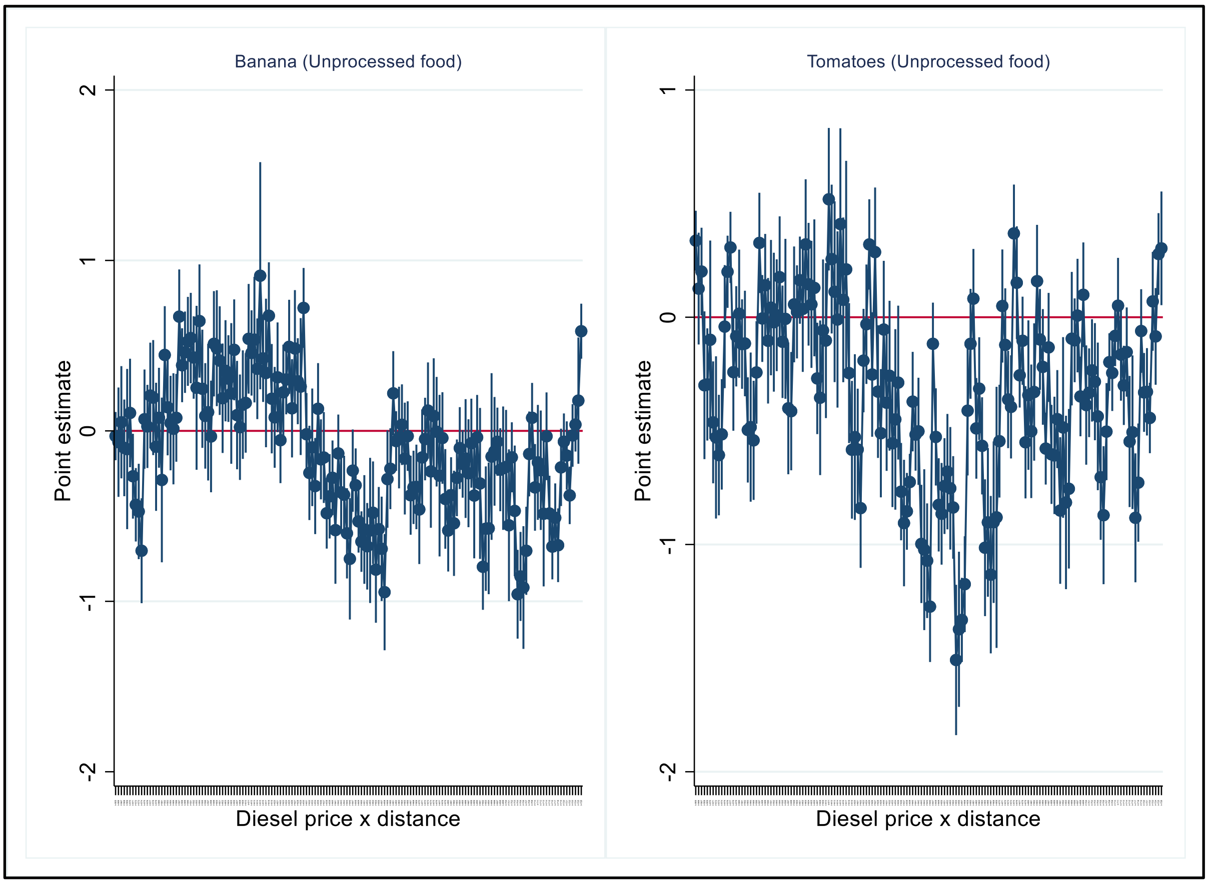

Figure A.14: Time dummies after fixed effects Poisson estimator for fruits

Table A.7: The association between transport costs and market price dispersion of staples including roots and tubers

Dependent variable: $\|P_{xt} - P_{yt}\|$	Cassava	Maize flour dehulled	Maize grain (private)	Maize grain (ADMARC)	Potatoes	Rice grain	Sweet potatoes	White bread	White buns
Diesel price x distance	-7.183**	16.39***	1.851	5.968***	1.896	-1.587	-2.584	-0.999	10.50
	(3.280)	(4.894)	(1.265)	(1.938)	(3.725)	(4.431)	(2.255)	(1.650)	(9.694)
Distance-specific linear time trends	-19.51***	2.723	-1.583	5.157	-12.66**	15.24*	-5.077	-0.928	-2.168
	(4.750)	(14.63)	(2.368)	(3.406)	(5.686)	(7.902)	(3.420)	(3.463)	(22.30)
Distance-specific quadratic time trends	0.0305***	-0.00341	0.00237	-0.00842	0.0200**	-0.0229*	0.00809	0.00138	0.00344
	(0.00750)	(0.0228)	(0.00372)	(0.00538)	(0.00883)	(0.0122)	(0.00537)	(0.00541)	(0.0339)
Distance-specific cubic time trends	-0.0000158***	0.00000129	-0.00000120	0.00000453	-0.0000105**	0.0000114*	-0.00000425	-0.000000678	-0.00000183
	(0.00000392)	(0.0000118)	(0.00000194)	(0.00000282)	(0.00000454)	(0.00000629)	(0.00000279)	(0.00000280)	(0.0000171)
Difference in population density	-1.174	-5.121	0.192	-0.369	-1.720	3.413	-0.807	3.672*	-58.87***
	(3.587)	(11.06)	(2.296)	(1.411)	(5.203)	(3.296)	(0.897)	(2.144)	(12.14)
=1 if one of the markets experienced flood shocks	1.492*	0.874	1.059	0.0423	-0.571	-0.299	0.590	0.458	-3.289
	(0.810)	(1.962)	(0.693)	(0.579)	(0.952)	(1.568)	(0.817)	(0.366)	(2.334)
=1 if one of the markets experienced drought shocks	-1.977**	-3.028	0.379	-0.743	1.914**	-5.366***	-0.195	1.101	-6.169
	(0.861)	(2.171)	(0.526)	(0.570)	(0.838)	(1.338)	(0.538)	(0.755)	(4.485)
Difference in local production	21.53	141.5	46.35*	26.21	5.848	35.49	-11.99	-	-
	(50.65)	(132.3)	(25.81)	(32.16)	(20.05)	(22.53)	(36.21)	-	-
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Market-pair FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	44023	57195	79971	70099	77243	69666	59931	76513	58832

Note: Linear regression absorbing multiple levels of fixed effects (reghdfe) estimator results for each food items are presented in each column. The dependent variable is absolute value of price difference. Route distance over paved road between the market pairs is measured in hundreds of kilometres, and diesel fuel price in thousands of Malawian Kwacha. Dyadic clustered standard errors at the market in parentheses * p<0.10, ** p<0.05, *** p<0.010

Table A.8: The association between transport costs and market price dispersion of legumes and nuts

Dependent variable: $\lvert P_{xt} - P_{yt} \rvert$	Brown beans	Groundnuts	White beans
Diesel price x distance	13.54**	-11.26	-0.526
	(5.864)	(6.942)	(6.375)
Distance-specific linear time trends	21.29**	-9.914	-13.53
	(8.596)	(9.432)	(8.816)
Distance-specific quadratic time trends	-0.0333**	0.0164	0.0216
	(0.0136)	(0.0149)	(0.0138)
Distance-specific cubic time trends	0.0000172**	-0.00000897	-0.0000115
	(0.00000711)	(0.00000783)	(0.00000716)
Difference in population density	-17.03	-18.28	-21.10
	(13.31)	(17.06)	(18.40)
=1 if one of the markets experienced flood shocks	1.633	3.272	1.799
	(1.947)	(2.100)	(1.777)
=1 if one of the markets experienced drought shocks	-5.715***	-5.100**	-3.024*
	(2.073)	(2.186)	(1.769)
Difference in local production	-28.16	-7.016	90.01
	(152.7)	(10.34)	(94.52)
Month FE	Yes	Yes	Yes
Time FE	Yes	Yes	Yes
Market-pair FE	Yes	Yes	Yes
N	79826	67104	60546

Note: Linear regression absorbing multiple levels of fixed effects (reghdfe) estimator results for each food items are presented in each column. The dependent variable is absolute value of price difference. Route distance over paved road between the market pairs is measured in hundreds of kilometres, and diesel fuel price in thousands of Malawian Kwacha. Dyadic clustered standard errors at the market in parentheses * p<0.10, ** p<0.05, *** p<0.010

Table A.9: The association between transport costs and market price dispersion of animal source foods

Dependent variable: $\|P_{xt} - P_{yt}\|$	Beef	Eggs	Ultra-pasteurized milk	Goat meat	Powdered milk	Usipa	Utaka
Diesel price x distance	-21.76***	21.04*	1.708	-12.20*	-239.2**	27.98	-120.4**
	(7.786)	(11.53)	(2.774)	(6.442)	(96.13)	(48.95)	(52.96)
Distance-specific linear time trends	18.69*	-1.427	0.714	8.602	-182.5	32.55	-180.8***
	(9.169)	(12.05)	(7.016)	(11.23)	(168.9)	(85.81)	(58.59)
Distance-specific quadratic time trends	-0.0309**	0.00173	-0.00142	-0.0139	0.284	-0.0510	0.289***
	(0.0146)	(0.0189)	(0.0112)	(0.0178)	(0.268)	(0.135)	(0.0905)
Distance-specific cubic time trends	0.0000171**	-0.000000728	0.000000908	0.00000750	-0.000144	0.0000266	-0.000153***
	(0.00000779)	(0.00000980)	(0.00000589)	(0.00000935)	(0.000141)	(0.0000700)	(0.0000464)
Difference in population density	0.834	1.728	5.144	3.267	200.7**	-55.27	-135.5
	(10.64)	(4.320)	(5.717)	(9.423)	(95.11)	(78.11)	(101.7)
=1 if one of the markets experienced flood shocks	1.376	-2.191	-0.638	1.258	-58.51***	24.12***	3.710
	(3.257)	(1.713)	(0.677)	(2.375)	(16.03)	(8.230)	(16.68)
=1 if one of the markets experienced drought shocks	-0.492	4.160	2.073	-0.251	19.98	-18.78*	10.77
	(3.535)	(2.464)	(1.401)	(2.178)	(31.01)	(10.18)	(16.82)
Difference in local production	433.6	-267.5	-	542.2	-	34.41	281.4**
	(1000.0)	(475.8)	-	(328.9)	-	(316.5)	(109.5)
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Market-pair FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	74127	71621	62329	74588	40866	78279	56714

Note: Linear regression absorbing multiple levels of fixed effects (reghdfe) estimator results for each food items are presented in each column. The dependent variable is absolute value of price difference. Route distance over paved road between the market pairs is measured in hundreds of kilometres, and diesel fuel price in thousands of Malawian Kwacha. Dyadic clustered standard errors at the market in parentheses * p<0.10, ** p<0.05, *** p<0.010

Table A.10: The association between transport costs and market price dispersion of vegetables

Dependent variable: $\lvert P_{xt} - P_{yt} \rvert$	Cabbage	Okra	Onions	Pumpkin leaves	Rape leaves
Diesel price x distance	-1.039 (2.624)	8.277 (13.59)	1.024 (5.122)	4.049 (3.380)	0.517 (2.192)
Distance-specific linear time trends	-0.950 (3.416)	-0.609 (15.25)	-0.457 (8.918)	21.11* (11.99)	2.452 (4.063)
Distance-specific quadratic time trends	0.00194 (0.00532)	0.00166 (0.0236)	0.00120 (0.0138)	-0.0336* (0.0190)	-0.00380 (0.00631)
Distance-specific cubic time trends	-0.00000123 (0.00000274)	-0.00000124 (0.0000120)	-0.000000871 (0.00000710)	0.0000177* (0.00000994)	0.00000197 (0.00000324)
Difference in population density	-3.534* (1.738)	0.700 (8.749)	-13.94* (7.873)	9.355* (5.126)	0.800 (4.565)
=1 if one of the markets experienced flood shocks	0.508 (0.475)	1.830 (2.636)	-2.280 (1.722)	-0.436 (1.295)	1.485 (1.133)
=1 if one of the markets experienced drought shocks	2.027** (0.751)	3.685 (2.275)	8.995*** (2.541)	-2.686* (1.426)	0.706 (1.040)
Difference in local production	-0.939 (62.06)	226.5 (137.4)	-1113.7* (552.6)	145.6 (108.4)	111.1* (56.85)
Month FE	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes
Market-pair FE	Yes	Yes	Yes	Yes	Yes
N	72133	53953	80090	64588	74680

Note: Linear regression absorbing multiple levels of fixed effects (reghdfe) estimator results for each food items are presented in each column. The dependent variable is absolute value of price difference. Route distance over paved road between the market pairs is measured in hundreds of kilometres, and diesel fuel price in thousands of Malawian Kwacha. Dyadic clustered standard errors at the market in parentheses * p<0.10, ** p<0.05, *** p<0.010

Table A.11: The association between transport costs and market price dispersion of fruits

Dependent variable: $\lvert P_{xt} - P_{yt} \rvert$	Bananas	Tomatoes
Diesel price x distance	-3.978*	-1.401
	(2.008)	(2.939)
Distance-specific linear time trends	-0.748	-5.147
	(6.311)	(6.877)
Distance-specific quadratic time trends	0.00109	0.00879
	(0.00982)	(0.0108)
Distance-specific cubic time trends	-0.000000480	-0.00000494
	(0.00000507)	(0.00000561)
Difference in population density	-2.025	-2.224
	(3.225)	(5.578)
=1 if one of the markets experienced flood shocks	0.470	-0.0747
	(1.078)	(1.845)
=1 if one of the markets experienced drought shocks	1.049	0.502
	(0.698)	(1.694)
Difference in local production	-139.3	-768.1***
	(905.5)	(192.7)
Month FE	Yes	Yes
Time FE	Yes	Yes
Market-pair FE	Yes	Yes
N	66053	80558

Note: Linear regression absorbing multiple levels of fixed effects (reghdfe) estimator results for each food items are presented in each column. The dependent variable is absolute value of price difference. Route distance over paved road between the market pairs is measured in hundreds of kilometres, and diesel fuel price in thousands of Malawian Kwacha. Dyadic clustered standard errors at the market in parentheses * p<0.10, ** p<0.05, *** p<0.010

Table A.12: The association between transport costs and market price dispersion of staples including roots and tubers

Dependent variable: $\lvert P_{xt} - P_{yt} \rvert$	Cassava	Maize flour dehulled	Maize grain (private)	Maize grain (ADMARC)	Potatoes	Rice grain	Sweet potatoes	White bread	White buns
Diesel price x distance	-0.0574	0.0829***	0.0748**	0.323***	0.0364	-0.0978**	-0.0501	-0.0388	0.0489
	(0.0590)	(0.0321)	(0.0291)	(0.104)	(0.0450)	(0.0470)	(0.0627)	(0.0562)	(0.0451)
Distance-specific linear time trends	-0.283***	0.126	-0.0904	0.355**	-0.0843	0.227***	-0.0341	0.00667	0.0546
	(0.0708)	(0.0831)	(0.0583)	(0.164)	(0.0703)	(0.0781)	(0.0703)	(0.113)	(0.117)
Distance-specific quadratic time trends	0.000437***	-0.000186	0.000132	-0.000570**	0.000135	-0.000339***	0.0000606	-0.0000107	-0.0000812
	(0.000110)	(0.000129)	(0.0000888)	(0.000258)	(0.000108)	(0.000120)	(0.000110)	(0.000174)	(0.000178)
Distance-specific cubic time trends	-0.000000224***	9.09e-08	-6.49e-08	0.000000302**	-7.21e-08	0.000000168***	-3.47e-08	6.00e-09	3.99e-08
	(5.64e-08)	(6.61e-08)	(4.49e-08)	(0.000000135)	(5.50e-08)	(6.13e-08)	(5.67e-08)	(8.92e-08)	(8.96e-08)
Difference in population density	-0.0213	-0.0464	-0.00916	0.000927	-0.0527	0.0181	-0.0262	0.0675	-0.266***
	(0.0638)	(0.0727)	(0.0445)	(0.0533)	(0.0580)	(0.0343)	(0.0276)	(0.0608)	(0.0747)
=1 if one of the markets experienced flood shocks	0.0151	0.00543	0.0287*	0.0100	-0.00832	0.00000164	0.0228	0.00951	-0.0162
	(0.0123)	(0.0117)	(0.0166)	(0.0269)	(0.0138)	(0.0156)	(0.0198)	(0.0150)	(0.0118)
=1 if one of the markets experienced drought shocks	-0.0485***	-0.0220	0.00643	-0.0419	0.0242**	-0.0538***	-0.0104	0.0345	-0.0244
	(0.0181)	(0.0152)	(0.0148)	(0.0322)	(0.0105)	(0.0132)	(0.0137)	(0.0226)	(0.0241)
Difference in local production	-0.886	1.238	0.556	1.075	-1.315***	0.513	-0.854	-	-
	(1.009)	(0.899)	(0.539)	(1.239)	(0.452)	(0.329)	(1.058)	-	-
Market-Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Market-pair FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	44023	57195	79971	64135	77243	69666	59931	76513	58832

Note: Poisson pseudo-likelihood regression with multiple levels of fixed effects (ppmlhdfe) estimator results for each food items are presented in each column. The dependent variable is absolute value of price difference. Route distance over paved road between the market pairs is measured in hundreds of kilometres, and diesel fuel price in thousands of Malawian Kwacha. Dyadic clustered standard errors at the market in parentheses * p<0.10, ** p<0.05, *** p<0.010

Table A.13: The association between transport costs and market price dispersion of legumes and nuts

Dependent variable: $\lvert P_{xt} - P_{yt} \rvert$	Brown beans	Groundnuts	White beans
Diesel price x distance	0.0885**	-0.0487	0.0288
	(0.0388)	(0.0397)	(0.0422)
Distance-specific linear time trends	0.121**	-0.0238	-0.0523
	(0.0502)	(0.0434)	(0.0563)
Distance-specific quadratic time trends	-0.000188**	0.0000411	0.0000826
	(0.0000777)	(0.0000672)	(0.0000871)
Distance-specific cubic time trends	9.64e-08**	-2.33e-08	-4.36e-08
	(4.00e-08)	(3.46e-08)	(4.47e-08)
Difference in population density	-0.0662	-0.0489	-0.0781
	(0.0547)	(0.0567)	(0.0906)
=1 if one of the markets experienced flood shocks	-0.00275	0.00700	0.000555
	(0.0103)	(0.0118)	(0.0110)
=1 if one of the markets experienced drought shocks	-0.0340***	-0.0195	-0.0273**
	(0.0132)	(0.0131)	(0.0137)
Difference in local production	0.897*	-0.265***	1.341**
	(0.512)	(0.0740)	(0.527)
Market-Month FE	Yes	Yes	Yes
Time FE	Yes	Yes	Yes
Market-pair FE	Yes	Yes	Yes
N	79826	67104	60546

Note: Poisson pseudo-likelihood regression with multiple levels of fixed effects (ppmlhdfe) estimator results for each food items are presented in each column. The dependent variable is absolute value of price difference. Route distance over paved road between the market pairs is measured in hundreds of kilometres, and diesel fuel price in thousands of Malawian Kwacha. Dyadic clustered standard errors at the market in parentheses * p<0.10, ** p<0.05, *** p<0.010

Table A.14: The association between transport costs and market price dispersion of animal source foods

Dependent variable: $\lvert P_{xt} - P_{yt} \rvert$	Beef	Eggs	Ultra-pasteurized milk	Goat meat	Powdered milk	Usipa	Utaka
Diesel price x distance	-0.0862**	0.108**	0.0383	-0.105***	-0.207**	0.0409	-0.0791
	(0.0341)	(0.0544)	(0.0377)	(0.0295)	(0.0826)	(0.0470)	(0.0638)
Distance-specific linear time trends	0.0118	0.00423	0.155	-0.0349	0.00512	0.0503	-0.0922
	(0.0357)	(0.0761)	(0.116)	(0.0615)	(0.117)	(0.0743)	(0.0610)
Distance-specific quadratic time trends	-0.0000298	-0.00000742	-0.000241	0.0000502	-0.00000250	-0.0000764	0.000154*
	(0.0000559)	(0.000118)	(0.000180)	(0.0000950)	(0.000183)	(0.000116)	(0.0000932)
Distance-specific cubic time trends	2.18e-08	3.95e-09	0.000000124	-2.37e-08	2.87e-10	3.86e-08	-8.40e-08*
	(2.91e-08)	(6.12e-08)	(9.35e-08)	(4.87e-08)	(9.53e-08)	(6.03e-08)	(4.73e-08)
Difference in population density	-0.0231	-0.00511	0.0242	-0.000880	0.196**	-0.0237	-0.113
	(0.0555)	(0.0290)	(0.0953)	(0.0234)	(0.0772)	(0.0825)	(0.0896)
=1 if one of the markets experienced flood shocks	-0.00673	-0.0250**	-0.0174	0.00650	-0.0495***	0.0348***	0.000465
	(0.0139)	(0.0113)	(0.0125)	(0.0123)	(0.0148)	(0.00809)	(0.0133)
=1 if one of the markets experienced drought shocks	0.00820	0.0275*	0.0345	-0.00831	-0.000130	-0.0179	0.00984
	(0.0162)	(0.0144)	(0.0229)	(0.0137)	(0.0262)	(0.0125)	(0.0152)
Difference in local production	1.320	-1.794	-	2.303*	-	0.0859	0.111
	(4.669)	(2.666)	-	(1.281)	-	(0.245)	(0.0978)
Market-Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Market-pair FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	74127	71621	62329	74588	40866	78279	56714

Note: Poisson pseudo-likelihood regression with multiple levels of fixed effects (ppmlhdfe) estimator results for each food items are presented in each column. The dependent variable is absolute value of price difference. Route distance over paved road between the market pairs is measured in hundreds of kilometres, and diesel fuel price in thousands of Malawian Kwacha. Dyadic clustered standard errors at the market in parentheses * p<0.10, ** p<0.05, *** p<0.010

Table A.15: The association between transport costs and market price dispersion of vegetables

Dependent variable: $\lvert P_{xt} - P_{yt} \rvert$	Cabbage	Okra	Onions	Pumpkin leaves	Rape leaves
Diesel price x distance	-0.0493 (0.0600)	0.0190 (0.0734)	0.00961 (0.0318)	0.00168 (0.0290)	0.0179 (0.0343)
Distance-specific linear time trends	0.0936 (0.0828)	0.0199 (0.0950)	0.0237 (0.0491)	0.144** (0.0673)	0.0726 (0.0546)
Distance-specific quadratic time trends	-0.000134 (0.000127)	-0.0000304 (0.000146)	-0.0000334 (0.0000757)	-0.000227** (0.000106)	-0.000111 (0.0000841)
Distance-specific cubic time trends	6.30e-08 (6.47e-08)	1.51e-08 (7.42e-08)	1.55e-08 (3.87e-08)	0.000000119** (5.51e-08)	5.62e-08 (4.29e-08)
Difference in population density	-0.0731 (0.0491)	-0.00835 (0.0369)	-0.0885* (0.0528)	0.0872* (0.0494)	0.0102 (0.0588)
=1 if one of the markets experienced flood shocks	0.0170 (0.0119)	0.00541 (0.0153)	-0.0180* (0.0106)	-0.0109 (0.0107)	0.0137 (0.0135)
=1 if one of the markets experienced drought shocks	0.0634*** (0.0167)	0.0216 (0.0152)	0.0571*** (0.0160)	-0.0323** (0.0129)	0.00583 (0.0133)
Difference in local production	0.00132 (1.134)	1.967 (2.277)	-4.232* (2.262)	2.799 (2.067)	3.721** (1.615)
Market-Month FE	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes
Market-pair FE	Yes	Yes	Yes	Yes	Yes
N	72133	53953	80090	64588	74680

Note: Poisson pseudo-likelihood regression with multiple levels of fixed effects (ppmlhdfe) estimator results for each food items are presented in each column. The dependent variable is absolute value of price difference. Route distance over paved road between the market pairs is measured in hundreds of kilometres, and diesel fuel price in thousands of Malawian Kwacha. Dyadic clustered standard errors at the market in parentheses * p<0.10, ** p<0.05, *** p<0.010

Table A.16: The association between transport costs and market price dispersion of fruits

Dependent variable: $\lvert P_{xt} - P_{yt} \rvert$	Bananas	Tomatoes
Diesel price x distance	-0.0911***	-0.00904
	(0.0229)	(0.0194)
Distance-specific linear time trends	-0.00962	0.0557
	(0.0838)	(0.0488)
Distance-specific quadratic time trends	0.0000173	-0.0000780
	(0.000129)	(0.0000748)
Distance-specific cubic time trends	-9.68e-09	3.59e-08
	(6.60e-08)	(3.81e-08)
Difference in population density	-0.0302	-0.0119
	(0.0436)	(0.0391)
=1 if one of the markets experienced flood shocks	0.0140	-0.00181
	(0.0147)	(0.0138)
=1 if one of the markets experienced drought shocks	0.0139	-0.00365
	(0.0128)	(0.0126)
Difference in local production	0.672	-4.048***
	(6.908)	(1.114)
Market-Month FE	Yes	Yes
Time FE	Yes	Yes
Market-pair FE	Yes	Yes
N	66053	80558

Note: Poisson pseudo-likelihood regression with multiple levels of fixed effects (ppmlhdfe) estimator results for each food items are presented in each column. The dependent variable is absolute value of price difference. Route distance over paved road between the market pairs is measured in hundreds of kilometres, and diesel fuel price in thousands of Malawian Kwacha. Dyadic clustered standard errors at the market in parentheses * p<0.10, ** p<0.05, *** p<0.010

Table A.17: The association between transport costs and market price dispersion of staples including roots and tubers

Dependent variable: $\lvert P_{xt} - P_{yt} \rvert$	Cassava	Maize flour dehulled	Maize grain (private)	Maize grain (ADMARC)	Potatoes	Rice grain	Sweet potatoes	White bread	White buns
Diesel price x distance	-0.142***	0.0809***	0.0684**	0.284***	0.0119	-0.0402	-0.0539	-0.0414	0.0728
	(0.0537)	(0.0298)	(0.0297)	(0.109)	(0.0434)	(0.0612)	(0.0353)	(0.0482)	(0.0446)
Distance-specific linear time trends	-0.311***	0.149*	-0.0711	0.272*	-0.0829	0.223***	-0.108	0.00960	0.0483
	(0.0743)	(0.0898)	(0.0579)	(0.148)	(0.0696)	(0.0613)	(0.0815)	(0.103)	(0.0992)
Distance-specific quadratic time trends	0.000482***	-0.000219	0.000103	-0.000444*	0.000134	-0.000334***	0.000173	-0.0000163	-0.0000731
	(0.000115)	(0.000138)	(0.0000884)	(0.000234)	(0.000107)	(0.0000939)	(0.000126)	(0.000160)	(0.000150)
Distance-specific cubic time trends	-0.000000247***	0.000000107	-4.98e-08	0.000000239*	-7.20e-08	0.000000166***	-9.17e-08	9.36e-09	3.66e-08
	(5.91e-08)	(7.03e-08)	(4.49e-08)	(0.000000123)	(5.42e-08)	(4.77e-08)	(6.44e-08)	(8.22e-08)	(7.50e-08)
Difference in population density	0.0206	-0.0314	0.00263	-0.0104	-0.0415	0.00504	-0.00414	0.0721	-0.251***
	(0.0525)	(0.0647)	(0.0424)	(0.0522)	(0.0456)	(0.0331)	(0.0309)	(0.0561)	(0.0733)
=1 if one of the markets experienced flood shocks	0.00763	-0.00528	0.0120	-0.00284	-0.0179	0.00172	0.0135	0.00771	-0.0160
	(0.00944)	(0.00962)	(0.0162)	(0.0246)	(0.0123)	(0.0154)	(0.0184)	(0.0146)	(0.0109)
=1 if one of the markets experienced drought shocks	-0.0298***	-0.0347***	0.0202	-0.0257	0.0228**	-0.0372***	-0.00382	0.0388*	-0.0139
	(0.00993)	(0.0102)	(0.0162)	(0.0256)	(0.0104)	(0.0128)	(0.0112)	(0.0223)	(0.0208)
Difference in local production	-0.158	0.726	0.536	1.155	-1.513***	0.293	-0.903	-	-
	(0.962)	(0.853)	(0.511)	(1.211)	(0.513)	(0.294)	(0.738)	-	-
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Market-pair FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	75376	70455	82840	72118	82840	76720	80032	80158	71158

Note: Poisson pseudo-likelihood regression with multiple levels of fixed effects (ppmlhdfe) estimator results for each food items are presented in each column. The dependent variable is absolute value of price difference with missing price differences mapped to their maximum price differences in each year. Route distance over paved road between the market pairs is measured in hundreds of kilometres, and diesel fuel price in thousands of Malawian Kwacha. Dyadic clustered standard errors at the market in parentheses * p<0.10, ** p<0.05, *** p<0.010

Table A.18: The association between transport costs and market price dispersion of legumes and nuts

Dependent variable: $\lvert P_{xt} - P_{yt} \rvert$	Brown beans	Groundnuts	White beans
Diesel price x distance	0.0547	-0.0606*	-0.0343
	(0.0346)	(0.0332)	(0.0323)
Distance-specific linear time trends	0.0887*	-0.0704	0.0165
	(0.0497)	(0.0519)	(0.0626)
Distance-specific quadratic time trends	-0.000137*	0.000113	-0.0000195
	(0.0000767)	(0.0000807)	(0.0000961)
Distance-specific cubic time trends	6.96e-08*	-5.97e-08	7.08e-09
	(3.93e-08)	(4.17e-08)	(4.90e-08)
Difference in population density	-0.0657	-0.0433	-0.00526
	(0.0528)	(0.0466)	(0.0661)
=1 if one of the markets experienced flood shocks	0.00543	-0.00626	0.00751
	(0.0106)	(0.00975)	(0.0135)
=1 if one of the markets experienced drought shocks	-0.0144	-0.00217	-0.0157
	(0.0140)	(0.0170)	(0.0223)
Difference in local production	0.859	-0.223***	1.292**
	(0.535)	(0.0815)	(0.503)
Month FE	Yes	Yes	Yes
Time FE	Yes	Yes	Yes
Market-pair FE	Yes	Yes	Yes
N	82840	80013	78215

Note: Poisson pseudo-likelihood regression with multiple levels of fixed effects (ppmlhdfe) estimator results for each food items are presented in each column. The dependent variable is absolute value of price difference with missing price differences mapped to their maximum price differences in each year. Route distance over paved road between the market pairs is measured in hundreds of kilometres, and diesel fuel price in thousands of Malawian Kwacha. Dyadic clustered standard errors at the market in parentheses * p<0.10, ** p<0.05, *** p<0.010

Table A.19: The association between transport costs and market price dispersion of animal source foods

Dependent variable: $\lvert P_{xt} - P_{yt} \rvert$	Beef	Eggs	Ultra-pasteurized milk	Goat meat	Powdered milk	Usipa	Utaka
Diesel price x distance	-0.0986***	0.105**	0.0261	-0.118***	-0.127*	0.000233	-0.0805
	(0.0354)	(0.0497)	(0.0361)	(0.0281)	(0.0739)	(0.0393)	(0.0623)
Distance-specific linear time trends	0.00289	0.0231	0.152	-0.0444	0.0898	-0.0248	-0.0438
	(0.0346)	(0.0684)	(0.108)	(0.0616)	(0.107)	(0.0504)	(0.0972)
Distance-specific quadratic time trends	-0.0000161	-0.0000380	-0.000236	0.0000650	-0.000141	0.0000407	0.0000783
	(0.0000543)	(0.000107)	(0.000169)	(0.0000953)	(0.000166)	(0.0000790)	(0.000148)
Distance-specific cubic time trends	1.47e-08	2.03e-08	0.000000122	-3.12e-08	7.49e-08	-2.18e-08	-4.48e-08
	(2.84e-08)	(5.50e-08)	(8.77e-08)	(4.89e-08)	(8.60e-08)	(4.10e-08)	(7.47e-08)
Difference in population density	-0.0219	-0.0243	0.0295	0.00455	0.132	-0.0184	-0.0931
	(0.0556)	(0.0303)	(0.0872)	(0.0243)	(0.0884)	(0.0742)	(0.0736)
=1 if one of the markets experienced flood shocks	-0.00398	-0.00284	-0.0120	0.0133	-0.0303**	0.0135	-0.0106
	(0.0135)	(0.0120)	(0.0121)	(0.0107)	(0.0125)	(0.0120)	(0.00789)
=1 if one of the markets experienced drought shocks	0.00404	0.0394**	0.0281	-0.00287	0.0247	-0.0376***	-0.00489
	(0.0166)	(0.0159)	(0.0187)	(0.0138)	(0.0166)	(0.0105)	(0.0182)
Difference in local production	1.220	-3.677	-	2.430*	-	0.0248	0.126
	(4.715)	(2.783)	-	(1.294)	-	(0.208)	(0.0832)
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Market-pair FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	77170	79334	70886	78286	56386	83212	74113

Note: Poisson pseudo-likelihood regression with multiple levels of fixed effects (ppmlhdfe) estimator results for each food items are presented in each column. The dependent variable is absolute value of price difference with missing price differences mapped to their maximum price differences in each year. Route distance over paved road between the market pairs is measured in hundreds of kilometres, and diesel fuel price in thousands of Malawian Kwacha. Dyadic clustered standard errors at the market in parentheses * p<0.10, ** p<0.05, *** p<0.010

Table A.20: The association between transport costs and market price dispersion of vegetables

Dependent variable: $\lvert P_{xt} - P_{yt} \rvert$	Cabbage	Okra	Onions	Pumpkin leaves	Rape leaves
Diesel price x distance	-0.103 (0.0661)	0.0144 (0.0539)	0.00796 (0.0257)	-0.0630** (0.0280)	0.0158 (0.0347)
Distance-specific linear time trends	0.0511 (0.0911)	0.0130 (0.0756)	0.0293 (0.0503)	0.0904 (0.0861)	0.0822 (0.0560)
Distance-specific quadratic time trends	-0.0000667 (0.000141)	-0.0000187 (0.000115)	-0.0000427 (0.0000777)	-0.000141 (0.000135)	-0.000126 (0.0000860)
Distance-specific cubic time trends	2.82e-08 (7.19e-08)	8.67e-09 (5.83e-08)	2.06e-08 (3.98e-08)	7.40e-08 (7.03e-08)	6.39e-08 (4.38e-08)
Difference in population density	-0.0724* (0.0413)	-0.0186 (0.0351)	-0.0786 (0.0504)	0.0764* (0.0408)	0.0205 (0.0566)
=1 if one of the markets experienced flood shocks	0.0103 (0.0141)	-0.0175* (0.00969)	-0.0140 (0.0103)	-0.00160 (0.0118)	0.0246* (0.0147)
=1 if one of the markets experienced drought shocks	0.0546*** (0.0184)	0.0195 (0.0147)	0.0509*** (0.0172)	-0.00723 (0.0117)	0.0135 (0.0143)
Difference in local production	1.782 (1.415)	3.222 (2.105)	-4.789*** (1.800)	1.682 (1.890)	2.475 (1.633)
Month FE	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes
Market-pair FE	Yes	Yes	Yes	Yes	Yes
N	81705	74088	83212	79492	81700

Note: Poisson pseudo-likelihood regression with multiple levels of fixed effects (ppmlhdfe) estimator results for each food items are presented in each column. The dependent variable is absolute value of price difference with missing price differences mapped to their maximum price differences in each year. Route distance over paved road between the market pairs is measured in hundreds of kilometres, and diesel fuel price in thousands of Malawian Kwacha. Dyadic clustered standard errors at the market in parentheses * p<0.10, ** p<0.05, *** p<0.010

Table A.21: The association between transport costs and market price dispersion of fruits

Dependent variable: $\lvert P_{xt} - P_{yt} \rvert$	Bananas	Tomatoes
Diesel price x distance	-0.0715***	-0.0109
	(0.0253)	(0.0215)
Distance-specific linear time trends	0.0286	0.0588
	(0.0827)	(0.0527)
Distance-specific quadratic time trends	-0.0000412	-0.0000832
	(0.000128)	(0.0000812)
Distance-specific cubic time trends	2.01e-08	3.88e-08
	(6.55e-08)	(4.16e-08)
Difference in population density	-0.0338	-0.00254
	(0.0323)	(0.0389)
=1 if one of the markets experienced flood shocks	0.0135	-0.00333
	(0.0133)	(0.0143)
=1 if one of the markets experienced drought shocks	0.0267*	0.00462
	(0.0141)	(0.0148)
Difference in local production	-1.349	-4.646***
	(6.950)	(1.298)
Month FE	Yes	Yes
Time FE	Yes	Yes
Market-pair FE	Yes	Yes
N	80442	83212

Note: Poisson pseudo-likelihood regression with multiple levels of fixed effects (ppmlhdfe) estimator results for each food items are presented in each column. The dependent variable is absolute value of price difference with missing price differences mapped to their maximum price differences in each year. Route distance over paved road between the market pairs is measured in hundreds of kilometres, and diesel fuel price in thousands of Malawian Kwacha. Dyadic clustered standard errors at the market in parentheses * p<0.10, ** p<0.05, *** p<0.010

Table A.22: The association between transport costs and market price dispersion of staples including roots and tubers

Dependent variable: $\lvert P_{xt} - P_{yt} \rvert$	Cassava	Maize flour dehulled	Maize grain (private)	Maize grain (ADMARC)	Potatoes	Rice grain	Sweet potatoes	White bread	White buns
Diesel price x distance	-0.178**	0.275**	0.154	0.366***	-0.0368	-0.143**	-0.0253	0.106	-0.322**
	(0.0741)	(0.120)	(0.101)	(0.0959)	(0.0881)	(0.0670)	(0.0879)	(0.105)	(0.158)
Diesel price x (distance)^2	0.00927**	-0.0176*	-0.00700	-0.00218	0.00640	0.00349	-0.00336	-0.0133**	0.0307***
	(0.00378)	(0.0101)	(0.00756)	(0.00823)	(0.00584)	(0.00621)	(0.00572)	(0.00665)	(0.0107)
Distance-specific linear time trends	-0.271***	0.0947	-0.0986*	0.380**	-0.0784	0.229***	-0.0485	-0.0131	0.118
	(0.0767)	(0.0777)	(0.0596)	(0.168)	(0.0704)	(0.0718)	(0.0683)	(0.113)	(0.100)
Distance-specific quadratic time trends	0.000420***	-0.000140	0.000144	-0.000610**	0.000127	-0.000341***	0.0000827	0.0000172	-0.000170
	(0.000119)	(0.000121)	(0.0000903)	(0.000264)	(0.000108)	(0.000111)	(0.000107)	(0.000175)	(0.000152)
Distance-specific cubic time trends	-0.000000216***	6.86e-08	-7.03e-08	0.000000323**	-6.84e-08	0.000000169***	-4.59e-08	-7.12e-09	8.16e-08
	(6.10e-08)	(6.24e-08)	(4.55e-08)	(0.000000137)	(5.49e-08)	(5.67e-08)	(5.51e-08)	(8.98e-08)	(7.67e-08)
Difference in population density	-0.0213	-0.0499	-0.00947	-0.00299	-0.0526	0.0217	-0.0245	0.0673	-0.259***
	(0.0670)	(0.0720)	(0.0456)	(0.0515)	(0.0591)	(0.0354)	(0.0285)	(0.0613)	(0.0760)
=1 if one of the markets experienced flood shocks	0.0276*	0.00463	0.0240	0.00177	-0.00438	0.00208	0.0236	0.0113	-0.0114
	(0.0146)	(0.0119)	(0.0167)	(0.0266)	(0.0138)	(0.0158)	(0.0213)	(0.0140)	(0.0112)
=1 if one of the markets experienced drought shocks	-0.0451**	-0.0285*	0.0132	-0.0329	0.0295***	-0.0570***	-0.0143	0.0294	-0.0109
	(0.0186)	(0.0152)	(0.0144)	(0.0307)	(0.0103)	(0.0149)	(0.0136)	(0.0243)	(0.0239)
Difference in local production	-1.007	1.339	0.603	0.994	-1.311***	0.494	-0.881	-	-
	(0.981)	(0.938)	(0.540)	(1.280)	(0.434)	(0.328)	(1.037)	-	-
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Market-pair FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	44023	57195	79971	64135	77243	69666	59931	76513	58832

Note: Poisson pseudo-likelihood regression with multiple levels of fixed effects (ppmlhdfe) estimator results for each food items are presented in each column while controlling for the non-linear effects of transport costs. The dependent variable is absolute value of price difference. Route distance over paved road between the market pairs is measured in hundreds of kilometres, and diesel fuel price in thousands of Malawian Kwacha. Dyadic clustered standard errors at the market in parentheses * p<0.10, ** p<0.05, *** p<0.010

Table A.23: The association between transport costs and market price dispersion of legumes and nuts

Dependent variable: $\lvert P_{xt} - P_{yt} \rvert$	Brown beans	Groundnuts	White beans
Diesel price x distance	0.104*	-0.0293	0.0536
	(0.0545)	(0.0925)	(0.0705)
Diesel price x (distance)^2	-0.00141	-0.00222	-0.00229
	(0.00447)	(0.00653)	(0.00471)
Distance-specific linear time trends	0.120**	-0.0276	-0.0561
	(0.0522)	(0.0452)	(0.0581)
Distance-specific quadratic time trends	-0.000186**	0.0000467	0.0000881
	(0.0000805)	(0.0000696)	(0.0000898)
Distance-specific cubic time trends	9.58e-08**	-2.61e-08	-4.62e-08
	(4.13e-08)	(3.57e-08)	(4.60e-08)
Difference in population density	-0.0657	-0.0491	-0.0771
	(0.0548)	(0.0571)	(0.0915)
=1 if one of the markets experienced flood shocks	0.000362	0.00558	0.00236
	(0.0107)	(0.0105)	(0.0109)
=1 if one of the markets experienced drought shocks	-0.0323**	-0.0204	-0.0223
	(0.0132)	(0.0133)	(0.0138)
Difference in local production	0.880*	-0.262***	1.288**
	(0.491)	(0.0752)	(0.532)
Month FE	Yes	Yes	Yes
Time FE	Yes	Yes	Yes
Market-pair FE	Yes	Yes	Yes
N	79826	67104	60546

Note: Poisson pseudo-likelihood regression with multiple levels of fixed effects (ppmlhdfe) estimator results for each food items are presented in each column while controlling for the non-linear effects of transport costs. The dependent variable is absolute value of price difference. Route distance over paved road between the market pairs is measured in hundreds of kilometres, and diesel fuel price in thousands of Malawian Kwacha. Dyadic clustered standard errors at the market in parentheses * p<0.10, ** p<0.05, *** p<0.010

Table A.24: The association between transport costs and market price dispersion of animal source foods

Dependent variable: $\lvert P_{xt} - P_{yt} \rvert$	Beef	Eggs	Ultra-pasteurized milk	Goat meat	Powdered milk	Usipa	Utaka
Diesel price x distance	-0.203***	0.321***	0.110	-0.424***	-0.346**	0.0491	-0.00925
	(0.0700)	(0.0735)	(0.0905)	(0.103)	(0.152)	(0.0375)	(0.0607)
Diesel price x (distance)^2	0.00940**	-0.0178***	-0.00627	0.0255***	0.0107	-0.00116	-0.00846*
	(0.00451)	(0.00543)	(0.00695)	(0.00821)	(0.00898)	(0.00432)	(0.00480)
Distance-specific linear time trends	0.0117	-0.0235	0.140	-0.0196	0.0190	0.0508	-0.0758
	(0.0369)	(0.0693)	(0.110)	(0.0652)	(0.124)	(0.0779)	(0.0594)
Distance-specific quadratic time trends	-0.0000276	0.0000308	-0.000219	0.0000317	-0.0000206	-0.0000771	0.000128
	(0.0000572)	(0.000109)	(0.000172)	(0.0000994)	(0.000191)	(0.000122)	(0.0000903)
Distance-specific cubic time trends	1.97e-08	-1.35e-08	0.000000114	-1.66e-08	8.02e-09	3.91e-08	-7.05e-08
	(2.96e-08)	(5.63e-08)	(8.95e-08)	(5.04e-08)	(9.85e-08)	(6.30e-08)	(4.56e-08)
Difference in population density	-0.0222	-0.0101	0.0251	0.00375	0.195**	-0.0241	-0.112
	(0.0557)	(0.0302)	(0.0951)	(0.0252)	(0.0784)	(0.0826)	(0.0900)
=1 if one of the markets experienced flood shocks	-0.00324	-0.0199*	-0.0201*	0.0111	-0.0405***	0.0237***	-0.00598
	(0.0136)	(0.0112)	(0.0117)	(0.0111)	(0.0130)	(0.00898)	(0.0131)
=1 if one of the markets experienced drought shocks	0.00823	0.0269*	0.0341	-0.00230	0.00738	-0.0199	0.0137
	(0.0162)	(0.0161)	(0.0219)	(0.0130)	(0.0269)	(0.0128)	(0.0179)
Difference in local production	0.720	-1.577	-	1.864	-	0.0550	0.163*
	(4.691)	(2.703)	-	(1.419)	-	(0.246)	(0.0948)
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Market-pair FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	74127	71621	62329	74588	40866	78279	56714

Note: Poisson pseudo-likelihood regression with multiple levels of fixed effects (ppmlhdfe) estimator results for each food items are presented in each column while controlling for the non-linear effects of transport costs. The dependent variable is absolute value of price difference. Route distance over paved road between the market pairs is measured in hundreds of kilometres, and diesel fuel price in thousands of Malawian Kwacha. Dyadic clustered standard errors at the market in parentheses * p<0.10, ** p<0.05, *** p<0.010

Table A.25: The association between transport costs and market price dispersion of vegetables

Dependent variable: $\lvert P_{xt} - P_{yt} \rvert$	Cabbage	Okra	Onions	Pumpkin leaves	Rape leaves
Diesel price x distance	-0.0419 (0.0836)	-0.331** (0.134)	-0.116 (0.0760)	0.0415 (0.0519)	0.0899 (0.0663)
Diesel price x (distance)^2	-0.00369 (0.00494)	0.0294*** (0.00669)	0.0105* (0.00557)	-0.00293 (0.00340)	-0.00785 (0.00510)
Distance-specific linear time trends	0.0735 (0.0855)	0.0652 (0.0966)	0.0330 (0.0530)	0.148** (0.0703)	0.0606 (0.0569)
Distance-specific quadratic time trends	-0.000102 (0.000131)	-0.0000944 (0.000148)	-0.0000457 (0.0000813)	-0.000234** (0.000110)	-0.0000928 (0.0000872)
Distance-specific cubic time trends	4.61e-08 (6.69e-08)	4.52e-08 (7.51e-08)	2.08e-08 (4.13e-08)	0.000000123** (5.72e-08)	4.73e-08 (4.43e-08)
Difference in population density	-0.0749 (0.0496)	-0.00283 (0.0375)	-0.0848 (0.0543)	0.0864* (0.0485)	0.00959 (0.0587)
=1 if one of the markets experienced flood shocks	0.0197 (0.0133)	0.00859 (0.0160)	-0.0125 (0.0101)	0.000456 (0.0107)	0.0313** (0.0139)
=1 if one of the markets experienced drought shocks	0.0450*** (0.0155)	0.0351** (0.0161)	0.0543*** (0.0160)	-0.0255** (0.0125)	-0.00903 (0.0157)
Difference in local production	0.278 (1.201)	0.826 (2.383)	-4.561** (2.181)	2.741 (2.049)	4.023** (1.644)
Month FE	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes
Market-pair FE	Yes	Yes	Yes	Yes	Yes
N	72133	53953	80090	64588	74680

Note: Poisson pseudo-likelihood regression with multiple levels of fixed effects (ppmlhdfe) estimator results for each food items are presented in each column while controlling for the non-linear effects of transport costs. The dependent variable is absolute value of price difference. Route distance over paved road between the market pairs is measured in hundreds of kilometres, and diesel fuel price in thousands of Malawian Kwacha. Dyadic clustered standard errors at the market in parentheses * p<0.10, ** p<0.05, *** p<0.010

Table A.26: The association between transport costs and market price dispersion of fruits

Dependent variable: $\lvert P_{xt} - P_{yt} \rvert$	Bananas	Tomatoes
Diesel price x distance	-0.399***	-0.00837
	(0.0998)	(0.0631)
Diesel price x (distance)^2	0.0264***	-0.000694
	(0.00768)	(0.00419)
Distance-specific linear time trends	0.0282	0.0499
	(0.0864)	(0.0476)
Distance-specific quadratic time trends	-0.0000362	-0.0000692
	(0.000133)	(0.0000734)
Distance-specific cubic time trends	1.55e-08	3.15e-08
	(6.75e-08)	(3.75e-08)
Difference in population density	-0.0226	-0.0113
	(0.0456)	(0.0399)
=1 if one of the markets experienced flood shocks	0.00872	-0.00165
	(0.0143)	(0.0148)
=1 if one of the markets experienced drought shocks	0.0265**	0.00170
	(0.0107)	(0.0150)
Difference in local production	0.473	-4.220***
	(7.198)	(1.195)
Month FE	Yes	Yes
Time FE	Yes	Yes
Market-pair FE	Yes	Yes
N	66053	80558

Note: Poisson pseudo-likelihood regression with multiple levels of fixed effects (ppmlhdfe) estimator results for each food items are presented in each column while controlling for the non-linear effects of transport costs. The dependent variable is absolute value of price difference. Route distance over paved road between the market pairs is measured in hundreds of kilometres, and diesel fuel price in thousands of Malawian Kwacha. Dyadic clustered standard errors at the market in parentheses * p<0.10, ** p<0.05, *** p<0.010

Table A.27: The association between transport costs and market price dispersion of staples including roots and tubers

Dependent variable: $\lvert P_{xt} - P_{yt} \rvert$	Cassava	Maize flour dehulled	Maize grain (private)	Maize grain (ADMARC)	Potatoes	Rice grain	Sweet potatoes	White bread	White buns
Diesel price x distance	-0.0785	0.122**	-0.0184	0.131	0.0177	0.00271	0.0685	-0.0217	0.00275
	(0.0878)	(0.0536)	(0.0469)	(0.195)	(0.0549)	(0.0751)	(0.0687)	(0.0576)	(0.0504)
Lagged diesel price x distance	0.00784	-0.0535	0.106**	0.234	0.0189	-0.118	-0.148*	-0.0175	0.0562
	(0.107)	(0.0452)	(0.0503)	(0.147)	(0.0397)	(0.0953)	(0.0897)	(0.0497)	(0.0372)
Distance-specific linear time trends	-0.282***	0.115	-0.0788	0.406**	-0.0780	0.217***	-0.0547	0.000937	0.0666
	(0.0804)	(0.0795)	(0.0622)	(0.166)	(0.0718)	(0.0711)	(0.0777)	(0.115)	(0.119)
Distance-specific quadratic time trends	0.000435***	-0.000168	0.000114	-0.000649**	0.000126	-0.000321***	0.0000933	-0.00000200	-0.0000997
	(0.000125)	(0.000123)	(0.0000948)	(0.000260)	(0.000110)	(0.000109)	(0.000121)	(0.000178)	(0.000180)
Distance-specific cubic time trends	-0.000000223***	8.12e-08	-5.53e-08	0.000000343**	-6.73e-08	0.000000159***	-5.19e-08	1.59e-09	4.94e-08
	(6.45e-08)	(6.31e-08)	(4.81e-08)	(0.000000136)	(5.61e-08)	(5.56e-08)	(6.25e-08)	(9.14e-08)	(9.03e-08)
Difference in population density	-0.0227	-0.0467	-0.00791	-0.00258	-0.0548	0.0202	-0.0240	0.0705	-0.266***
	(0.0672)	(0.0727)	(0.0447)	(0.0519)	(0.0580)	(0.0350)	(0.0285)	(0.0609)	(0.0746)
=1 if one of the markets experienced flood shocks	0.0272*	0.00498	0.0241	0.00343	-0.00471	0.00184	0.0239	0.0113	-0.0182*
	(0.0145)	(0.0118)	(0.0167)	(0.0263)	(0.0137)	(0.0156)	(0.0212)	(0.0141)	(0.0109)
=1 if one of the markets experienced drought shocks	-0.0465**	-0.0268*	0.0147	-0.0325	0.0283***	-0.0569***	-0.0141	0.0318	-0.0184
	(0.0184)	(0.0154)	(0.0146)	(0.0307)	(0.0101)	(0.0147)	(0.0139)	(0.0235)	(0.0220)
Difference in local production	-0.892	1.217	0.554	1.000	-1.371***	0.492	-0.882	-	-
	(1.008)	(0.904)	(0.534)	(1.267)	(0.465)	(0.344)	(1.068)	-	-
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Market-pair FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	43639	56761	79475	63981	76778	69285	59578	76107	58424

Note: Poisson pseudo-likelihood regression with multiple levels of fixed effects (ppmlhdfe) estimator results for each food items are presented in each column. The dependent variable is absolute value of price difference. Route distance over paved road between the market pairs is measured in hundreds of kilometres, and diesel fuel price in thousands of Malawian Kwacha. Dyadic clustered standard errors at the market in parentheses * p<0.10, ** p<0.05, *** p<0.010

Table A.28: The association between transport costs and market price dispersion of legumes and nuts

Dependent variable: $|P_{xt} - P_{yt}|$	Brown beans	Groundnuts	White beans
Diesel price x distance	0.156***	-0.0183	0.178***
	(0.0450)	(0.0574)	(0.0538)
Lagged diesel price x distance	-0.0757*	-0.0405	-0.165***
	(0.0433)	(0.0690)	(0.0420)
Distance-specific linear time trends	0.116**	-0.0326	-0.0666
	(0.0519)	(0.0466)	(0.0588)
Distance-specific quadratic time trends	-0.000179**	0.0000547	0.000105
	(0.0000804)	(0.0000720)	(0.0000910)
Distance-specific cubic time trends	9.19e-08**	-3.03e-08	-5.53e-08
	(4.14e-08)	(3.70e-08)	(4.67e-08)
Difference in population density	-0.0655	-0.0491	-0.0768
	(0.0549)	(0.0569)	(0.0914)
=1 if one of the markets experienced flood shocks	0.000355	0.00576	0.00218
	(0.0107)	(0.0106)	(0.0108)
=1 if one of the markets experienced drought shocks	-0.0323**	-0.0199	-0.0220
	(0.0133)	(0.0129)	(0.0139)
Difference in local production	0.894*	-0.264***	1.310**
	(0.501)	(0.0745)	(0.527)
Month FE	Yes	Yes	Yes
Time FE	Yes	Yes	Yes
Market-pair FE	Yes	Yes	Yes
N	79330	66726	60113

Note: Poisson pseudo-likelihood regression with multiple levels of fixed effects (ppmlhdfe) estimator results for each food items are presented in each column. The dependent variable is absolute value of price difference. Route distance over paved road between the market pairs is measured in hundreds of kilometres, and diesel fuel price in thousands of Malawian Kwacha. Dyadic clustered standard errors at the market in parentheses * p<0.10, ** p<0.05, *** p<0.010

Table A.29: The association between transport costs and market price dispersion of animal source foods

Dependent variable: $\lvert P_{xt} - P_{yt} \rvert$	Beef	Eggs	Ultra-pasteurized milk	Goat meat	Powdered milk	Usipa	Utaka
Diesel price x distance	-0.0388	0.0553	-0.0444	-0.104***	-0.103*	0.0875*	-0.00341
	(0.0339)	(0.0559)	(0.0509)	(0.0269)	(0.0610)	(0.0462)	(0.0906)
Lagged diesel price x distance	-0.0599**	0.0665	0.0857*	-0.00937	-0.121**	-0.0585	-0.105*
	(0.0246)	(0.0678)	(0.0449)	(0.0257)	(0.0565)	(0.0522)	(0.0609)
Distance-specific linear time trends	-0.00748	0.0188	0.161	-0.0483	-0.0194	0.0481	-0.0786
	(0.0336)	(0.0788)	(0.118)	(0.0589)	(0.120)	(0.0793)	(0.0608)
Distance-specific quadratic time trends	9.69e-08	-0.0000303	-0.000249	0.0000709	0.0000359	-0.0000725	0.000134
	(0.0000525)	(0.000122)	(0.000184)	(0.0000909)	(0.000188)	(0.000124)	(0.0000924)
Distance-specific cubic time trends	6.34e-09	1.58e-08	0.000000128	-3.43e-08	-1.96e-08	3.64e-08	-7.45e-08
	(2.73e-08)	(6.30e-08)	(9.52e-08)	(4.66e-08)	(9.77e-08)	(6.43e-08)	(4.67e-08)
Difference in population density	-0.0224	-0.00613	0.0271	-0.00179	0.191**	-0.0226	-0.110
	(0.0553)	(0.0291)	(0.0952)	(0.0235)	(0.0782)	(0.0822)	(0.0900)
=1 if one of the markets experienced flood shocks	-0.00395	-0.0190*	-0.0192*	0.00871	-0.0397***	0.0241***	-0.00532
	(0.0136)	(0.0115)	(0.0116)	(0.0111)	(0.0128)	(0.00898)	(0.0132)
=1 if one of the markets experienced drought shocks	0.00605	0.0305*	0.0352	-0.00803	0.00533	-0.0198	0.0139
	(0.0165)	(0.0156)	(0.0221)	(0.0134)	(0.0262)	(0.0125)	(0.0180)
Difference in local production	1.057	-1.788	-	2.380*	-	0.0548	0.163*
	(4.643)	(2.657)	-	(1.291)	-	(0.246)	(0.0952)
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Market-pair FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	73744	71296	62022	74182	40645	77814	56308

Note: Poisson pseudo-likelihood regression with multiple levels of fixed effects (ppmlhdfe) estimator results for each food items are presented in each column. The dependent variable is absolute value of price difference. Route distance over paved road between the market pairs is measured in hundreds of kilometres, and diesel fuel price in thousands of Malawian Kwacha. Dyadic clustered standard errors at the market in parentheses * p<0.10, ** p<0.05, *** p<0.010

Table A.30: The association between transport costs and market price dispersion of vegetables

Dependent variable: $\lvert P_{xt} - P_{yt} \rvert$	Cabbage	Okra	Onions	Pumpkin leaves	Rape leaves
Diesel price x distance	-0.0393 (0.0546)	0.0144 (0.0626)	0.00924 (0.0392)	-0.0222 (0.0613)	-0.0198 (0.0384)
Lagged diesel price x distance	-0.0499 (0.0445)	0.00825 (0.0584)	-0.0103 (0.0447)	0.0326 (0.0581)	0.0188 (0.0520)
Distance-specific linear time trends	0.0696 (0.0914)	0.0137 (0.0987)	0.0281 (0.0520)	0.155** (0.0749)	0.0768 (0.0553)
Distance-specific quadratic time trends	-0.0000951 (0.000140)	-0.0000216 (0.000152)	-0.0000394 (0.0000803)	-0.000244** (0.000117)	-0.000116 (0.0000851)
Distance-specific cubic time trends	4.25e-08 (7.15e-08)	1.09e-08 (7.72e-08)	1.82e-08 (4.11e-08)	0.000000128** (6.11e-08)	5.84e-08 (4.35e-08)
Difference in population density	-0.0739 (0.0490)	-0.00994 (0.0370)	-0.0873 (0.0532)	0.0869* (0.0489)	0.0121 (0.0589)
=1 if one of the markets experienced flood shocks	0.0197 (0.0133)	0.00613 (0.0155)	-0.0128 (0.0101)	0.000368 (0.0106)	0.0315** (0.0138)
=1 if one of the markets experienced drought shocks	0.0453*** (0.0154)	0.0296* (0.0158)	0.0527*** (0.0158)	-0.0252** (0.0127)	-0.00814 (0.0156)
Difference in local production	0.304 (1.201)	2.004 (2.349)	-4.652** (2.207)	2.749 (2.092)	3.470** (1.650)
Month FE	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes
Market-pair FE	Yes	Yes	Yes	Yes	Yes
N	71779	53511	79625	64155	74215

Note: Poisson pseudo-likelihood regression with multiple levels of fixed effects (ppmlhdfe) estimator results for each food items are presented in each column. The dependent variable is absolute value of price difference. Route distance over paved road between the market pairs is measured in hundreds of kilometres, and diesel fuel price in thousands of Malawian Kwacha. Dyadic clustered standard errors at the market in parentheses * p<0.10, ** p<0.05, *** p<0.010

Table A.31: The association between transport costs and market price dispersion of fruits

Dependent variable: $\lvert P_{xt} - P_{yt} \rvert$	Bananas	Tomatoes
Diesel price x distance	-0.136***	0.0301
	(0.0258)	(0.0315)
Lagged diesel price x distance	0.0606**	-0.0531*
	(0.0307)	(0.0280)
Distance-specific linear time trends	0.00463	0.0507
	(0.0832)	(0.0492)
Distance-specific quadratic time trends	-0.00000527	-0.0000699
	(0.000128)	(0.0000756)
Distance-specific cubic time trends	2.14e-09	3.16e-08
	(6.54e-08)	(3.86e-08)
Difference in population density	-0.0276	-0.0106
	(0.0437)	(0.0397)
=1 if one of the markets experienced flood shocks	0.00634	-0.00151
	(0.0144)	(0.0147)
=1 if one of the markets experienced drought shocks	0.0221*	0.00161
	(0.0113)	(0.0150)
Difference in local production	0.253	-4.220***
	(6.983)	(1.196)
Month FE	Yes	Yes
Time FE	Yes	Yes
Market-pair FE	Yes	Yes
N	65618	80062

Note: Poisson pseudo-likelihood regression with multiple levels of fixed effects (ppmlhdfe) estimator results for each food items are presented in each column. The dependent variable is absolute value of price difference. Route distance over paved road between the market pairs is measured in hundreds of kilometres, and diesel fuel price in thousands of Malawian Kwacha. Dyadic clustered standard errors at the market in parentheses * p<0.10, ** p<0.05, *** p<0.010

Table A.32: The association between transport costs and market price dispersion of staples including roots and tubers

Dependent variable: $\lvert P_{xt} - P_{yt}\rvert$	Cassava	Maize flour dehulled	Maize grain (private)	Maize grain (ADMARC)	Potatoes	Rice grain	Sweet potatoes	White bread	White buns
Diesel price x distance (102 - 151)	-0.160 (0.418)	0.0678 (0.332)	-0.842** (0.327)	0.0316 (0.466)	0.121 (0.316)	-0.129 (0.180)	-0.230 (0.383)	-0.286 (0.211)	-0.206 (0.296)
Diesel price x distance (152-200)	-0.0512 (0.389)	0.466** (0.195)	-0.393 (0.324)	-0.356 (0.701)	0.321 (0.326)	-0.392** (0.156)	0.123 (0.327)	-0.530** (0.248)	0.232 (0.322)
Diesel price x distance (201 - 267)	-0.0687 (0.377)	0.134 (0.353)	-0.261 (0.273)	-0.398 (0.640)	0.291 (0.286)	-0.340 (0.245)	0.171 (0.319)	-0.635*** (0.182)	-0.295 (0.316)
Diesel price x distance (268 – 331)	0.177 (0.340)	0.812*** (0.273)	-0.194 (0.367)	-0.213 (0.591)	0.412 (0.316)	-0.529*** (0.205)	0.118 (0.359)	-0.177 (0.222)	-0.245 (0.260)
Diesel price x distance (332 – 393)	-0.179 (0.273)	0.332 (0.236)	-0.158 (0.362)	0.0908 (0.407)	-0.0808 (0.290)	-0.328 (0.242)	-0.118 (0.415)	-0.217 (0.233)	0.0536 (0.339)
Diesel price x distance (394 – 479)	0.209 (0.351)	0.432 (0.265)	0.287 (0.285)	-0.255 (0.587)	0.408 (0.291)	-0.572** (0.290)	-0.602* (0.365)	0.0819 (0.266)	-0.327 (0.519)
Diesel price x distance (480 – 600)	0.123 (0.334)	0.486* (0.253)	0.187 (0.290)	-0.00840 (0.699)	0.346 (0.314)	-1.139*** (0.236)	-0.352 (0.368)	-0.367 (0.235)	0.0154 (0.376)
Diesel price x distance (601 – 734)	-0.563 (0.390)	0.471* (0.277)	0.124 (0.274)	0.468 (0.591)	0.346 (0.281)	-1.178*** (0.229)	-0.0666 (0.477)	-0.554** (0.239)	-0.0217 (0.306)
Diesel price x distance (735 – 1098)	0.118 (0.408)	0.351 (0.324)	0.321 (0.251)	1.585** (0.737)	0.709 (0.462)	-1.342*** (0.298)	-0.449 (0.476)	-0.616 (0.398)	0.458* (0.278)
Distance (102 - 151) x squared time trends	-0.00669 (0.0303)	0.0628* (0.0331)	0.0625** (0.0264)	0.0310 (0.0502)	0.0195 (0.0235)	-0.00440 (0.0146)	0.0223 (0.0160)	-0.00358 (0.0237)	-0.0327 (0.0288)
Distance (152-200) x squared time trends	-0.0365 (0.0297)	0.0416 (0.0356)	0.0429* (0.0228)	0.0461 (0.0721)	0.0162 (0.0179)	-0.00109 (0.0140)	0.0124 (0.0210)	-0.0223 (0.0299)	-0.0209 (0.0355)
Distance (201 - 267) x squared time trends	-0.0446** (0.0225)	0.0546 (0.0436)	0.0154 (0.0202)	0.0257 (0.0535)	0.00750 (0.0221)	0.00359 (0.0192)	-0.00583 (0.0224)	0.0134 (0.0305)	-0.0347 (0.0296)
Distance (268 – 331) x squared time trends	0.00132 (0.0222)	0.0482 (0.0327)	0.0250 (0.0181)	-0.0993 (0.0748)	-0.0166 (0.0164)	0.0331** (0.0162)	0.00983 (0.0170)	-0.0106 (0.0278)	0.0388 (0.0366)
Distance (332 – 393) x squared time trends	-0.0127 (0.0171)	0.0195 (0.0278)	0.0132 (0.0243)	-0.118** (0.0511)	0.00452 (0.0149)	0.0268 (0.0166)	0.0191 (0.0197)	-0.00900 (0.0304)	-0.00183 (0.0263)
Distance (394 – 479) x squared time trends	-0.0187 (0.0193)	0.0421 (0.0356)	0.0180 (0.0219)	-0.144*** (0.0540)	0.0159 (0.0173)	0.0550** (0.0221)	0.0871*** (0.0271)	0.0237 (0.0241)	0.0370 (0.0303)
Distance (480 – 600) x squared time trends	0.00438 (0.0317)	0.103*** (0.0382)	-0.0292 (0.0226)	-0.175*** (0.0559)	0.0259 (0.0174)	0.0741*** (0.0183)	0.0824*** (0.0215)	0.0165 (0.0273)	0.0201 (0.0305)
Distance (601 – 734) x squared time trends	0.0322 (0.0334)	0.103** (0.0417)	-0.0251 (0.0254)	-0.216*** (0.0533)	0.0536** (0.0213)	0.0849*** (0.0160)	0.0687*** (0.0224)	-0.0311 (0.0274)	0.0238 (0.0312)
Distance (735 – 1098) x squared time trends	-0.00285 (0.0462)	0.101*** (0.0338)	-0.0212 (0.0309)	-0.129 (0.0827)	0.0640** (0.0267)	0.100*** (0.0151)	0.0732** (0.0295)	-0.0124 (0.0334)	0.0181 (0.0348)
Distance (102 - 151) x cubic time trends	0.00000608 (0.0000216)	-0.0000496** (0.0000249)	-0.0000448** (0.0000196)	-0.0000260 (0.0000390)	-0.0000161 (0.0000175)	0.00000290 (0.0000109)	-0.0000162 (0.0000117)	0.00000424 (0.0000186)	0.0000247 (0.0000225)
Distance (152-200) x cubic time trends	0.0000281 (0.0000215)	-0.0000326 (0.0000275)	-0.0000300* (0.0000167)	-0.0000364 (0.0000563)	-0.0000140 (0.0000130)	0.00000143 (0.0000109)	-0.00000934 (0.0000154)	0.0000199 (0.0000226)	0.0000144 (0.0000268)
Distance (201 - 267) x cubic time trends	0.0000339** (0.0000160)	-0.0000418 (0.0000331)	-0.0000102 (0.0000154)	-0.0000190 (0.0000397)	-0.00000724 (0.0000174)	-0.00000277 (0.0000147)	0.00000348 (0.0000166)	-0.00000727 (0.0000235)	0.0000279 (0.0000228)
Distance (268 – 331) x cubic time trends	-0.00000167 (0.0000157)	-0.0000381 (0.0000251)	-0.0000177 (0.0000134)	0.0000768 (0.0000572)	0.0000124 (0.0000123)	-0.0000248** (0.0000125)	-0.00000689 (0.0000120)	0.00000988 (0.0000214)	-0.0000304 (0.0000278)

Note: Poisson pseudo-likelihood regression with multiple levels of fixed effects (ppmlhdfe) estimator results for each food items are presented in each column. The dependent variable is absolute value of price difference. Route distance over paved road between the market pairs is measured in kilometres, and diesel fuel price in thousands of Malawian Kwacha. Base category is distance below 102 km. Dyadic clustered standard errors at the market in parentheses * p<0.10, ** p<0.05, *** p<0.010

Table A.32 continued…

Dependent variable: $\lvert P_{xt} - P_{yt} \rvert$	Cassava	Maize flour dehulled	Maize grain (private)	Maize grain (ADMARC)	Potatoes	Rice grain	Sweet potatoes	White bread	White buns
Distance (332 – 393) x cubic time trends	0.00000937	-0.0000157	-0.00000952	0.0000906**	-0.00000357	-0.0000199	-0.0000142	0.00000934	-0.000000986
	(0.0000127)	(0.0000214)	(0.0000178)	(0.0000393)	(0.0000114)	(0.0000125)	(0.0000142)	(0.0000236)	(0.0000202)
Distance (394 – 479) x cubic time trends	0.0000134	-0.0000341	-0.0000151	0.000111***	-0.0000144	-0.0000410**	-0.0000648***	-0.0000167	-0.0000328
	(0.0000135)	(0.0000274)	(0.0000165)	(0.0000417)	(0.0000135)	(0.0000170)	(0.0000201)	(0.0000185)	(0.0000239)
Distance (480 – 600) x cubic time trends	-0.00000290	-0.0000802***	0.0000212	0.000135***	-0.0000217*	-0.0000542***	-0.0000610***	-0.0000102	-0.0000169
	(0.0000234)	(0.0000293)	(0.0000166)	(0.0000436)	(0.0000128)	(0.0000141)	(0.0000157)	(0.0000212)	(0.0000240)
Distance (601 – 734) x cubic time trends	-0.0000216	-0.0000792**	0.0000176	0.000165***	-0.0000429***	-0.0000617***	-0.0000511***	0.0000269	-0.0000188
	(0.0000244)	(0.0000322)	(0.0000187)	(0.0000400)	(0.0000158)	(0.0000125)	(0.0000159)	(0.0000219)	(0.0000241)
Distance (735 – 1098) x cubic time trends	0.00000325	-0.0000788***	0.0000143	0.0000915	-0.0000501***	-0.0000730***	-0.0000529**	0.0000133	-0.0000135
	(0.0000339)	(0.0000259)	(0.0000224)	(0.0000625)	(0.0000193)	(0.0000116)	(0.0000210)	(0.0000274)	(0.0000266)
Difference in population density	-0.0114	-0.0518	-0.00838	-0.0310	-0.0493	0.0233	-0.0193	0.0719	-0.256***
	(0.0638)	(0.0698)	(0.0452)	(0.0538)	(0.0526)	(0.0358)	(0.0269)	(0.0617)	(0.0769)
=1 if one of the markets experienced flood shocks	0.0270*	0.00407	0.0221	0.00402	-0.00559	-0.000649	0.0232	0.00947	-0.0155
	(0.0147)	(0.0114)	(0.0164)	(0.0269)	(0.0130)	(0.0157)	(0.0213)	(0.0135)	(0.0101)
=1 if one of the markets experienced drought shocks	-0.0427**	-0.0251*	0.0155	-0.0325	0.0309***	-0.0577***	-0.0152	0.0307	-0.0143
	(0.0191)	(0.0141)	(0.0146)	(0.0288)	(0.0103)	(0.0143)	(0.0138)	(0.0238)	(0.0235)
Difference in local production	-0.921	1.348	0.489	1.329	-1.298***	0.466	-0.838	-	-
	(1.009)	(0.948)	(0.529)	(1.279)	(0.374)	(0.311)	(1.020)	-	-
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Market-pair FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	44023	57195	79971	64135	77243	69666	59931	76513	58832

Table A.33: The association between transport costs and market price dispersion of legumes and nuts

Dependent variable: $\lvert P_{xt} - P_{yt}\rvert$	Brown beans PD	Groundnuts PD	White beans PD
Diesel price x distance (102 - 151)	0.154 (0.177)	-0.267 (0.244)	0.575** (0.277)
Diesel price x distance (152-200)	0.0392 (0.152)	0.0708 (0.178)	0.662*** (0.239)
Diesel price x distance (201 - 267)	-0.0787 (0.202)	-0.0829 (0.225)	0.317 (0.204)
Diesel price x distance (268 – 331)	0.344** (0.169)	-0.509* (0.303)	0.706*** (0.158)
Diesel price x distance (332 – 393)	-0.0128 (0.235)	-0.532* (0.282)	0.362 (0.282)
Diesel price x distance (394 – 479)	0.252 (0.207)	-0.604* (0.341)	0.540* (0.280)
Diesel price x distance (480 – 600)	0.211 (0.205)	-0.354 (0.250)	0.640*** (0.228)
Diesel price x distance (601 – 734)	0.185 (0.197)	-0.466* (0.259)	0.519* (0.270)
Diesel price x distance (735 – 1098)	0.434 (0.280)	-0.175 (0.361)	0.769** (0.340)
Distance (102 - 151) x squared time trends	-0.0194 (0.0147)	-0.00468 (0.0204)	-0.0234 (0.0223)
Distance (152-200) x squared time trends	-0.00597 (0.0159)	0.00531 (0.0245)	-0.0175 (0.0239)
Distance (201 - 267) x squared time trends	-0.00239 (0.0144)	-0.0241 (0.0264)	-0.0465** (0.0227)
Distance (268 – 331) x squared time trends	-0.0293*** (0.00965)	0.00481 (0.0266)	-0.0470 (0.0288)
Distance (332 – 393) x squared time trends	-0.0347** (0.0151)	-0.0170 (0.0315)	-0.0514* (0.0287)
Distance (394 – 479) x squared time trends	-0.0309 (0.0225)	0.000370 (0.0276)	-0.0225 (0.0337)
Distance (480 – 600) x squared time trends	-0.00527 (0.0162)	-0.0116 (0.0220)	-0.0290 (0.0220)
Distance (601 – 734) x squared time trends	0.000279 (0.0247)	0.0114 (0.0188)	-0.00291 (0.0222)
Distance (735 – 1098) x squared time trends	-0.0211 (0.0296)	0.0432 (0.0419)	0.00551 (0.0196)
Distance (102 - 151) x cubic time trends	0.0000145 (0.0000109)	0.00000330 (0.0000158)	0.0000151 (0.0000165)
Distance (152-200) x cubic time trends	0.00000360 (0.0000120)	-0.00000554 (0.0000188)	0.00000908 (0.0000177)
Distance (201 - 267) x cubic time trends	0.00000224 (0.0000106)	0.0000180 (0.0000201)	0.0000329** (0.0000168)
Distance (268 – 331) x cubic time trends	0.0000215*** (0.00000723)	-0.00000235 (0.0000204)	0.0000326 (0.0000212)

Note: Poisson pseudo-likelihood regression with multiple levels of fixed effects (ppmlhdfe) estimator results for each food items are presented in each column. The dependent variable is absolute value of price difference. Route distance over paved road between the market pairs is measured in kilometres, and diesel fuel price in thousands of Malawian Kwacha. Base category is distance below 102 km. Dyadic clustered standard errors at the market in parentheses * p<0.10, ** p<0.05, *** p<0.010

Table A.33 continued…

Dependent variable: $\lvert P_{xt} - P_{yt} \rvert$	Brown beans PD	Groundnuts PD	White beans PD
Distance (332 – 393) x cubic time trends	0.0000262**	0.0000137	0.0000368*
	(0.0000111)	(0.0000238)	(0.0000209)
Distance (394 – 479) x cubic time trends	0.0000220	0.000000284	0.0000140
	(0.0000165)	(0.0000207)	(0.0000250)
Distance (480 – 600) x cubic time trends	0.00000213	0.00000878	0.0000178
	(0.0000122)	(0.0000168)	(0.0000165)
Distance (601 – 734) x cubic time trends	-0.00000119	-0.00000800	-0.00000105
	(0.0000183)	(0.0000142)	(0.0000166)
Distance (735 – 1098) x cubic time trends	0.0000143	-0.0000351	-0.00000854
	(0.0000225)	(0.0000308)	(0.0000151)
Difference in population density	-0.0649	-0.0525	-0.0688
	(0.0528)	(0.0536)	(0.0862)
=1 if one of the markets experienced flood shocks	-0.0000338	0.00736	0.00321
	(0.0104)	(0.0108)	(0.0106)
=1 if one of the markets experienced drought shocks	-0.0307**	-0.0157	-0.0199
	(0.0132)	(0.0127)	(0.0143)
Difference in local production	0.836*	-0.222***	1.211**
	(0.464)	(0.0717)	(0.542)
Month FE	Yes	Yes	Yes
Time FE	Yes	Yes	Yes
Market-pair FE	Yes	Yes	Yes
N	79826	67104	60546

Table A.34: The association between transport costs and market price dispersion of animal source foods

Dependent variable: $\lvert P_{xt} - P_{yt} \rvert$	Beef	Eggs	Ultra- pasteurized milk	Goat meat	Powdered milk	Usipa	Utaka
Diesel price x distance (102 - 151)	-0.216 (0.228)	0.574* (0.337)	0.185 (0.335)	-0.308* (0.169)	-1.495** (0.717)	0.823*** (0.235)	0.152 (0.217)
Diesel price x distance (152-200)	0.192 (0.331)	0.722*** (0.266)	0.287 (0.259)	-0.438* (0.240)	-1.541** (0.668)	0.529** (0.217)	0.317 (0.202)
Diesel price x distance (201 - 267)	-0.380* (0.212)	0.571** (0.255)	0.154 (0.285)	-0.638*** (0.229)	-1.343** (0.588)	0.546*** (0.204)	0.230 (0.179)
Diesel price x distance (268 – 331)	-0.164 (0.188)	0.447* (0.257)	0.0241 (0.299)	-0.449* (0.230)	-1.318 (0.871)	0.196 (0.289)	0.169 (0.302)
Diesel price x distance (332 – 393)	-0.252 (0.244)	0.355 (0.260)	0.0549 (0.258)	-0.685*** (0.184)	-0.581 (0.770)	0.428* (0.241)	0.268 (0.177)
Diesel price x distance (394 – 479)	-0.358 (0.247)	0.686* (0.403)	0.159 (0.390)	-0.734*** (0.178)	-1.230** (0.579)	0.229 (0.237)	0.139 (0.283)
Diesel price x distance (480 – 600)	-0.577** (0.240)	0.596** (0.265)	0.307 (0.437)	-1.021*** (0.205)	-1.062* (0.616)	0.141 (0.233)	-0.119 (0.333)
Diesel price x distance (601 – 734)	-0.754*** (0.280)	0.808** (0.376)	0.142 (0.320)	-1.279*** (0.183)	-1.571** (0.622)	0.151 (0.192)	0.336 (0.381)
Diesel price x distance (735 – 1098)	-0.740** (0.307)	1.304*** (0.341)	0.115 (0.412)	-0.666** (0.263)	-2.724*** (0.734)	0.871*** (0.250)	-1.329*** (0.487)
Distance (102 - 151) x squared time trends	-0.0171 (0.0296)	0.0130 (0.0180)	-0.0602 (0.0435)	-0.0403 (0.0271)	0.00335 (0.0377)	-0.0195 (0.0150)	0.0260 (0.0182)
Distance (152-200) x squared time trends	-0.0794** (0.0374)	-0.00650 (0.0240)	-0.0452 (0.0518)	-0.0110 (0.0240)	0.00567 (0.0263)	-0.00594 (0.0102)	0.0339** (0.0165)
Distance (201 - 267) x squared time trends	-0.0490* (0.0278)	-0.0303 (0.0217)	-0.0122 (0.0433)	-0.00278 (0.0227)	-0.0465 (0.0353)	0.00438 (0.0131)	0.0196 (0.0135)
Distance (268 – 331) x squared time trends	-0.0696** (0.0292)	-0.0323 (0.0205)	-0.00703 (0.0451)	-0.0276 (0.0276)	-0.00408 (0.0664)	0.00325 (0.0150)	0.0448** (0.0181)
Distance (332 – 393) x squared time trends	-0.0817*** (0.0316)	-0.0380* (0.0205)	-0.0232 (0.0403)	-0.0211 (0.0296)	-0.0429 (0.0345)	-0.00921 (0.0176)	0.0250* (0.0144)
Distance (394 – 479) x squared time trends	-0.0818* (0.0440)	0.0143 (0.0262)	-0.0120 (0.0516)	-0.0383 (0.0266)	-0.0201 (0.0289)	0.00568 (0.0149)	0.0636** (0.0267)
Distance (480 – 600) x squared time trends	-0.137*** (0.0257)	-0.0117 (0.0355)	-0.0132 (0.0547)	-0.0523* (0.0310)	-0.0432 (0.0365)	0.0104 (0.0121)	0.0795*** (0.0208)
Distance (601 – 734) x squared time trends	-0.133*** (0.0338)	-0.0226 (0.0395)	-0.0608 (0.0478)	-0.0547* (0.0327)	-0.00520 (0.0359)	0.0326** (0.0143)	0.0551*** (0.0214)
Distance (735 – 1098) x squared time trends	-0.107*** (0.0336)	-0.0205 (0.0314)	-0.0353 (0.0581)	-0.0363 (0.0342)	0.0521 (0.0658)	-0.0141 (0.0156)	0.194*** (0.0259)
Distance (102 - 151) x cubic time trends	0.0000147 (0.0000230)	-0.0000120 (0.0000134)	0.0000478 (0.0000330)	0.0000305 (0.0000206)	0.000000183 (0.0000289)	0.0000130 (0.0000112)	-0.0000197 (0.0000133)
Distance (152-200) x cubic time trends	0.0000619** (0.0000289)	0.00000428 (0.0000179)	0.0000392 (0.0000390)	0.00000826 (0.0000179)	-0.00000139 (0.0000205)	0.00000347 (0.00000766)	-0.0000257** (0.0000125)
Distance (201 - 267) x cubic time trends	0.0000406* (0.0000216)	0.0000229 (0.0000167)	0.0000118 (0.0000327)	0.00000294 (0.0000174)	0.0000371 (0.0000270)	-0.00000446 (0.00000974)	-0.0000154 (0.0000101)
Distance (268 – 331) x cubic time trends	0.0000548** (0.0000229)	0.0000257 (0.0000161)	0.00000834 (0.0000344)	0.0000202 (0.0000207)	0.00000748 (0.0000488)	-0.00000196 (0.0000111)	-0.0000342*** (0.0000132)

Note: Poisson pseudo-likelihood regression with multiple levels of fixed effects (ppmlhdfe) estimator results for each food items are presented in each column. The dependent variable is absolute value of price difference. Route distance over paved road between the market pairs is measured in kilometres, and diesel fuel price in thousands of Malawian Kwacha. Base category is distance below 102 km. Dyadic clustered standard errors at the market in parentheses * p<0.10, ** p<0.05, *** p<0.010

Table A.34 continued…

Dependent variable: $\lvert P_{xt} - P_{yt} \rvert$	Beef	Eggs	Ultra-pasteurized milk	Goat meat	Powdered milk	Usipa	Utaka
Distance (332 – 393) x cubic time trends	0.0000640**	0.0000296*	0.0000197	0.0000159	0.0000328	0.00000668	-0.0000196*
	(0.0000249)	(0.0000161)	(0.0000307)	(0.0000228)	(0.0000268)	(0.0000131)	(0.0000108)
Distance (394 – 479) x cubic time trends	0.0000657*	-0.0000105	0.0000124	0.0000292	0.0000217	-0.00000371	-0.0000478**
	(0.0000344)	(0.0000198)	(0.0000384)	(0.0000206)	(0.0000208)	(0.0000112)	(0.0000197)
Distance (480 – 600) x cubic time trends	0.000109***	0.00000855	0.0000137	0.0000409*	0.0000389	-0.00000727	-0.0000590***
	(0.0000205)	(0.0000273)	(0.0000411)	(0.0000240)	(0.0000276)	(0.00000904)	(0.0000154)
Distance (601 – 734) x cubic time trends	0.000106***	0.0000153	0.0000500	0.0000433*	0.0000129	-0.0000233**	-0.0000407***
	(0.0000271)	(0.0000300)	(0.0000365)	(0.0000253)	(0.0000276)	(0.0000108)	(0.0000156)
Distance (735 – 1098) x cubic time trends	0.0000876***	0.0000116	0.0000327	0.0000283	-0.0000254	0.00000988	-0.000141***
	(0.0000268)	(0.0000241)	(0.0000449)	(0.0000266)	(0.0000492)	(0.0000117)	(0.0000190)
Difference in population density	-0.0231	-0.00673	0.0342	0.00660	0.203***	-0.0253	-0.106
	(0.0569)	(0.0333)	(0.0931)	(0.0244)	(0.0761)	(0.0803)	(0.0884)
=1 if one of the markets experienced flood shocks	-0.00395	-0.0169	-0.0183	0.00984	-0.0399***	0.0250***	-0.00644
	(0.0134)	(0.0109)	(0.0118)	(0.0107)	(0.0124)	(0.00919)	(0.0137)
=1 if one of the markets experienced drought shocks	0.00692	0.0313*	0.0355*	-0.00614	0.00767	-0.0193	0.0128
	(0.0163)	(0.0160)	(0.0215)	(0.0133)	(0.0272)	(0.0133)	(0.0175)
Difference in local production	0.249	-1.796	-	1.979	-	0.0698	0.124
	(4.429)	(2.491)	-	(1.248)	-	(0.263)	(0.0866)
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Market-pair FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	74127	71621	62329	74588	40866	78279	56714

Table A.35: The association between transport costs and market price dispersion of vegetables

Dependent variable: $\|P_{xt} - P_{yt}\|$	Cabbage	Okra	Onions	Pumpkin leaves	Rape leaves
Diesel price x distance (102 - 151)	-0.0189	-0.0169	0.223	0.0801	0.487
	(0.239)	(0.254)	(0.171)	(0.336)	(0.390)
Diesel price x distance (152-200)	-0.272	-0.104	0.263	-0.0798	0.240
	(0.199)	(0.253)	(0.165)	(0.338)	(0.258)
Diesel price x distance (201 - 267)	-0.339	-0.0147	-0.0465	0.134	0.258
	(0.289)	(0.253)	(0.240)	(0.348)	(0.239)
Diesel price x distance (268 – 331)	-0.146	-0.423	0.0819	0.230	0.299
	(0.324)	(0.299)	(0.210)	(0.338)	(0.285)
Diesel price x distance (332 – 393)	-0.310	-0.0483	-0.0378	0.108	0.156
	(0.227)	(0.318)	(0.174)	(0.368)	(0.281)
Diesel price x distance (394 – 479)	-0.355	0.282	-0.296	0.219	0.0543
	(0.322)	(0.365)	(0.270)	(0.306)	(0.312)
Diesel price x distance (480 – 600)	-0.552	-0.218	-0.183	0.105	0.00877
	(0.344)	(0.403)	(0.261)	(0.362)	(0.304)
Diesel price x distance (601 – 734)	-0.440	-0.106	-0.300	0.133	-0.143
	(0.303)	(0.420)	(0.245)	(0.326)	(0.216)
Diesel price x distance (735 – 1098)	-0.972**	0.0786	0.352*	-0.535*	0.164
	(0.397)	(0.574)	(0.203)	(0.317)	(0.309)
Distance (102 - 151) x squared time trends	0.00565	0.0174	0.0120	-0.0105	-0.0113
	(0.0109)	(0.0191)	(0.0156)	(0.0240)	(0.0234)
Distance (152-200) x squared time trends	0.0250*	0.0149	-0.0328**	0.00172	-0.0110
	(0.0151)	(0.0255)	(0.0160)	(0.0247)	(0.0197)
Distance (201 - 267) x squared time trends	0.0153	0.0267	0.00166	-0.0140	-0.0310
	(0.0133)	(0.0280)	(0.0111)	(0.0228)	(0.0196)
Distance (268 – 331) x squared time trends	0.0193	0.0328	-0.000210	-0.00182	-0.0265
	(0.0130)	(0.0260)	(0.0212)	(0.0231)	(0.0183)
Distance (332 – 393) x squared time trends	0.0175	-0.00812	-0.00599	-0.00634	-0.0235
	(0.0147)	(0.0263)	(0.0167)	(0.0275)	(0.0199)
Distance (394 – 479) x squared time trends	0.0250	-0.00595	0.0138	-0.0269	-0.0189
	(0.0230)	(0.0310)	(0.0187)	(0.0265)	(0.0196)
Distance (480 – 600) x squared time trends	0.0852***	0.0374	0.0334*	-0.0425	-0.0205
	(0.0190)	(0.0349)	(0.0186)	(0.0265)	(0.0247)
Distance (601 – 734) x squared time trends	0.0744***	0.0280	0.0415**	-0.0411*	0.00860
	(0.0286)	(0.0293)	(0.0189)	(0.0245)	(0.0173)
Distance (735 – 1098) x squared time trends	0.125***	0.0228	0.0147	-0.0443	0.0102
	(0.0164)	(0.0315)	(0.0187)	(0.0341)	(0.0214)
Distance (102 - 151) x cubic time trends	-0.00000518	-0.0000140	-0.00000997	0.00000790	0.00000722
	(0.00000793)	(0.0000146)	(0.0000117)	(0.0000177)	(0.0000172)
Distance (152-200) x cubic time trends	-0.0000188	-0.0000126	0.0000239**	-0.000000586	0.00000806
	(0.0000115)	(0.0000198)	(0.0000121)	(0.0000182)	(0.0000148)
Distance (201 - 267) x cubic time trends	-0.0000116	-0.0000219	-0.00000187	0.0000109	0.0000226
	(0.00000972)	(0.0000212)	(0.00000865)	(0.0000165)	(0.0000145)
Distance (268 – 331) x cubic time trends	-0.0000149	-0.0000264	-0.00000105	0.00000169	0.0000198
	(0.00000950)	(0.0000198)	(0.0000161)	(0.0000170)	(0.0000135)

Note: Poisson pseudo-likelihood regression with multiple levels of fixed effects (ppmlhdfe) estimator results for each food items are presented in each column. The dependent variable is absolute value of price difference. Route distance over paved road between the market pairs is measured in kilometres, and diesel fuel price in thousands of Malawian Kwacha. Base category is distance below 102 km. Dyadic clustered standard errors at the market in parentheses * p<0.10, ** p<0.05, *** p<0.010

Table A.35 continued…

Dependent variable: $\lvert P_{xt} - P_{yt} \rvert$	Cabbage	Okra	Onions	Pumpkin leaves	Rape leaves
Distance (332 – 393) x cubic time trends	-0.0000139	0.00000336	0.00000326	0.00000445	0.0000175
	(0.0000110)	(0.0000199)	(0.0000126)	(0.0000201)	(0.0000149)
Distance (394 – 479) x cubic time trends	-0.0000200	0.000000414	-0.0000104	0.0000198	0.0000146
	(0.0000171)	(0.0000231)	(0.0000142)	(0.0000195)	(0.0000147)
Distance (480 – 600) x cubic time trends	-0.0000653***	-0.0000324	-0.0000261*	0.0000326*	0.0000170
	(0.0000137)	(0.0000263)	(0.0000142)	(0.0000195)	(0.0000185)
Distance (601 – 734) x cubic time trends	-0.0000578***	-0.0000244	-0.0000317**	0.0000330*	-0.00000498
	(0.0000215)	(0.0000219)	(0.0000141)	(0.0000180)	(0.0000131)
Distance (735 – 1098) x cubic time trends	-0.0000949***	-0.0000188	-0.0000118	0.0000373	-0.00000770
	(0.0000121)	(0.0000228)	(0.0000140)	(0.0000267)	(0.0000157)
Difference in population density	-0.0745	0.00533	-0.0760	0.0895*	0.00547
	(0.0492)	(0.0379)	(0.0548)	(0.0475)	(0.0585)
=1 if one of the markets experienced flood shocks	0.0192	0.00493	-0.0122	-0.00166	0.0302**
	(0.0132)	(0.0152)	(0.0103)	(0.0104)	(0.0136)
=1 if one of the markets experienced drought shocks	0.0467***	0.0302*	0.0535***	-0.0277**	-0.00618
	(0.0148)	(0.0169)	(0.0161)	(0.0129)	(0.0155)
Difference in local production	0.400	0.763	-4.590**	2.938	3.639**
	(1.133)	(2.463)	(2.166)	(1.890)	(1.530)
Month FE	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes
Market-pair FE	Yes	Yes	Yes	Yes	Yes
N	72133	53953	80090	64588	74680

Table A.36: The association between transport costs and market price dispersion of fruits

Dependent variable: $\lvert P_{xt} - P_{yt} \rvert$	Bananas	Tomatoes
Diesel price x distance (102 - 151)	0.227	-0.0566
	(0.214)	(0.208)
Diesel price x distance (152-200)	0.170	-0.234
	(0.202)	(0.267)
Diesel price x distance (201 - 267)	0.198	-0.357
	(0.270)	(0.260)
Diesel price x distance (268 – 331)	0.157	-0.149
	(0.243)	(0.368)
Diesel price x distance (332 – 393)	0.00761	-0.223
	(0.222)	(0.298)
Diesel price x distance (394 – 479)	0.283	-0.867**
	(0.278)	(0.387)
Diesel price x distance (480 – 600)	-0.245	-0.459
	(0.309)	(0.314)
Diesel price x distance (601 – 734)	-0.677***	-0.444
	(0.252)	(0.278)
Diesel price x distance (735 – 1098)	-0.268	-0.171
	(0.241)	(0.234)
Distance (102 - 151) x squared time trends	-0.0465**	0.00868
	(0.0222)	(0.0158)
Distance (152-200) x squared time trends	-0.0501**	0.00823
	(0.0253)	(0.0163)
Distance (201 - 267) x squared time trends	-0.0667**	0.00196
	(0.0283)	(0.0214)
Distance (268 – 331) x squared time trends	-0.0606**	0.00148
	(0.0281)	(0.0212)
Distance (332 – 393) x squared time trends	-0.0549*	-0.00887
	(0.0297)	(0.0182)
Distance (394 – 479) x squared time trends	-0.0975***	0.0220
	(0.0303)	(0.0183)
Distance (480 – 600) x squared time trends	-0.0318	0.0254
	(0.0203)	(0.0174)
Distance (601 – 734) x squared time trends	-0.00735	0.0581***
	(0.0334)	(0.0208)
Distance (735 – 1098) x squared time trends	-0.0216	0.0795***
	(0.0254)	(0.0213)
Distance (102 - 151) x cubic time trends	0.0000348**	-0.00000603
	(0.0000165)	(0.0000119)
Distance (152-200) x cubic time trends	0.0000368*	-0.00000536
	(0.0000188)	(0.0000119)
Distance (201 - 267) x cubic time trends	0.0000491**	-0.000000628
	(0.0000210)	(0.0000160)
Distance (268 – 331) x cubic time trends	0.0000440**	-0.000000112
	(0.0000208)	(0.0000158)

Note: Poisson pseudo-likelihood regression with multiple levels of fixed effects (ppmlhdfe) estimator results for each food items are presented in each column. The dependent variable is absolute value of price difference. Route distance over paved road between the market pairs is measured in kilometres, and diesel fuel price in thousands of Malawian Kwacha. Base category is distance below 102 km. Dyadic clustered standard errors at the market in parentheses * p<0.10, ** p<0.05, *** p<0.010

Table A.36 continued…

Dependent variable: $\lvert P_{xt} - P_{yt} \rvert$	Bananas	Tomatoes
Distance (332 – 393) x cubic time trends	0.0000396*	0.00000709
	(0.0000221)	(0.0000136)
Distance (394 – 479) x cubic time trends	0.0000715***	-0.0000145
	(0.0000228)	(0.0000133)
Distance (480 – 600) x cubic time trends	0.0000234	-0.0000181
	(0.0000151)	(0.0000128)
Distance (601 – 734) x cubic time trends	0.00000628	-0.0000432***
	(0.0000259)	(0.0000154)
Distance (735 – 1098) x cubic time trends	0.0000191	-0.0000603***
	(0.0000191)	(0.0000161)
Difference in population density	-0.0134	-0.0138
	(0.0461)	(0.0382)
=1 if one of the markets experienced flood shocks	0.00482	0.000198
	(0.0144)	(0.0151)
=1 if one of the markets experienced drought shocks	0.0231**	0.00617
	(0.00983)	(0.0148)
Difference in local production	0.213	-4.236***
	(7.082)	(1.259)
Month FE	Yes	Yes
Time FE	Yes	Yes
Market-pair FE	Yes	Yes
N	66053	80558

Table A.37: The association between transport costs and market price dispersion of staples including roots and tubers

Dependent variable: $\lvert P_{xt} - P_{yt}\rvert$	Cassava	Maize flour dehulled	Maize grain (private)	Maize grain (ADMARC)	Potatoes	Rice grain	Sweet potatoes	White bread	White buns
Lag of price difference	0.00671***	0.00387***	0.0126***	0.0214***	0.00598***	0.00525***	0.00964***	0.0147***	0.00344***
	(0.000850)	(0.0000814)	(0.000404)	(0.000616)	(0.000124)	(0.000109)	(0.000345)	(0.00133)	(0.000140)
Diesel price x distance	0.0420***	0.0470***	0.0193***	0.0278***	0.0318***	0.0201***	0.0230***	0.0334***	0.0369***
	(0.00826)	(0.00510)	(0.00368)	(0.00864)	(0.00530)	(0.00258)	(0.00599)	(0.00677)	(0.00662)
Distance-specific linear time trends	0.143**	0.189***	-0.0583	0.0788**	0.0862***	0.0508**	0.145***	0.218***	0.161***
	(0.0697)	(0.0309)	(0.0374)	(0.0342)	(0.0306)	(0.0208)	(0.0370)	(0.0342)	(0.0306)
Distance-specific quadratic time trends	-0.000204*	-0.000280***	0.0000951*	-0.000129**	-0.000120***	-0.0000703**	-0.000208***	-0.000330***	-0.000235***
	(0.000105)	(0.0000473)	(0.0000566)	(0.0000527)	(0.0000463)	(0.0000316)	(0.0000557)	(0.0000527)	(0.0000467)
Distance-specific cubic time trends	9.70e-08*	0.000000138***	-5.11e-08*	7.00e-08***	5.51e-08**	3.24e-08**	9.95e-08***	0.000000166***	0.000000114***
	(5.23e-08)	(2.41e-08)	(2.85e-08)	(2.70e-08)	(2.33e-08)	(1.60e-08)	(2.79e-08)	(2.70e-08)	(2.37e-08)
Difference in population density	0.0305***	0.00853**	0.00629**	-0.000313	0.00918**	0.00864**	0.000267	0.0533***	0.00924*
	(0.00925)	(0.00376)	(0.00285)	(0.00630)	(0.00427)	(0.00357)	(0.00480)	(0.0181)	(0.00497)
=1 if one of the markets experienced flood shocks	0.0452*	0.00835	0.0250*	0.00348	-0.00857	-0.0106	-0.0191	0.0141	0.0122
	(0.0231)	(0.0103)	(0.0149)	(0.0131)	(0.0127)	(0.0157)	(0.0158)	(0.0170)	(0.00742)
=1 if one of the markets experienced drought shocks	0.0147	-0.00617	0.0182*	-0.0429**	0.0278**	0.0175	0.0317**	-0.00807	0.0176*
	(0.0173)	(0.0121)	(0.0104)	(0.0204)	(0.0111)	(0.0167)	(0.0146)	(0.0222)	(0.00984)
Difference in local production	-0.131	0.161	-0.176**	-0.212	0.0954**	-0.0511	0.747***	-	-
	(0.171)	(0.209)	(0.0759)	(0.167)	(0.0436)	(0.0632)	(0.244)	-	-
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Market-pair FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	43639	56761	79475	63981	76778	69285	59578	76107	58424

Note: Instrumental variable Poisson (ivpoisson) estimator with additive errors results for each food items are presented in each column. The dependent variable is absolute value of price difference. Route distance over paved road between the market pairs is measured in hundreds of kilometres, and diesel fuel price in thousands of Malawian Kwacha. The second to forth lag of the price differences are used as instruments for the lagged price differences (*i.e.*, over identified case). Dyadic clustered standard errors at the market in parentheses * p<0.10, ** p<0.05, *** p<0.010

Table A.38: The association between transport costs and market price dispersion of legumes and nuts

Dependent variable: $\lvert P_{xt} - P_{yt} \rvert$	Brown beans	Groundnuts	White beans
Lag of price difference	0.00276***	0.00230***	0.00281***
	(0.0000945)	(0.000130)	(0.000181)
Diesel price x distance	0.00786*	0.0185**	0.0164**
	(0.00438)	(0.00770)	(0.00771)
Distance-specific linear time trends	0.0893***	0.0378	0.110***
	(0.0210)	(0.0348)	(0.0348)
Distance-specific quadratic time trends	-0.000126***	-0.0000502	-0.000164***
	(0.0000323)	(0.0000528)	(0.0000531)
Distance-specific cubic time trends	5.94e-08***	2.21e-08	8.12e-08***
	(1.65e-08)	(2.66e-08)	(2.69e-08)
Difference in population density	0.0208***	0.0237*	0.0438***
	(0.00551)	(0.0125)	(0.00722)
=1 if one of the markets experienced flood shocks	-0.0407***	-0.0121	-0.0202
	(0.0120)	(0.0114)	(0.0164)
=1 if one of the markets experienced drought shocks	0.0115	-0.00445	-0.00289
	(0.0142)	(0.0128)	(0.0138)
Difference in local production	0.689***	-0.0622	0.826***
	(0.256)	(0.0656)	(0.194)
Month FE	Yes	Yes	Yes
Market-pair FE	Yes	Yes	Yes
N	71593	50484	42337

Note: Instrumental variable Poisson (ivpoisson) estimator with additive errors results for each food items are presented in each column. The dependent variable is absolute value of price difference. Route distance over paved road between the market pairs is measured in hundreds of kilometres, and diesel fuel price in thousands of Malawian Kwacha. The second to forth lag of the price differences are used as instruments for the lagged price differences (*i.e.*, over identified case). Dyadic clustered standard errors at the market in parentheses * p<0.10, ** p<0.05, *** p<0.010

Table A.39: The association between transport costs and market price dispersion of animal source foods

Dependent variable: $\lvert P_{xt} - P_{yt} \rvert$	Beef	Eggs	Ultra- pasteurized milk	Goat meat	Powdered milk	Usipa	Utaka
Lag of price difference	0.00292***	0.00391***	0.00839***	0.00314***	0.000537***	0.000465***	0.000446***
	(0.000163)	(0.000250)	(0.000401)	(0.000248)	(0.0000166)	(0.0000116)	(0.0000162)
Diesel price x distance	0.0232***	0.0184**	0.0300**	0.0383***	0.0658***	0.00814*	-0.000797
	(0.00767)	(0.00837)	(0.0123)	(0.00793)	(0.0112)	(0.00419)	(0.00994)
Distance-specific linear time trends	0.0730**	0.265***	0.153***	0.0618	0.153***	-0.0809***	-0.203***
	(0.0345)	(0.0663)	(0.0401)	(0.0422)	(0.0559)	(0.0272)	(0.0453)
Distance-specific quadratic time trends	- 0.000107**	-0.000401***	-0.000222***	-0.0000863	-0.000225***	0.000134***	0.000325***
	(0.0000532)	(0.000100)	(0.0000599)	(0.0000640)	(0.0000848)	(0.0000421)	(0.0000686)
Distance-specific cubic time trends	5.25e-08*	0.000000201***	0.000000107***	3.99e-08	0.000000110**	-7.28e-08***	- 0.000000171***
	(2.72e-08)	(5.03e-08)	(2.99e-08)	(3.24e-08)	(4.28e-08)	(2.16e-08)	(3.46e-08)
Difference in population density	-0.00623	0.0120**	0.0240	-0.0139**	-0.00722	0.0363***	0.0558***
	(0.00549)	(0.00531)	(0.0157)	(0.00692)	(0.00572)	(0.00465)	(0.00454)
=1 if one of the markets experienced flood shocks	0.0171***	0.0204	-0.00454	0.0185**	-0.000544	0.00575	0.0230**
	(0.00575)	(0.0401)	(0.0101)	(0.00887)	(0.00984)	(0.0133)	(0.0116)
=1 if one of the markets experienced drought shocks	-0.0239**	-0.0176	0.00548	0.00182	-0.00357	-0.0213	-0.0377**
	(0.0101)	(0.0299)	(0.0105)	(0.0139)	(0.0162)	(0.0134)	(0.0192)
Difference in local production	-0.272	0.405***	-	0.518***	-	-0.127**	-0.0800***
	(0.231)	(0.115)	-	(0.143)	-	(0.0598)	(0.0278)
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Market-pair FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	69164	62639	52783	68845	31186	65157	39629

Note: Instrumental variable Poisson (ivpoisson) estimator with additive errors results for each food items are presented in each column. The dependent variable is absolute value of price difference. Route distance over paved road between the market pairs is measured in hundreds of kilometres, and diesel fuel price in thousands of Malawian Kwacha. The second to forth lag of the price differences are used as instruments for the lagged price differences (*i.e.*, over identified case). Dyadic clustered standard errors at the market in parentheses * p<0.10, ** p<0.05, *** p<0.010

Table A.40: The association between transport costs and market price dispersion of vegetables

Dependent variable: $\lvert P_{xt} - P_{yt} \rvert$	Cabbage	Okra	Onions	Pumpkin leaves	Rape leaves
Lag of price difference	0.0132***	0.00331***	0.00293***	0.00451***	0.00527***
	(0.000308)	(0.000110)	(0.0000857)	(0.000174)	(0.000170)
Diesel price x distance	0.0519***	0.0635***	0.0235***	0.0169***	0.0200***
	(0.00681)	(0.0116)	(0.00326)	(0.00484)	(0.00709)
Distance-specific linear time trends	0.0480	0.106**	0.126***	0.145***	0.127***
	(0.0312)	(0.0414)	(0.0254)	(0.0190)	(0.0213)
Distance-specific quadratic time trends	-0.0000670	-0.000152**	-0.000184***	-0.000212***	-0.000181***
	(0.0000482)	(0.0000637)	(0.0000390)	(0.0000287)	(0.0000321)
Distance-specific cubic time trends	3.10e-08	7.24e-08**	8.88e-08***	0.000000104***	8.60e-08***
	(2.47e-08)	(3.25e-08)	(1.99e-08)	(1.44e-08)	(1.61e-08)
Difference in population density	-0.00436	-0.00871	0.0266***	0.0254***	0.0212***
	(0.00484)	(0.00715)	(0.00347)	(0.00481)	(0.00479)
=1 if one of the markets experienced flood shocks	0.0125	0.0338	0.00755	0.0161	-0.00454
	(0.0138)	(0.0247)	(0.0139)	(0.0127)	(0.0103)
=1 if one of the markets experienced drought shocks	0.0558***	0.0233	-0.00921	0.0305***	-0.00793
	(0.0203)	(0.0185)	(0.0113)	(0.0112)	(0.0102)
Difference in local production	-0.182**	-0.292	0.0154	-0.740**	-0.480
	(0.0760)	(0.503)	(0.0140)	(0.293)	(0.448)
Month FE	Yes	Yes	Yes	Yes	Yes
Market-pair FE	Yes	Yes	Yes	Yes	Yes
N	55174	31842	70571	42357	60474

Note: Instrumental variable Poisson (ivpoisson) estimator with additive errors results for each food items are presented in each column. The dependent variable is absolute value of price difference. Route distance over paved road between the market pairs is measured in hundreds of kilometres, and diesel fuel price in thousands of Malawian Kwacha. The second to forth lag of the price differences are used as instruments for the lagged price differences (*i.e.*, over identified case). Dyadic clustered standard errors at the market in parentheses * p<0.10, ** p<0.05, *** p<0.010

Table A.41: The association between transport costs and market price dispersion of fruits

Dependent variable: $\lvert P_{xt} - P_{yt} \rvert$	Bananas	Tomatoes
	0.00672***	0.00364***
	(0.000354)	(0.0000818)
Diesel price x distance	0.0411***	0.00949*
	(0.0103)	(0.00528)
Distance-specific linear time trends	0.106**	0.0521**
	(0.0456)	(0.0237)
Distance-specific quadratic time trends	-0.000146**	-0.0000679*
	(0.0000702)	(0.0000365)
Distance-specific cubic time trends	6.67e-08*	2.94e-08
	(3.59e-08)	(1.86e-08)
Difference in population density	0.00867	0.0126***
	(0.00581)	(0.00263)
=1 if one of the markets experienced flood shocks	0.000195	0.000957
	(0.0132)	(0.0129)
=1 if one of the markets experienced drought shocks	0.0146	0.0326***
	(0.00927)	(0.0114)
Difference in local production	0.579	0.358**
	(0.359)	(0.148)
Month FE	Yes	Yes
Market-pair FE	Yes	Yes
N	44336	73517

Note: Instrumental variable Poisson (ivpoisson) estimator with additive errors results for each food items are presented in each column. The dependent variable is absolute value of price difference. Route distance over paved road between the market pairs is measured in hundreds of kilometres, and diesel fuel price in thousands of Malawian Kwacha. The second to forth lag of the price differences are used as instruments for the lagged price differences (*i.e.*, over identified case). Dyadic clustered standard errors at the market in parentheses * p<0.10, ** p<0.05, *** p<0.010

Table A.42: The association between transport costs and market price dispersion of staples including roots and tubers

Dependent variable: $\lvert P_{xt} - P_{yt} \rvert$	Cassava	Maize flour dehulled	Maize grain (private)	Maize grain (ADMARC)	Potatoes	Rice grain	Sweet potatoes	White bread	White buns
Lag of price difference	0.364***	0.303***	0.300***	0.746***	0.341***	0.331***	0.262***	0.480***	0.513***
	(0.0224)	(0.0130)	(0.00841)	(0.00727)	(0.0105)	(0.00774)	(0.0133)	(0.0225)	(0.0151)
Diesel price x distance	15.97***	10.82***	1.055	4.424***	5.296**	6.515***	-3.870***	2.892***	1.135
	(2.850)	(2.987)	(0.995)	(0.567)	(2.282)	(2.256)	(1.376)	(0.578)	(1.729)
Distance-specific linear time trends	23.26***	26.80***	-5.075***	-6.735***	-17.33***	10.16***	5.018*	4.766***	5.926
	(6.124)	(5.227)	(1.256)	(0.898)	(3.389)	(3.069)	(2.924)	(1.030)	(4.771)
Distance-specific quadratic time trends	-0.0368***	-0.0412***	0.00794***	0.0104***	0.0273***	-0.0153***	-0.00816*	-0.00732***	-0.00912
	(0.00987)	(0.00820)	(0.00198)	(0.00136)	(0.00535)	(0.00476)	(0.00461)	(0.00160)	(0.00737)
Distance-specific cubic time trends	0.0000193***	0.0000210***	-0.00000411***	-0.00000534***	-0.0000143***	0.00000769***	0.00000442*	0.00000372***	0.00000464
	(0.00000529)	(0.00000426)	(0.00000104)	(0.000000686)	(0.00000280)	(0.00000244)	(0.00000241)	(0.000000827)	(0.00000377)
Difference in population density	14.46**	7.144	-8.561***	1.109	4.848	-27.63***	5.493	-7.814***	-33.95**
	(7.001)	(14.43)	(2.212)	(1.690)	(9.126)	(9.671)	(3.468)	(1.897)	(15.24)
=1 if one of the markets experienced flood shocks	-0.391	0.639	0.625*	-0.551**	-0.0112	-3.127***	0.289	-0.150	1.805***
	(0.688)	(1.101)	(0.350)	(0.234)	(0.628)	(0.935)	(0.449)	(0.209)	(0.698)
=1 if one of the markets experienced drought shocks	0.177	0.455	0.729**	-0.361	-0.255	0.168	-1.285***	-0.732***	-0.977
	(0.503)	(1.113)	(0.334)	(0.256)	(0.589)	(0.919)	(0.461)	(0.204)	(0.854)
Difference in local production	-67.28	-335.8***	-207.4***	-85.57*	14.50	36.02	-42.96	-	-
	(43.54)	(85.95)	(31.57)	(43.73)	(37.75)	(38.24)	(43.27)	-	-
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Market-pair FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	30453	49545	75529	65286	70241	64222	46963	72352	53025

Note: Arellano–Bond estimator results for each food items are presented in each column. The dependent variable is absolute value of price difference. Route distance over paved road between the market pairs is measured in hundreds of kilometres, and diesel fuel price in thousands of Malawian Kwacha. Robust standard errors in parentheses * p<0.10, ** p<0.05, *** p<0.010

Table A.43: The association between transport costs and market price dispersion of legumes and nuts

Dependent variable: $\lvert P_{xt} - P_{yt} \rvert$	Brown beans	Groundnuts	White beans
Lag of price difference	0.112***	0.266***	0.234***
	(0.00780)	(0.0154)	(0.0124)
Diesel price x distance	39.51***	47.15***	24.08***
	(4.690)	(5.423)	(4.614)
Distance-specific linear time trends	-17.64***	16.42**	60.88***
	(4.960)	(7.599)	(10.24)
Distance-specific quadratic time trends	0.0288***	-0.0248**	-0.0955***
	(0.00775)	(0.0120)	(0.0164)
Distance-specific cubic time trends	-0.0000157***	0.0000123**	0.0000495***
	(0.00000401)	(0.00000627)	(0.00000869)
Difference in population density	-27.03**	-10.75	-23.33
	(11.47)	(11.84)	(14.71)
=1 if one of the markets experienced flood shocks	-3.657**	0.999	-5.505***
	(1.454)	(1.591)	(1.355)
=1 if one of the markets experienced drought shocks	1.064	-1.404	-0.402
	(1.666)	(1.304)	(1.212)
Difference in local production	23.47	1.880	126.9
	(171.8)	(5.891)	(140.6)
Month FE	Yes	Yes	Yes
Time FE	Yes	Yes	Yes
Market-pair FE	Yes	Yes	Yes
N	75514	57488	49701

Note: Arellano–Bond estimator results for each food items are presented in each column. The dependent variable is absolute value of price difference. Route distance over paved road between the market pairs is measured in hundreds of kilometres, and diesel fuel price in thousands of Malawian Kwacha. Robust standard errors in parentheses * p<0.10, ** p<0.05, *** p<0.010

Table A.44: The association between transport costs and market price dispersion of animal source foods

Dependent variable: $\lvert P_{xt} - P_{yt} \rvert$	Beef	Eggs	Ultra-pasteurized milk	Goat meat	Powdered milk	Usipa	Utaka
Lag of price difference	0.742***	0.366***	0.785***	0.729***	0.724***	0.258***	0.150***
	(0.00832)	(0.0119)	(0.00844)	(0.00787)	(0.00943)	(0.00889)	(0.0104)
Diesel price x distance	-3.014**	9.132***	-0.742	9.656***	16.47	3.121	336.0***
	(1.263)	(2.562)	(0.574)	(1.792)	(10.32)	(26.12)	(60.72)
Distance-specific linear time trends	-12.48***	-4.624	-7.785***	-6.246***	172.2***	-271.3***	-367.0***
	(2.038)	(4.185)	(1.076)	(2.035)	(51.49)	(47.34)	(88.62)
Distance-specific quadratic time trends	0.0202***	0.00884	0.0124***	0.00980***	-0.259***	0.420***	0.586***
	(0.00318)	(0.00646)	(0.00168)	(0.00318)	(0.0790)	(0.0750)	(0.140)
Distance-specific cubic time trends	-0.0000108***	-0.00000539	-0.00000658***	-0.00000507***	0.000129***	-0.000216***	-0.000311***
	(0.00000164)	(0.00000331)	(0.000000873)	(0.00000164)	(0.0000401)	(0.0000395)	(0.0000733)
Difference in population density	-22.87*	-1.859	24.15***	-25.04**	218.3	-108.0	107.2
	(11.69)	(24.15)	(5.714)	(10.14)	(141.6)	(92.89)	(84.26)
=1 if one of the markets experienced flood shocks	-0.832	0.970	-1.156***	-1.275*	-20.14**	4.848	-6.226
	(0.705)	(1.023)	(0.289)	(0.710)	(8.713)	(7.868)	(10.81)
=1 if one of the markets experienced drought shocks	0.313	0.341	-1.027***	0.615	14.54	-29.24***	14.68
	(0.766)	(1.245)	(0.326)	(0.740)	(11.14)	(7.500)	(10.85)
Difference in local production	-1052.1	-1080.9***	-	-799.1***	-	810.4	-656.3
	(838.8)	(263.6)	-	(176.2)	-	(711.3)	(455.7)
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Market-pair FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	71570	66795	57156	71619	35550	71103	46055

Note: Arellano–Bond estimator results for each food items are presented in each column. The dependent variable is absolute value of price difference. Route distance over paved road between the market pairs is measured in hundreds of kilometres, and diesel fuel price in thousands of Malawian Kwacha. Robust standard errors in parentheses * p<0.10, ** p<0.05, *** p<0.010

Table A.45: The association between transport costs and market price dispersion of vegetables

Dependent variable: $\lvert P_{xt} - P_{yt} \rvert$	Cabbage	Okra	Onions	Pumpkin leaves	Rape leaves
Lag of price difference	0.318***	0.237***	0.289***	0.256***	0.265***
	(0.0101)	(0.0155)	(0.00938)	(0.0161)	(0.0105)
Diesel price x distance	2.724**	-28.03***	9.577***	10.06***	1.474
	(1.213)	(5.683)	(3.386)	(3.203)	(2.014)
Distance-specific linear time trends	-8.409***	-48.63***	-18.63***	15.86**	3.424
	(2.192)	(14.58)	(6.169)	(6.660)	(3.318)
Distance-specific quadratic time trends	0.0130***	0.0775***	0.0312***	-0.0250**	-0.00510
	(0.00346)	(0.0233)	(0.00973)	(0.0105)	(0.00522)
Distance-specific cubic time trends	-0.00000667***	-0.0000407***	-0.0000172***	0.0000131**	0.00000248
	(0.00000182)	(0.0000124)	(0.00000510)	(0.00000548)	(0.00000272)
Difference in population density	-4.286*	-21.12**	-27.46*	4.761	36.49***
	(2.501)	(10.25)	(15.31)	(13.87)	(11.65)
=1 if one of the markets experienced flood shocks	-0.312	1.365	-2.659*	-2.244**	-0.508
	(0.393)	(1.631)	(1.445)	(1.128)	(0.680)
=1 if one of the markets experienced drought shocks	1.074***	5.850***	3.605***	2.618***	0.117
	(0.313)	(1.333)	(1.342)	(1.003)	(0.563)
Difference in local production	165.0	-423.6***	-1108.5	-531.2***	-73.23
	(245.9)	(133.2)	(941.2)	(102.3)	(53.76)
Month FE	Yes	Yes	Yes	Yes	Yes
Market-pair FE	Yes	Yes	Yes	Yes	Yes
N	62557	40669	74940	51618	66971

Note: Arellano–Bond estimator results for each food items are presented in each column. The dependent variable is absolute value of price difference. Route distance over paved road between the market pairs is measured in hundreds of kilometres, and diesel fuel price in thousands of Malawian Kwacha. Robust standard errors in parentheses * p<0.10, ** p<0.05, *** p<0.010

Table A.46: The association between transport costs and market price dispersion of fruits

Dependent variable: $\lvert P_{xt} - P_{yt} \rvert$	Bananas	Tomatoes
Lag of price difference	0.185***	0.307***
	(0.0149)	(0.00807)
Diesel price x distance	6.879***	-5.055
	(2.122)	(3.562)
Distance-specific linear time trends	-14.75***	-16.30***
	(4.489)	(4.192)
Distance-specific quadratic time trends	0.0232***	0.0249***
	(0.00714)	(0.00662)
Distance-specific cubic time trends	-0.0000120***	-0.0000125***
	(0.00000377)	(0.00000346)
Difference in population density	16.17*	42.35***
	(8.452)	(14.37)
=1 if one of the markets experienced flood shocks	0.756	1.922
	(0.617)	(1.190)
=1 if one of the markets experienced drought shocks	1.107**	6.389***
	(0.476)	(1.184)
Difference in local production	322.4	-620.6***
	(539.4)	(222.2)
Month FE	Yes	Yes
Market-pair FE	Yes	Yes
N	53232	76879

Note: Arellano–Bond estimator results for each food items are presented in each column. The dependent variable is absolute value of price difference. Route distance over paved road between the market pairs is measured in hundreds of kilometres, and diesel fuel price in thousands of Malawian Kwacha. Robust standard errors in parentheses * p<0.10, ** p<0.05, *** p<0.010

Chapter 3

What are distributional effects of the fuel subsidy removal on household maize production and welfare in Malawi?

3.1 Introduction

Enhancing market participation through lower transport and input costs is often perceived to improve households' well-being in developing countries (Christiaensen & Demery, 2018; Headey et al., 2019; Jones, 2017; Koppmair et al., 2017; Nkegbe & Abdul Mumin, 2022; Olabisi et al., 2021; Sibhatu et al., 2015; Stifel & Minten, 2017). According to Christiaensen & Demery (2018), improved households' access to markets has the potential to increase agricultural productivity through increased use of modern inputs such as inorganic fertilisers and hybrid seeds, increase returns, and improve nutrition outcomes in rural areas. Previous studies confirm that transport and input costs are major barriers that reduce market participation among households in rural areas (Damania et al., 2016; Minten et al., 2013; Omamo, 1998; Stifel & Minten, 2017; Vandercasteelen et al., 2018). To reduce domestic transport costs, most sub-Saharan African (SSA) governments such as Ethiopia, Ghana, Malawi, Namibia, and Niger adopted fuel subsidies prior to the 2007/08 international oil price shock. During this period, international fuel price pass-through to domestic fuel price was lower in SSA than in most developed countries (International Monetary Fund, 2013). However, the surge in international oil prices in 2007/08 increased fiscal costs and led to removal or reduction in fuel subsidies, which increased domestic fuel prices and transport costs in most countries. The removal or reduction in fuel subsidies was not uniform across SSA countries (International Monetary Fund, 2013). For instance, Ethiopia removed its fuel subsidy immediately in 2008, while Malawi continued to implement its fuel subsidy until May 2012.

The objective of this chapter is to provide insights on how the removal of the fuel subsidy differentially affected households in Malawi. We do this by estimating both the immediate and persistent differential effects of the reform to the fuel policy adopted in 2012 on staple maize production and consumption. The Government re-introduced automatic price adjustment mechanism in 2012 that was abandoned in 2004 to sustain fuel supply, which led to an increase in fuel prices and transport costs. Prior to the fuel price reform, liquid fuels were being sold at subsidised prices and events of fuel shortages were common across the country. Thus, the reform increased the cost of transporting produce from the farm to the market or consumption centre, and inputs from the market to the farm using motorised transportation. We hypothesise that the policy reform has immediate differential effects on staple maize production and consumption, but the differential effects do not persist over time once the policy is adopted. Thus, we expect the differential effects of the policy reform to go away from its initial impact as households adapt to cope with the effects dampening off over time. Consistent with the previous literature, we anticipate a heterogeneous differential impact of the fuel price reform on households that varies with household market position as a net-seller, a net-buyer or self-sufficient in staple maize grain, and market access (Deaton, 1989; Fuje, 2019; Hasan, 2016; Minot & Goletti, 1998; Omamo, 1998). Net sellers are households whose quantity of staple maize grain sold on the market is greater than the quantity of staple maize grain purchased, net buyers are households whose quantity of staple maize grain purchased on the market is greater than the quantity of staple maize grain sold, and self-sufficient households are those that do not purchase or sell any staple maize grain on the market (Aksoy & Isik-Dikmelik, 2008; World Food Programme, 2009). Our measure of market access, which involves transport costs from households' location to the market, is the minimum distance to the closest local large agricultural market or consumption centre in kilometres as the crow flies. We anticipate the policy to have a larger effect on net sellers and net buyers relative to self-sufficient households of staple maize grain that varies with the level of market access.

To estimate the immediate and persistent differential impacts of the fuel policy reform on households, we use three waves of nationally representative panel data from the Integrated Household Panel Survey (IHPS), which were implemented in 2010, 2013, and 2016 as part of the Living Standards Measurement Study-Integrated Surveys of Agriculture (LSMS-ISA) for Malawi. The enactment of the policy in 2012 provides us with a natural experimental setting to conduct our analysis, where the 2010 data represents the period before the reform while the 2013 and 2016 data represent the period after the reform. Thus, the period after the fuel subsidy removal represents the period of higher transport costs for households than the period before the fuel policy reform to access markets. We explicitly examine whether any pre-existing differences on staple maize production and consumption across household groups persisted after the policy reform. Any break in pre-existing differences in the level or trend of staple maize production and consumption closer to the time of the reform in 2012 and then further away from the reform period is our estimate of the causal impact of the policy reform (Finkelstein, 2007; Sun & Shapiro, 2022). We use the data between 2010 and 2013 (*i.e.*, one year after the policy reform) to estimate immediate differential effects of the policy reform, and the data between 2010 and 2016 (*i.e.*, four years after the policy reform) to estimate persistent differential effects on maize production and consumption among households using a fixed effects estimator.

This present chapter adds an important dimension to the literature on transport costs in SSA by estimating how the fuel subsidy removal differentially affected staple maize production and consumption. There are only a limited number of studies that investigate the effects of transport costs on production and consumption of farm produce by households in developing countries using observational cross-sectional data via econometric estimation. Using various measures of transport costs, these studies find that an increase in transport costs results in an increase in the production of food for own consumption among net buyers, reduces the supply of food on the market among net sellers, increases the price farmers pay for inputs but reduces profitability and the intensification of input use, agricultural productivity, farm mechanisation, crop revenue, food security, and consumption of more diversified diets in the most remote areas (Damania et al., 2016; Fuje, 2019; Minten et al., 2013; Omamo, 1998; Vandercasteelen et al., 2018). The closest study to ours is Fuje (2019) that examines the impact of fuel subsidy reforms on real incomes for net buyers and net sellers of grains (*i.e.*, "teff", wheat, maize, sorghum, and barley) via a non-parametric approach in Ethiopia. The author finds mixed results of the impact of the subsidy removal on households where most rural households lost in terms of real income because most grain stayed local which lowered grain prices, others gained or did not experience any change in their real income, and fewer households in urban areas had a decrease in real income. Building on Fuje (2019) and the previous studies, ours is the first study to estimate the differential impacts of the fuel policy reform on households using nationally representative observational panel data from SSA via a parametric approach.

Our analysis confirms that there are heterogeneous differential impacts of the fuel price reform on households that vary with household status and market access. In contrast with our expectations, we find that there are both short- and long-term consequences of the fuel policy reform on staple maize production and consumption. Thus, households do not dampen off the effects of increasing transport costs in the long run. Overall, our results indicate that households that are in autarky in remote areas increased maize production more than those closer to the market but lost in consumption due to the increase in transport costs of accessing markets. Households that are net buyers that reside closer to the market increased maize production, consumption, and became less prone to maize insecurity, while those that reside in remote locations lost in non-food consumption and became more prone to maize insecurity relative to households that are in autarky. Conversely, households that are net sellers that reside in remote locations lost in non-food consumption and maize consumption, while those that reside closer to the market lost in consumption, non-food consumption and non-maize food consumption relative to households that are in autarky. These differential effects are less sensitive to how market access is measured.

The rest of the chapter is organised as follows. The next section presents an overview of agriculture in Malawi. Section 3 provides an overview of fuel pricing in Malawi, while section 4 reviews related literature. Section 5 describes the theoretical framework, empirical strategy, and data used in the study. Findings are presented in section 6 and section 7 concludes.

3.2 An overview of agriculture in Malawi

Agriculture remains the main source of livelihood, food security and nutrition in Malawi (Government of Malawi, 2016a). National Statistical Office (2013) reports that about 64 percent of the labour is employed in Agriculture. Agriculture is dominated by smallholder farming households who cultivate food crops such as cereals (maize, rice, and sorghum), roots and tubers (potatoes, cassava, and sweet potatoes), and legumes (beans, groundnuts, soybeans) on less than one hectare mainly under rain-fed production to meet their food and cash needs (Benson & Weerdt, 2023; Chirwa, 2006; Government of Malawi, 2016a). Maize grain is Malawi's staple food; hence, important for food security among households. For instance, over 90 percent of the households cultivated maize during the 2012/13 cropping season (National Statistical Office, 2014a; Sibande et al., 2017). Smallholder farming households usually intercrop maize with other crops such as beans, and peas. Thus, smallholder farming is diversified as demonstrated in the previous studies using integrated household survey data (Jones et al., 2014; Snapp & Fisher, 2015).

Agricultural productivity among households is lower due to limited access to modern inputs such as inorganic fertilisers and hybrid seeds (Benson & Weerdt, 2023; Government of Malawi, 2016a). To increase agricultural productivity of maize, the government of Malawi has been providing farmers with subsidised inputs such as inorganic fertiliser and hybrid seed through the Input Subsidy Programme since 2005. Sheahan & Barrett (2017) find an increase in average fertiliser application rate of about 146 kg/ha in a maize field using the 2010/11 Integrated Household Survey data, which is within the recommended application rate of about 100-250 kg/ha in Malawi (Benson, 1999). According to Government of Malawi (2016), subsidised inputs have increased maize productivity from 1,300 kg per ha to 2,000 kg per ha (*i.e.*, 54%) among programme beneficiaries and availability of maize for consumption across the country since its inception.

Despite an increase in available maize for consumption, diets in Malawi are not diversified, child-malnutrition is high, and there are food safety issues relating to high levels of aflatoxin that build up in maize and groundnuts (Government of Malawi, 2016a). Food expenditure shares have increased by 0.9 percentage points (*i.e.*, from 61.7% to 62.6%) among households between 2004 and 2011 (Pauw et al., 2015, 2016), suggesting that food expenditure shares are larger across households than non-food expenditure shares in Malawi. Further, Pauw et al. (2015) find increased consumption of staple maize grain, fruits, and animal source foods, and reduced consumption of vegetables, and cassava between 2004 and 2011 using integrated household survey data. A recent poverty analysis report shows that households in urban centres consume more per capita (MWK395, 706) than those in rural areas (MWK185, 418) (Government of Malawi, 2020a). Further, Minot (2010) reports that per capita maize consumption is about 133 kg and contributes about 54 percent of households' consumed calories, which is relatively large in the eastern and southern Africa.

Household maize market position

Farmers' participation in agricultural markets is constrained by several factors that include inadequate transportation and market infrastructure, inadequate market information, and limited access to commercial services especially in the rural areas (Government of Malawi, 2016a). As a result, farmers in the rural areas face higher costs to access markets and lower returns than those in the urban centres. Analysis of integrated household survey (IHS) data has shown that fewer households participate in maize markets as net sellers than as net buyers and autarkic households. The proportion of households that participate in the market as net sellers, net buyers, and autarkic households varies over time. For instance, Chirwa (2006) finds that from 56.2 percent of the households that cultivated maize only 9.2 percent participated in the maize market as net sellers using the 1998 Malawi's IHS data. Using the 2004 IHS data, the proportion of farming households participating in maize market as net sellers increased (14.5%) during the 2003/2004 agricultural season (Chirwa, 2009). Dorward et al., (2008) find that about 10 percent of farming households are net sellers, 60 percent are net buyers, and 30 percent are maize autarkic households using the 2007 household survey data that the National Statistical Office collected by re-interviewing some of the households that were part of the 2004 IHS in May through to June 2007 across the country. Further, Sibande et al. (2017) find that about 11 percent of the households that cultivated maize participated in the maize market as net sellers between 2010 and 2013 using IHPS data for Malawi. These findings suggest that there are fewer faming households that produce maize surplus, and most farming households are net buyers of maize in Malawi compared to other developing countries such as Zambia, Kenya, and Ethiopia. For instance, Jayne et al., (2006) find that 26 percent are maize net sellers, 36 are net buyers, and 39 are autarkic households in Zambia, whereas 26 percent are maize net sellers, 62 percent are net buyers, and 8 percent are autarkic households in Kenya, and 25 percent are maize and teff net sellers, 73 percent are net buyers, and 2 percent are autarkic households in Ethiopia. Similarly, Mason & Ricker-Gilbert (2013) find that 25 percent of the farming households are net sellers, 37 percent net buyers, and 25 percent are maize self-sufficient in Zambia.

3.3 Fuel pricing in Malawi

Malawi imports its fuel from countries such as United Arab Emirates, Kuwait, South Africa, India, and Switzerland (Cammack, 2012; Innovex Development Consulting Ltd, 2020). The fuel supplies come through the ports of Beira and Nacala in Mozambique, and Dar es Salaam in Tanzania. Being a landlocked country, road transportation is the main means of getting fuel into Malawi. Like most fuel importing countries, Malawi's fuel pump price is determined by the price of refined petroleum product on the international market, exchange rates, transport costs, excise duties, and domestic margins (Kpodar & Djiofack, 2009). In addition, the Government imposes various fuel levies to the fuel pump price such as energy regulatory levy to fund the Malawi Energy Regulatory Authority (MERA), road maintenance levy, Malawi Bureau of Standards access levy, safety net levy to finance development projects such as fertiliser subsidy program, levy to contribute to the Price Stabilization Fund (PSF), and in-bond landed cost recovery levy for compensating losses that fuel importers incur (Bacon & Kojima, 2006; Innovex Development Consulting Ltd, 2020; Kojima, 2013). Bacon & Kojima (2006) report that taxes comprised about 11-20 percent whereas levies comprised about 21-24 percent of the retail fuel price in June 2005, suggesting that levies make up a larger share of the fuel price than taxes.

According to Robinson & Wakeford (2013), Malawi has been blending its imported liquid fuel with ethanol using a blending ratio of 10 percent ethanol to 90 percent petrol since the 1970s energy crisis. During the early 1990s, fuel importation was fully privatized and Oil Marketing Companies (OMC) such as Puma Energy, Total Energies, and Petroda Malawi Limited through Industry Petroleum Supply Unit (IPSU) imported fuel for domestic consumption (Government of Malawi, 2021a). During that time, the Government established a Petroleum Control Commission (PCC) to regulate liquid fuel and the PSF to stabilize fuel pump prices and adopted an automatic pricing formula to determine the price of fuel (Bacon & Kojima, 2006). The price of fuel could be adjusted when the import fuel price in Malawi Kwacha changed by more than 5 percent. When the changes in the landed costs of fuel were smaller than 5 percent, PCC could use PSF to compensate the OMCs for losses to maintain the prevailing liquid fuel pump prices (Bacon & Kojima, 2006; Kojima, 2013). This means that the Government could deplete PSF when the international oil price increased by less than 5 percent (*i.e.*, pay importers when fuel is expensive globally) and could replenish PSF when the international oil price declined by less than 5 percent (*i.e.*, tax importers when fuel is cheaper globally) to stabilise liquid fuel pump prices. Thus, the PSF was expected to be self-financing (Kojima, 2013). However, PSF recorded the largest deficit in 2008 (Kojima, 2009; Kojima et al., 2010).

The flotation of the Malawi Kwacha resulted in IPSU incurring heavy exchange losses, which influenced the company to hand over fuel importation to PCC in 1994. Thus, PCC became the regulator and importer of liquid fuel at that time. Then, the private sector established a consortium of OMCs called Petroleum Importer Limited (PIL), which took over fuel importation from PCC over concerns that PCC combined the role of fuel regulation and importation, which was not in line with good governance in 1999. During this period, Malawi was vulnerable to supply chain shocks since the private sector did not have incentives to invest in fuel storage for more than 10 days (Kojima, 2013; Kojima et al., 2010; Robinson & Wakeford, 2013).

To increase the limited import cover for fuel, the Government built Strategic Fuel Reserves facilities with a 60-day demand cover to ensure both stability and security in the supply of fuel and gas products in the country. Then, it established the National Oil Company of Malawi (NOCMA) Limited to manage Strategic Fuel Reserves facilities in 2010 (Government of Malawi, 2021a). NOCMA is also involved in the liquid fuel imports and supply about 50% of its fuel reserves to OMCs through PIL and to other independent fuel retail business operators (Government of Malawi, 2021b). Since the establishment of NOCMA, PIL was allocated 50% quota to continue importing fuel and supplying its members. Further, to regulate the energy sector in Malawi, the Government established MERA to replace the PCC and the National Electricity Council in 2007. This means that MERA also regulates NOCMA.

Bacon & Kojima (2006) report that Malawi abandoned the automatic pricing formula in 2004, and fuel price adjustments were subjected to discretionary powers. The result of this policy change was that OMCs incurred losses whenever there was an increase in the international oil price because the Government delayed adjusting the pump prices while PSF accumulated surpluses. In May 2012, the Government re-introduced automatic price adjustment mechanism and floated the Malawi Kwacha to sustain fuel supply in the country following events of fuel shortages between 2009 and 2012, which culminated into national protests in July 2011 (Government of Malawi, 2018; Kojima, 2013). Figure 3.1 shows that fuel prices gradually increased in a step fashion over time. Immediately after the policy reform (depicted by the red line in Figure 3.1), fuel prices increased by 32 percent from April to May in 2012 and reached its maximum price in March 2013 (depicted by the blue line in Figure 3.1).[34] Thus, the increase in fuel prices was fully realised in March 2013 when fuel prices had increased by 93 percent following the enactment of the fuel policy reform. The Government through MERA

[34] Figure B.1 in the appendix shows that the increase in fuel price was sharper after the policy reform than before the reform.

regulates liquid pump prices, which are uniform throughout the country (Kojima, 2012, 2013; Kojima et al., 2010).

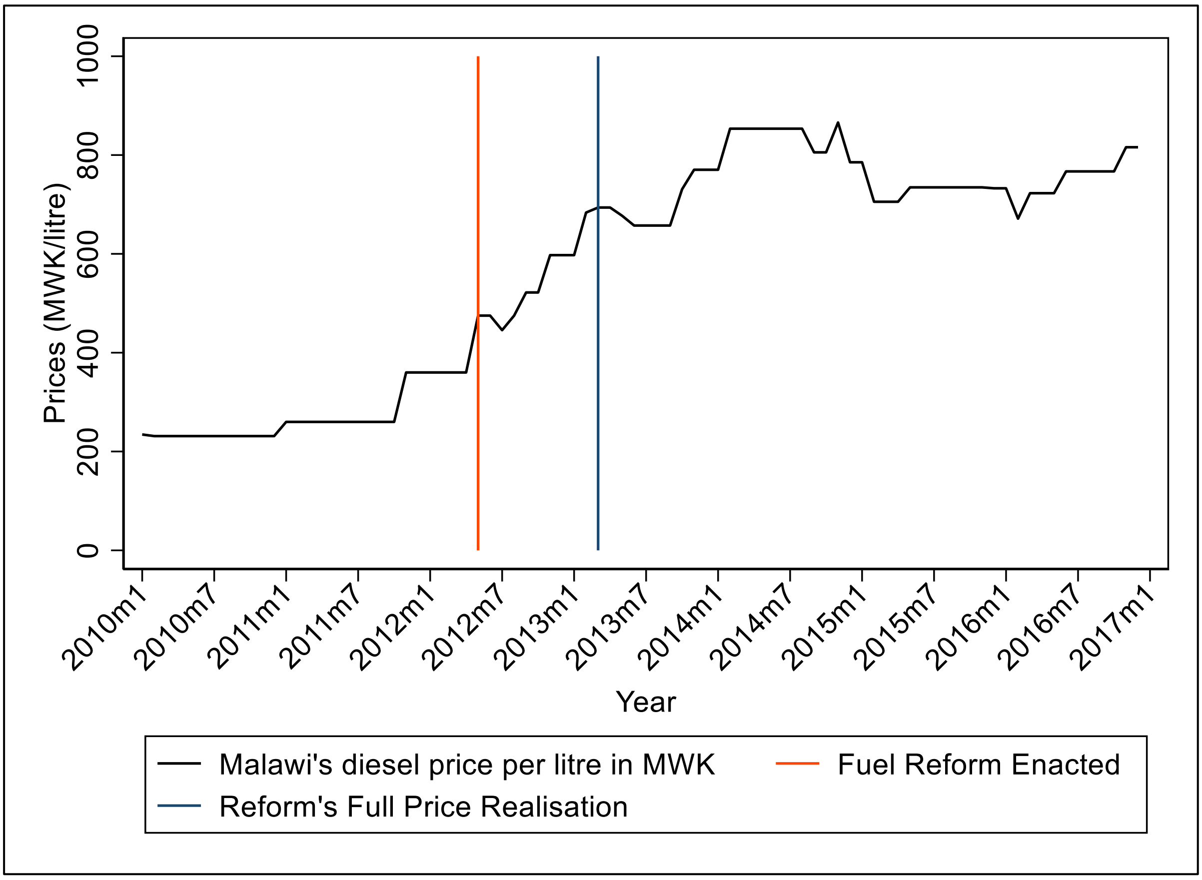

Figure 3.1: Variation in fuel prices between 2010 and 2017

3.4 Related literature

This section discusses the empirical method that we adopt in this study to examine distributional effects of fuel policy reforms on household welfare. Further, the section reviews previous studies that use transport costs as a measure of market access when examining how market access relates to various measures of household welfare indicators in developing countries.

3.4.1 Our empirical approach

We use a panel data econometric approach to estimate distributional effects of the fuel price reform on household welfare taking advantage of available household level panel data for Malawi before and after the policy change. The standard procedure in policy analysis is to estimate the average difference in outcome between households affected by the policy change (*i.e.*, treatment group) and households unaffected by the policy change (*i.e.*, control group) using a standard difference-in-differences (DiD) framework. However, national policy reforms such as the removal of fuel subsidies affect almost all households in the economy either directly through consumption of fuel products or indirectly through consumption of goods and services that use fuel as input such as farm inputs and marketing costs. As a result, it may be hard to get a pure control group.

In the absence of the pure control group, we follow the literature that exploits variation in the intensity of exposure to the event either geographically or in some other measure to estimate the treatment effect of the policy change (Sun & Shapiro, 2022). The regression equation involves household fixed effects, time fixed effects, and an interaction between event variable (*i.e.*, treatment time) and exposure variable before the policy change. Sun & Shapiro (2022) call this equation an exposure model. The interaction term captures the average difference in outcome due to differences in exposure to the event because of the policy change (see, Sun & Shapiro, 2022). The identification assumption of the exposure model is that pre-existing differences before the policy reform would persist on the same trends in the absence of the policy reform (Finkelstein, 2007; Sun & Shapiro, 2022). Finkelstein (2007) uses the exposure model to estimate effects of the introduction of Medicare health insurance on hospital expenditure in the US. Further, Dube & Vargas (2013) use the exposure model to estimate how income induced shocks by changes in international oil prices affected civil war in

Colombia, and Pierce & Schott (2016) apply the exposure model to examine the link between a sudden decline in US manufacturing employment after 2000 to trade policy reform that removed tariff increases on goods imported from China.

In accordance with the previous literature, we anticipate that there will be heterogeneous differential effects in the impact of the fuel price reform on households' welfare that vary with household status as a net-seller, a net-buyer or self-sufficient in staple maize grains, and proximity to consumption centres. Thus, a combination of household status and distance to the consumption centre before the policy change will be our measure of exposure to the fuel policy reform. Therefore, the exposure model allows us to measure distributional consequences of the fuel policy reform on net sellers and net buyers relative to self-sufficient households of staple maize grain that varies with distance to consumption centre.

3.4.2 Measures of transport costs at the household level

There are only a few studies that explicitly investigate the effects of transport costs on households in developing countries. This section reviews previous studies that use various measures of transport costs to investigate the effects of transport costs on crop production, marketing of farm output, and household welfare in developing countries using cross-sectional data.

3.4.2.1 Opportunity cost of travel time for a round trip

The opportunity cost of a round trip to the market is one of the proxy variables for transport costs that has been used at the household level in this literature. To our knowledge, Omamo (1998) is the only study that uses the opportunity cost of a round trip to purchase maize or sell cotton to the market to examine the effects of transport costs on cropping patterns among households in the Siaya District in Kenya. The author calculated the opportunity cost of a round trip as the travel time multiplied by daily wage to show that an increase in transport costs to the market is associated with an increase in local production of maize among net buyers, but a decrease in local production of maize among net sellers. Further, the author shows that an increase in transport costs is associated with the reduction in revenue from cotton production. Thus, an increase in transport costs is associated with the increase in production of food for own consumption among net buyers, but it is associated with the reduction in the supply of food on the market among net sellers leading to income losses. This study provides important insights on how the effects of transport costs at the household level vary with household market position (*i.e*., as a net-seller, a net-buyer or self-sufficient in staple maize grain).

3.4.2.2 Opportunity cost of travel time complemented with the cost of transporting a load

There are other studies that build on Omamo (1998) to investigate the effects of transport costs at the household level. These studies complement the opportunity cost of travel time with the cost of hiring a donkey for a round-trip to carry a 100 kg load to a community market in a geographic area. For instance, Minten et al. (2013) use a cost of hiring a donkey for a round-trip to carry a 100 kg load to a community market in Atsedemariam town to show that the increase in transaction and transport costs is associated with an increase in the price farmers pay for inputs, but it is associated with a reduction in profitability of fertiliser use, and the probability of using modern inputs and their intensity in the most remote areas using a unique cross-sectional dataset in Ethiopia. In a related study, Stifel & Minten (2017) use the same dataset and the measure of transport costs as in Minten et al. (2013) to show that the increase in transport costs of getting to the market in more remote areas is associated with lower household consumption per capita, food insecurity, less diversified diets, and with fewer children enrolled in schools than households that resided closer to the market. While these studies focus on the same market that serves the community as the only systematic difference in the area, their findings and conclusions cannot be applied or generalised to other areas without additional research. Further, these studies simply estimate the overall effect of transport costs on households, which vary with household market position.

3.4.2.3 Lowest cost of transporting a ton of goods across various routes

Usually, households have various routes to get to the market based on the road network in their locations. There are other studies that calculate the lowest travel costs from farmers' location to the market as the measure of transport costs at the household level. These studies use data from various sources such as road surveys, Geographic Information System roads networks, and the Highway Development Management Model (HDM-4) data to identify the lowest cost of transporting a ton of goods from each farmer location to the market. Given that road placement is endogenous, these studies use an instrumental variable approach to correct for endogeneity of road placement. For instance, Damania et al. (2016) show that mechanisation of yams and rice production decreases, whereas traditional production of maize and rice increases with rising transport costs to the market in Nigeria, while Vandercasteelen et al. (2018) show that the increase in transport costs from the farm to the city reduces the price of output that farmers in remote areas receive, use of improved inputs, and agricultural yield in Ethiopia. While the results from these studies are nationally representatives, these studies also estimate the overall effect of transport costs on households, which vary with household market position.

3.4.2.4 Change in the fuel policy

The change in fuel policy is another proxy measure for transport costs that has been used in this literature at the household level. This involves the use of Deaton's framework for analysing budget shares where changes in prices of a good that the household consume are assumed to be proportional to the share of expenditure on that good in total household budgets (A. Deaton, 1989).[35] This approach allows estimation of the effect of the fuel price change. The effect is computed as the budget share of each expenditure on fuel product before the fuel price change (i.e., the ratio of fuel expenditure to total household consumption) multiplied by the percentage change in fuel price (Arze del Granado et al., 2012; Coady et al., 2006; Groot & Oostveen, 2019). To our knowledge, Fuje (2019) is the only study that uses this framework to show that the effects of the subsidy removal on households are mixed for net buyers and net sellers of food (*i.e.*, "teff", wheat, maize, sorghum, and barley) in Ethiopia where most rural households lost in terms of real income because most grain stayed local which lowered grain prices, others gained or did not experience any change in their real income, and fewer households in urban areas had a decrease in real income using nationally representative data. While this approach is non-parametric, it also assumes that households do not adjust their fuel consumption, which overestimates the welfare impact (Coady et al., 2006).

[35] This framework has been widely used to estimate welfare effects of price shocks on households (see, for example, Arndt et al., 2008; Barrett & Dorosh, 1996; Buddy, 1993; Dimova & Gbakou, 2013; Minot & Dewina, 2015).

This chapter builds on these previous studies to examine how increase in transport costs arising from the removal of the fuel subsidy with distance to the market differentially affected staple maize production and consumption among households that are net-sellers, net-buyers or self-sufficient in staple maize grain in Malawi using nationally representative observational panel data via a parametric approach.

3.5 Methods

3.5.1 Theoretical framework

In most developing countries, an increase in fuel price arising from either movement in international oil prices or removal of fuel subsidies largely increases transport costs for both rural and urban population (Dillon & Barrett, 2016; Fuje, 2019, 2020). This increase in transport costs is exogenous to households (Omamo, 1998). Evidence shows that an increase in transport costs, which increase with distance, increases spatial price dispersion of most food commodities across locations (Fuje, 2019) or markets as established in the previous chapter. This means that increasing transport costs reduce inter-market or region trade of staple foods such that prices may remain low in surplus producing areas and high in low producing areas. This suggests that there are differential effects of increasing transport costs on household welfare, which depend on how households engage with food markets (*i.e.*, as a net buyer, a net seller or self-sufficient), whether households incur additional transport costs to access the goods and services (*i.e.*, household proximity to the market), and whether households consume goods and services that use fuel as an input (Fuje, 2019; Goetz, 1992; Kpodar & Djiofack, 2009; Omamo, 1998; Stifel & Minten, 2017). Recall, net sellers are households whose quantity of staple maize grain sold on the market is greater than the quantity of staple maize grain purchased, net buyers are households whose quantity of staple maize grain purchased on the market is greater than the quantity of staple maize grain sold, and self-sufficient households are those that do not purchase or sell any staple maize grain on the market.

To better understand distributional effects of the fuel price reform on maize production and consumption, it is important to identify markets that provide better terms of trade to households and are important for food security. Most food production takes place in rural areas rather than in urban areas. As a result, rural areas are considered as surplus locations while urban areas are considered as deficit locations. It has been established that producer prices are higher in deficit locations (*i.e.*, consumption centres) than in surplus locations (*i.e.*, rural areas) because demand is higher from people who find it costly to get their food more cheaply from smaller rural markets (Alene et al., 2008; Arndt et al., 2008; Benfica, 2014; Fuje, 2019; Goetz, 1992; Key et al., 2000; Minot & Dewina, 2015; Stifel & Minten, 2017).[36] Therefore, local large markets or consumption centres provide better terms of trade to farm households. Farm households have three options to sell their produce, namely local small markets, local large markets, and consumption centres. Local small markets are least active markets that serve the community or village in smaller quantities for both farm produce and consumer goods, but they are not available in every community (Koppmair et al., 2017). Local large markets are most active agricultural markets that the Famine Early Warning Systems Network-Malawi (FEWS NET-Malawi) monitors staple food prices across the country on a regular basis (National Statistical Office, 2013a), while consumption centres are rural towns with a population of greater than 20,000 (Jones, 2017).[37] Consistent with the literature, a key assumption is that locally produced food such as staple maize flow from the farm to local large market or consumption centres, whereas farm inputs such as inorganic fertilisers and pesticides including other food items that households cannot produce themselves flow from consumption centres to local small markets in rural areas (Alene et al., 2008; Jayne, 1994; Key et al., 2000; Stifel & Minten, 2017).

Net sellers of food finance their consumption expenditures with agricultural income and off-farm small-scale enterprises, which suggests that agricultural income is more important for these households. However, the decision of where to sell to maximise revenue (*i.e.*, sell locally at lower producer prices or sell at higher producer prices in the local large market or consumption centre) depends on the cost of transporting the output from the farm to the point of sale or consumption. Transport costs of accessing consumption centres or local large markets increase with distance to the rural areas. Table 3.1 summarises the expected effects

[36] Although consumption or urban areas are better connected to international markets and logistics are geared towards urban markets one might expect grain prices to be lower in urban areas than in rural areas. However, commercial imports of grain are limited in most developing countries including Malawi, except during the period of food crisis (Babu & Chapasuka, 1997; Dana et al., 2007; Derlagen, 2012; Ellis & Manda, 2012; Fuje, 2020). As a result, urban areas depend largely on grain supplies from rural areas, which involves motorised transportation leading to higher grain prices in urban areas than in rural areas.

[37] FEWS NET-Malawi collects food prices for beans, cassava, cowpeas, maize, pigeon peas, and rice in local large agricultural markets to monitor acute food insecurity in the country.

on production and welfare indicators on net buyers and net sellers relative to autarkic households. If the distance to the local large market or consumption centre is small, then with lower transport costs, net sellers may allocate more land for maize production to increase crop revenue from higher producer prices, which may lead to an increase in household consumption and quality of diet. Conversely, if the distance to the local large market or consumption centre is large, then with the higher transport costs it becomes harder for net sellers to sell their output in local large markets or consumption centres. As a result, net sellers in remote locations may allocate less land for maize production, which may reduce their crop revenue, consumption, and quality of diet (Damania et al., 2016; Fuje, 2019; Omamo, 1998; Stifel & Minten, 2017). Even if net sellers decide to sell their output locally to aggregators or small-scale traders that penetrate rural areas, they may still lose out in crop revenue due to lower producer prices that these traders offer because they must also incur additional transport costs to get their goods to the point of consumption (Alene et al., 2008; Key et al., 2000). However, there is a possibility that lower producer prices might spur net sellers in the rural area to produce and sell more maize if revenue falls below subsistence consumption levels. In any case, the increase in transport costs to the consumption centre or local large market would reduce crop revenue, consumption, and quality of diet among net selling households in the rural areas.

Table 3.1: Summary of expected effects of fuel policy shock on production and welfare by household status and market access.

Welfare-related indicators	**Net seller**		**Net buyer**	
	Closer to market	**Remote location**	**Closer to market**	**Remote location**
Production indicator				
Share of land allocated to maize	+	-	+	-
Consumption indicators				
per capita consumption	+	-	-	+
per capita non-food consumption	+	-	-	+
per capita food consumption	+	-	-	+
per capita maize consumption	+	-	-	+
Non-maize food consumption share	-	+	+	-
Maize consumption share	-	+	+	-
Dietary diversity index	+	-	-	+

Note: - represents a negative differential effect, while + represents a positive differential effect. The reference group is autarkic households.

The mechanism through which net buyers of food finance their consumption expenditures differ across urban and rural areas. Usually, urban net buyers solely finance their expenditures through labour supply for cash income because they do not have access to agricultural land. This indicates that agricultural income is less important for these urban net buyers. As a result, these households are unable to adjust their operations to take advantage of increasing producer prices in deficit locations and suffer most from these higher prices (Arndt et al., 2008; Benfica, 2014; Fuje, 2019; Goetz, 1992; Minot & Dewina, 2015). Conversely, rural net buyers partially produce their own food because they have limited access to agricultural land and productivity enhancing technologies such as fertiliser and improved seeds, and are geographically dispersed (Arndt et al., 2008; Burke et al., 2020; Ruel et al., 2010). These rural net buyers finance their consumption expenditures with cash income from agricultural wage labour and some off-farm small-scale enterprises. If the distance to the local large market or consumption centre is small, then net buyers may allocate more land for maize production to avoid higher producer maize prices arising from higher transport costs of getting the maize from the remote locations and consume more from own production (Omamo, 1998). Whether net buyers partially or entirely produce their own maize rely on market purchases, higher producer prices in local large markets or consumption centres would reduce household consumption and quality of diet among net buyers. Conversely, if the distance to the local large market or consumption centre is large, then net buyers may allocate less land for maize production to gain from lower producer prices as more maize stays local because of higher transport costs and consume more from the market. If an increase in transport costs does not affect non-farm earnings (Stifel & Minten, 2017), then net buyers may save in maize expenditures, which may allow them to increase consumption and their quality of diet.

Unlike locally produced food, farm inputs and some food items that farmers cannot produce themselves such as cooking oil, salt, and sugar move from local large markets or consumption centres to local small markets in rural areas. Minten et al. (2013) find that transport costs to the market were positively associated with higher input prices in Ethiopia, while Stifel & Minten (2017) using the same data find that transport costs to the market were positively associated with higher consumer prices for food items that are not locally produced. Similarly, if the distance to the local large market or consumption centre is small, then with lower transport costs, households in consumption centre or local large market would gain from lower prices for farm inputs and other food items. Conversely, if the distance to the local large market or consumption centre is large, then with higher transport costs, households in rural areas may lose out from higher prices for farm inputs and other food items. Thus, the effects of transport costs on prices for farm inputs and other food items vary with distance to local large market or consumption centre, but not with household status because households do not produce these items themselves. Since households use crop revenue to finance their consumption expenditures on farm inputs and other food items that households cannot produce themselves, poor crop revenue reduces ability of both net sellers and net buyers in rural areas to purchase these items.

While autarkic households in both local large markets or consumption centres and rural areas may not be affected by the changes in producer maize prices arising from changes in transport costs between the local large markets or consumption centre and rural areas, we expect the fuel policy reform to have a less of an effect on autarkic households than on net sellers and net buyers of maize staple grain. As a result, it is difficult to predict expected effects on autarkic households.

Ultimately, the extent to which the fuel price reform affect staple maize production and consumption is an empirical question that we describe and estimate in the following section.

3.5.2 Empirical strategy

Our empirical strategy is to examine whether there are differences in the effect of the policy reform on welfare-related outcomes by comparing households to which the policy reform should have a larger effect (*i.e.*, net sellers and net buyers) to households to which it should have a smaller effect (*i.e.*, self-sufficient households). We examine whether there are breaks in pre-existing differences in the level or trend of welfare-related outcomes closer to the time of reform in 2012 and then further away from the reform period. Any break in pre-existing differences in the level or trend of welfare-related outcomes is our estimate of the causal impact of the policy reform (Finkelstein, 2007; Sun & Shapiro, 2022). The identification assumption in this approach is that any pre-existing differences would have persisted on the same trends in the absence of the policy reform (Finkelstein, 2007; Sun & Shapiro, 2022).

Before estimating heterogeneous differential impact of the fuel price reform on welfare outcomes that varies with distance to consumption centre, we first examine whether there are differential effects between household groups. To examine immediate differential effects of fuel subsidy removal on welfare outcomes for household i in community c (*i.e.*, between the first wave in 2010 and second wave in 2013), we specify our estimating equation as follows:

$$W_{ict} = \beta_0 + \beta_1(P_t * B_{ic}) + \beta_2(P_t * S_{ic}) + \beta_3 X_{ict} + \tau_t + \vartheta_i + \varepsilon_{ict} \qquad (3.1)$$

where W_{ict} represents production or welfare-related outcome indicators (maize production and consumption) for household i in community c at time t. We use share of land allocated to maize as an indicator of maize production in each survey wave.[38] Further, we use logarithm of per capita consumption in MWK, logarithm of per capita non-food consumption in MWK, logarithm of per capita non-maize food consumption in MWK, per capita quantity of maize consumption in kilograms, consumption shares on non-maize food and maize, and dietary diversity score as indicators of household consumption in each survey wave (the next section describes in more detail how each indicator was constructed).

P_t is a dummy variable that takes on a value of 1 for the years after the fuel subsidy removal (2013 and 2016), and 0 for the year before the fuel reform (2010) (*i.e.*, reference period). We classified households based on pre-price-rise (*i.e.*, before the fuel price rise) maize market position as a net buyer, net seller, or autarkic (*i.e.*, self-sufficient) household in 2010 (the next section describes in more detail how household groups were constructed). B_{ic} is a dummy variable equal to one for households that are net buyers of staple maize grain in 2010 and zero otherwise. S_{ic} is a dummy variable that captures households that are net sellers of staple maize grain in 2010 and zero otherwise. In this specification, households that are in autarky in 2010 are our reference group. τ_t represents year dummies (*i.e.*, time-specific fixed effects) that capture common shocks to households such as economic conditions and weather, whereas ϑ_i represents household-specific fixed effects.[39] Since P_t only varies across time while S_{ic} and B_{ic} only vary across households, they will be picked up by the time dummies

[38] We also examine the effect of the fuel policy reform on maize cultivation at the extensive margin, where we use a dummy variable equal to one if all the land is allocated to maize cultivation, and zero otherwise.

[39] Note that the time dummy τ_t is effectively the same as the post-reform dummy, P_t, if we always only use two periods, one before and one after the reform. Hence, we could rewrite equation [3.1] as follows:

$$W_{ict} = \beta_0 + \beta_1(P_t * B_{ic}) + \beta_2(P_t * S_{ic}) + \beta_3 X_{ict} + \beta_4 P_t + \vartheta_i + \varepsilon_{ict}$$

where P_t could pick up the effect of the reform for households that are in autarky. However, P_t does not capture any causal effect of the reform on welfare related outcomes. Instead, it captures common shocks to households such as economic conditions and weather in the post versus pre period other than the fuel price reform itself. As a result, we cannot plausibly estimate the absolute effect of the reform for households that are in autarky. Instead, we focus on relative effects, β_1 and β_2.

and household-specific fixed effects, respectively. X_{ict} is a vector of time varying covariates that affect welfare-related outcomes such as household head's age, gender and education, household size, and value of assets (see table B.1 in the appendix for details). We also controlled for seasonality in consumption using monthly dummies in the specification for consumption-related indicators.[40] The parameter β_1 provides an estimate of the immediate differential effect of the policy reform on households that are net buyers relative to households that are in autarky, while the parameter β_2 captures an estimate of the immediate differential effect of the policy reform on households that are net sellers relative to households that are in autarky. The underlying assumption for identification of a causal impact of the reform, β_1 and β_2, is that everything else that changes welfare-related outcomes in the post versus pre reform period is the same for households that are net buyers and net sellers as for households that are in autarky. ε_{ict} represents a household random error term, and β_0, β_1 through to β_3 are additional parameters to estimate.[41] To examine whether there are persistent differential effects of fuel subsidy removal between household groups on welfare outcomes for household i in community c, equation [3.1] is re-estimated between the first wave in 2010 and third wave in 2016. Since predicted differential effects depend on distance as well, it is tricky to get clearer predictions in equation [3.1].

Then, we interact our policy variable and household status variables with distance to allow for heterogeneity of immediate differential impacts of the reform with geographic location in our baseline specification (*i.e.*, equation [3.1]) as follows:

$$W_{ict} = \gamma_0 + \gamma_1(P_t * B_{ic}) + \gamma_2(P_t * S_{ic}) + \gamma_3(P_t * Dist_{ic}) + \gamma_4(P_t * Dist_{ic} * B_{ic}) + \gamma_5(P_t * Dist_{ic} * S_{ic}) + \gamma_6 X_{ict} + \tau_t + \vartheta_i + \mu_{ict} \qquad (3.2)$$

where $Dist_{ic}$ is a household-specific distance to a market (*i.e.*, our measure of market access), which is measured as a minimum distance to the closest local large agricultural market or consumption centre in kilometres as the crow flies. Similarly, $Dist_{ic}$ along with its interaction with household status (*i.e.*, $(Dist_{ic} * B_{ic})$ and $(Dist_{ic} * S_{ic})$) will be picked up by the household fixed effects since they both do not vary over time. In equation [3.2], μ_{ict} represents

[40] We did not control for seasonality in maize production because most of the maize production is under rain-fed agriculture from November through to April (Government of Malawi, 2016a).

[41] Equation [3.1] is a difference-in-differences specification with common treatment timing.

household random error term, and γ_0, γ_1 through to γ_6 are parameters to estimate.[42] The parameter γ_3 provides an estimate of the immediate differential effect of the policy reform for households that are in autarky, the reference group, closer to the market relative to those in remote areas after the policy reform in 2010. Here, the assumption needed to get a causal effect of this parameter γ_3 is that other things (other than the fuel price reform) also change between the pre and post period in the same way for households close to or far away from the market; so that the changes in welfare-related outcomes driven by things other than the reform (economic conditions and weather) are not correlated with distance in any way. The parameter γ_4 captures the immediate effect of distance on the differential effect of the policy reform for households that are net buyers relative to households that are in autarky, whilst γ_5 measures the immediate effect of distance on the differential effect of the policy reform for households that are net sellers relative to households that are in autarky. The underlying assumption for identification of a causal impact of distance on the effect of the reform, γ_4 and γ_5, is that everything else that changes welfare-related outcomes in the post versus pre reform period varies by distance in the same way for households that are net buyers and net sellers as for households that are in autarky. A combination of $\gamma_1 + \gamma_4 Dist_{ic}$ captures the immediate differential effect of the policy reform on households that are net buyers relative to households that are in autarky that varies with distance, while $\gamma_2 + \gamma_5 Dist_{ic}$ measures the immediate differential effect of the policy on households that are net sellers relative to households that are in autarky that varies with distance.[43] $\gamma_1 + \gamma_4 Dist_{ic}$ and $\gamma_2 + \gamma_5 Dist_{ic}$ are our parameters of interest in this specification because they capture differential effects of the fuel price subsidy removal that varies with distance for households that are net buyers and net sellers relative to households that are in autarky. We plot how the differential effects of the policy reform on households vary with distance.

[42] Similarly, we can replace the time dummy τ_t with the post-reform dummy P_t since they are the same in equation [3.2].

[43] Note that any "total" effect of the fuel price reform for any group would depend on knowing how welfare-related outcome has changed over time for that group. The combination of $\gamma_1 + \gamma_3 Dist_{ic} + \gamma_4 Dist_{ic}$ could capture total effect of the policy for households that are net buyers, while $\gamma_2 + \gamma_3 Dist_{ic} + \gamma_5 Dist_{ic}$ could capture total effect of the policy for households that are net sellers relative to the period before the reform. Here, these combinations could include P_t if we could replace the time dummy τ_t with the post-reform dummy P_t. Therefore, we cannot plausibly separate these effects from any other changes over time in welfare-related outcome due to economic conditions and weather in a case with only two periods, one before and one after the reform. For this reason, we only focus on relative effects.

In accordance with Omamo (1998) and Fuje (2019), we expect the policy reform to increase amount of land allocated to maize production among households that are net sellers and net buyers that reside closer to the market relative to households that are in autarky due to higher producer maize prices arising from higher transport costs of getting the maize from the remote locations (*i.e.,* $\gamma_1 + \gamma_4 Dist_{ic} > 0$ and $\gamma_2 + \gamma_5 Dist_{ic} > 0$ when distance is small). Conversely, we expect the policy reform to reduce amount of land allocated to maize production among households that are net sellers and net buyers that reside in remote areas relative to households that are in autarky due to higher transport costs of accessing markets and lower producer maize prices (*i.e.,* $\gamma_1 + \gamma_4 Dist_{ic} < 0$ and $\gamma_2 + \gamma_5 Dist_{ic} < 0$ when distance is large). However, there is a possibility that households that are net sellers in remote locations might also allocate more land to maize cultivation relative to households that are in autarky to increase crop income if it falls below subsistence consumption levels. Regarding consumption, we anticipate the policy reform to reduce household consumption among households that are net buyers closer to the market (*i.e.*, lose out the most given that it gets harder to get maize grain from the remote locations, which increases producer prices) (*i.e.*, $\gamma_1 + \gamma_4 Dist_{ic} < 0$ when distance is small) and to increase household consumption among households that are net sellers closer to the market (*i.e.*, gain in crop revenue arising from higher producer prices and lower transport costs of accessing markets) relative to households that are self-sufficient (*i.e.*, $\gamma_2 + \gamma_5 Dist_{ic} > 0$ when distance is small). Conversely, we anticipate differential effect to be positive in remote locations (*i.e.*, $\gamma_1 + \gamma_4 Dist_{ic} > 0$ when distance is large) among households that are net buyers (*i.e.*, become better off or save in consumption given that more locally produced maize grain stays local, which reduces producer prices) and differential effect to be negative (*i.e.*, $\gamma_2 + \gamma_5 Dist_{ic} < 0$ when distance is large) among households that are net sellers in remote locations (*i.e.*, lose out more in crop revenue arising from lower producer prices) relative to households that are self-sufficient. The rest of the variables are the same as those in equation [3.1]. Similarly, to examine persistent differential effects of fuel subsidy removal that varies with distance on welfare outcomes for household i in community c, equation [3.2] is re-estimated between the first wave in 2010 and third wave in 2016.

Functional form

The next step is to determine the functional form and find an estimator for equation [3.1] where welfare-related outcomes, dependent variables, take different range of values. The share of land allocated for maize production, consumption shares of food and maize are substantively restricted between a zero and a one, whereas per capita quantity of maize consumption is piled up at zero. Further, per capita consumption, per capita non-food consumption, and per capita non-maize food consumption take continuous values, whereas the dietary diversity score is a count variable. In equation [3.1], we need to include the time-specific fixed effects, τ_t, and household-specific fixed effects, ϑ_i, for the differential impact of the policy reform on welfare-related outcomes among households that are net buyers and net sellers relative to households that are self-sufficient to be identified. Therefore, we use a linear fixed effects estimator to examine the impact of the policy reform on maize production and consumption outcomes that take continuous values and are restricted between a zero and a one.

For welfare-related outcomes that are piled at zero values, we treat the zero values in our data as observed values and not missing or unobserved observations. In our context, a corner solution model or a model for count data is more plausible than a Heckman selection model, which deals with incidental truncation when zeros are missing or unobserved values. Therefore, equation [3.1] can be estimated via either the tobit estimator (Tobin, 1958) and the more flexible double hurdle model (Cragg, 1971) for the corner solution model or Poisson regression estimator for the count data. We use the fixed effects Poisson estimator to examine the impact of the policy reform on welfare-related outcomes that are piled at zero values. The fixed effects Poisson estimator has attractive features in that it simply requires the conditional mean function of the dependent variable to be correctly specified for consistency, allows for arbitrarily dependence between unobserved effect and covariates, and the dependent variable does not need to be a count variable (Correia et al., 2019; Gourieroux et al., 1984; Hausman et al., 1984; Wooldridge, 1999, 2010). Thus, the dependent variable can be continuous or a corner solution (*i.e.*, specification of the distribution assumption of the dependent variable is not required or is unrestricted). When the conditional mean function of the dependent variable is correctly specified, Poisson regression comes to be the Poisson pseudo maximum likelihood (PPML) regression. Further, observations of the dependent variable that are equal to zero are naturally dealt with and do not cause the sample selection problem (Correia et al., 2019; Santos Silva & Tenreyro, 2006; Wooldridge, 1999, 2010). In addition, we use the fixed effects Poisson estimator to examine the impact of the policy reform on dietary diversity score, which is a count variable. We implement the fixed effects Poisson estimator (*i.e.*, PPML) using Stata's command for estimating (pseudo) Poisson regression models with multiple high-dimensional fixed effects (*i.e.*, PPMLHDFE) because it converges much faster in the presence of fixed effects and estimates are consistent in the presence of heteroskedasticity (Correia et al., 2019; Santos Silva & Tenreyro, 2006).

Robustness

As robustness checks, we re-estimate equation [3.2] with distance from the household to the district capital as an alternative measures of market access. We expect the direction of the differential effect of higher fuel costs to remain the same, but only differ in magnitude to estimates in our main specifications.

3.5.3 Data

3.5.3.1 Household data

We use panel data from the three waves of the nationally representative Integrated Household Panel Survey (IHPS) implemented in 2010, 2013, and 2016 as part of the Living Standards Measurement Study-Integrated Surveys of Agriculture (LSMS-ISA) for Malawi. The IHPS tracks a stratified two-stage random sub-sample of 3246 households that were part of the third Integrated Household Survey (IHS3) in 2010/2011 across 204 enumeration areas (EAs) (*i.e.*, communities) in 27 districts and 4 cities (*i.e.*, Lilongwe, Blantyre, Mzuzu, and Zomba Municipality). This indicates that the community has about 16 households (=3246/204) in the dataset, on average. IHS3 cross-sectional full survey was conducted from March 2010 through to March 2011 across 768 communities from a sample of 12271 (National Statistical Office, 2014a). In panel surveys, individuals that branched off from the original household and formed a new household were included into the IHPS sample, which indicates that the sample increases over time (National Statistical Office, 2014b). In 2012/13 panel survey, about 4000 households were successfully tracked of which 896 households (*i.e.*, 23.2 percent) branched off into 2 or more households whereas the remaining 3104 baseline households did not branch off (National Statistical Office, 2014b). In subsequent panel surveys after 2013, the number of communities to be tracked was reduced from 204 to 102 EAs due to financial and resource constraints (National Statistical Office of Malawi, 2017). In 2015/16 panel survey, about 2508 households were successfully tracked of which 984 households (*i.e.*, 39.2 percent) branched off into 2 or more households whereas the remaining 1524 baseline households did not branch off since the 2010 survey (National Statistical Office of Malawi, 2017).

The fieldwork for the 2012/13 survey took place from April to December 2013 (*i.e.*, a month after full realisation of the increase in fuel prices following the enactment of the fuel policy reform) (National Statistical Office, 2014a), while for the 2015/16 survey fieldwork took place from April 2016 to April 2017 (National Statistical Office, 2017). Thus, the 2012/13 and 2015/16 survey waves were conducted one year and four years after the policy reform, respectively. In this study, we exclude splitting-off households from the baseline households that went on to form new households after the policy reform. Therefore, our initial sample is an unbalanced panel with 7874 observations from the three survey waves. The panel survey has a stronger focus on household demographics, food consumption, education, health, and agriculture. The LSMS-ISA survey asks the respondent who is more knowledgeable about the food consumed in the household to recall and list all foods consumed during the previous seven days including their quantities and sources. The survey collects information on 135 variety of food items that households consumed in each wave. The food consumed is sourced from own production, market purchase, or received as gifts and other sources.

Classification of households

Values on food consumed and produced were collected using non-standard measurement units such as a plate or a bucket of maize, where the plate or the bucket has different sizes across geographical locations (World Bank, 2010). To obtain equivalent measures in kilograms, we use conversion factors that are published along with the data to estimate household food consumption by source and agricultural production by crop. Then, we use the quantity of staple maize consumed from the market in the household consumption module and the quantity of staple maize sold on the market using information in the household agriculture module to classify households into net buyers, net sellers, and self-sufficient at baseline (Aksoy & Isik-Dikmelik, 2008; World Food Programme, 2009). Recall, net buyers are households whose quantity of maize grain purchased is greater than the quantity sold on the market, net sellers are households whose quantity of maize grain sold is greater than the quantity purchased on the market, and self-sufficient households are those that do not purchase or sell maize grain on the market. The difference between the quantity of maize grain sold and purchased is net quantity of maize grain sold, which is positive among net sellers, negative among net buyers, and zero among autarkic households. Then, we create a categorical variable that takes on a value of zero for households with zero net quantity of maize grain sold (*i.e.*, self-sufficient households), a value of one for households with negative net quantity of maize grain sold (*i.e.*, net buyers), and a value of two for households with positive net quantity of maize grain sold (*i.e.*, net sellers). Using this procedure, we classified 337 households as net sellers at baseline (*i.e.*, 10%), 1236 households as net buyers at baseline (*i.e.*, 38%), and 1673 as self-sufficient households at baseline (*i.e.*, 52%) with respect to staple maize grain.[44,45,46] This shows that our sample has fewer net buyers at baseline and more autarkic households at baseline of staple maize grain than the previous studies (Dorward et al., 2008; Sibande et al., 2017).[47] We adopt this classification of households based on maize because it remains the most produced and consumed food crop compared to rice, millet, and sorghum in Malawi (Benson & Weerdt, 2023; Minot, 2010; National Statistical Office, 2014a; Pauw et al., 2015).

Construction of maize production indicator

We use share of land allocated to maize crop as the measure of household agricultural production. The area of land allocated to maize was collected in various units namely, hectares, acres, and square meters. We converted all the land that were measured in acres and square meters to hectares. Then, we construct the share of land allocated to maize crop as the ratio

[44] About 857 households (*i.e.*, 26 percent) are urban dwellers, while 2389 households (*i.e.*, 74 percent) are rural dwellers. Further, about 746 households (*i.e.*, 23 percent) are non-farming households, while 2500 households (*i.e.*, 77 percent) are farming households.

[45] All self-sufficient households are maize producing and consuming households. We have 33 households *(i.e.*, 1% of the baseline sample) that neither produce nor consume maize, which we have classified as net buyers of maize. Our findings are robust to exclusion of these households from the analysis.

[46] Note that households might change their market position over time. We find that about 39% of the households changed their market position between 2010 and 2013, while about 47% changed their market position between 2013 and 2016.

[47] Sibande et al. (2017) is the closest paper to ours to classify households into maize market position using IHPS data for Malawi. Our findings on maize market position remain the same when we pool the IHPS data to replicate Sibande et al. (2017) paper. However, we find similar results to theirs when we classify households based on all staple grains in the data (*i.e.*, maize, rice, millet, and sorghum) where 10 percent are net sellers, 52 percent are net buyers, and 38 percent are autarkic households with respect to staple grains.

of land allocated to maize crop to total land area (Chibwana et al., 2012; Omamo, 1998; Salazar-Espinoza et al., 2015).[48] The share ranges from zero to one. The share of zero indicates that households do not cultivate maize crop, while the share closer to one indicates a greater amount of land allocated to maize crop. To examine the differential effect of the fuel policy reform on maize cultivation at the extensive margin, we create a dummy variable that takes on a value of one if all the land is allocated for maize cultivation, and zero otherwise.

Construction of consumption indicators

Consumption indicators are comprised as follows: annual per capita consumption (including non-food and non-maize food); weekly quantity of per capita maize consumption); annual food and maize consumption shares; and weekly dietary diversity index (DD). Typical of household surveys, the recall period for expenditure and consumption data varies from a week (food items) to a year (durable assets). We impute the monetary value of non-purchased consumed items (*i.e.*, consumption from own production and in-kind transfers) using median consumer prices in each geographic location to derive expenditure equivalents (International Dietary Data Expansion Project, 2018; Schneider et al., 2023; Stifel & Minten, 2017; World Food Programme, 2009). Thus, non-maize food consumption is the monetary value of purchased and non-purchased food (*i.e.*, food consumed from own production and in-kind transfers). Then, we sum all expenditure equivalent values to get their corresponding annual values.[49] We use these annual values to compute total annual household equivalent expenditures, which comprises expenditures on durable assets, food, consumables, housing, utilities, health, education, and agriculture.[50] Expenditures on durable assets and housing are approximated annual use values (Adams & Cuecuecha, 2010; Schneider et al., 2023).

We divide total annual household expenditures by household adult equivalent scales to obtain annual per capita consumption for each household as illustrated in Smith & Subandoro (2007).[51] This is a four-step procedure that involves (i) categorising household members into

[48] Note that we retain households that do not cultivate any crop when constructing the shares. These households have a value of zero on crop shares.
[49] Expenditures are adjusted for inflation using consumer price index (CPI) from the Reserve Bank of Malawi.
[50] It is important to note that expenditure and consumption are used interchangeably in consumption literature.
[51] While consumption data is collected at the household level, we are interested in comparing the standard of living of individuals in different households. Therefore, accounting for differences in household size and composition allows us to compare the standard of living of individuals in different households (Deaton & Zaidi, 2002). For this reason, we use per capita consumption and not total household consumption indicator.

age by gender (*i.e.*, age-sex category), (ii) allocating an adult equivalent factor to each category based on energy requirements for moderate activity (2,900 kilocalories) relative to a male adult aged between 30 and 60 years old, (iii) multiplying the adult equivalent factor by the number of household members in each category, and (iv) add the number of adult equivalents from each category to get the household adult equivalent scale. We use this procedure to get annual per capita value of non-food and non-maize food consumption, and weekly quantity of per capita maize consumption.[52]

Then, we divide annual household food consumption by total annual household expenditures to get food consumption shares or consumption ratios (Adams & Cuecuecha, 2010; A. Deaton, 1989; Fuje, 2019; International Dietary Data Expansion Project, 2018; Schneider et al., 2023; Usman & Haile, 2022). Similarly, we construct maize consumption shares by dividing annual household maize consumption by total annual household expenditures. The shares range from zero to one. The value of the share closer to zero indicates lower risk to food insecurity, while the value closer to one signifies greater risk to food insecurity (Smith & Subandoro, 2007). We drop those observations with missing or unreasonable zero values of food consumption (*i.e.*, households that reported not to have consumed any food during the interviews) as part of data cleaning (Hasan, 2016; Olabisi et al., 2021). Our final sample is unbalanced panel data with 7872 observations from the three survey waves.

Our measure of dietary diversity is the sum of food groups consumed during the past seven days (Coates et al., 2007; Headey et al., 2019; Hirvonen et al., 2017; Olabisi et al., 2021; Swindale, 2005; Swindale & Bilinsky, 2006). We classified the 135 food items into 12 food groups, namely, (i) cereals & cereal products, (ii) roots & tubers (iii) legumes, nuts, and seeds (iv) eggs, (v) milk and milk products, (vi) stimulants, spices & condiment, (vii) meat, (viii) fish and seafood, (ix) sweets & confectionary, (x) vegetables, (xi) fruits., and (xii) oils & fats (Coates et al., 2007; Hoddinott & Yohannes, 2002; Swindale & Bilinsky, 2006).[53] The household gets a one if they consume food items from the food group and zero otherwise (*i.e.*, each food group consumed is defined as a binary indicator). The score closer to 12 means that the household

[52]Maize consumption is quantity of maize consumed from the market or other sources (*i.e.*, maize consumed from own production and in-kind transfers).
[53] Aggregation to 12 food groups that excludes consumption of beverages is widely used in this literature because it reflects the quality of diet in Africa thereby allowing comparison of food consumed across space and time (Food and Agriculture Organization of the United Nations, 1968; Swindale & Bilinsky, 2006).

consumes most of food items from the 12 food groups, while the score closer to 0 means that the household consumes fewer food items from the 12 food groups.

Construction of market access indicator

Malawi NSO also publishes Euclidean distance (*i.e.*, as the crow flies) in kilometres from the household to the nearest town (*i.e.*, consumption centre) with a population of greater than 20 000 and Euclidean distance from the household to the nearest local large agricultural market (National Statistical Office, 2013a). We select the closest distance from the household to the local large agricultural market or consumption centre as a measure of market access of each household in the community. We also use Euclidean distance from the household to the district capital, which are published along with the data as an alternative measures of market access.

Distribution of the data

Figure B.2 in the appendix provides the distribution of production and consumption indicators. The distribution of the data suggests presence of outliers in some dependent variables. While some functional forms are less sensitive to potential outlier observations such as constant elasticity models where logarithmic transformation narrows the range of the data (Wooldridge, 2020), we drop the top 5 percent of values of per capita consumption indicators (*i.e.*, per capita annual, non-food, non-maize food, and quantity of maize consumption) based on their distribution. Values of non-maize food expenditure shares closer to zero or equal to one seem unreasonable, hence, we drop the top and bottom 1 percent. Further, we drop the top 1 percent of values of maize consumption share and the bottom 1 percent of dietary diversity index (DD).[54]

3.5.3.2 Descriptive statistics

Table 3.2 presents summary statistics of outcome variables used in the analysis for each category of household status at baseline (table B.2 in the appendix provides summary statistics of outcome variables by wave). We use a multivariate test of means procedure that Moore (1998) developed to determine whether the means are different across household

[54] Overall, our results are not sensitive to the presence of outliers as discussed in the next section.

groups. Focusing initially upon the alternative dependent variables adopted in the analysis, table 3.2 reveals that all households allocate a large share of their land to maize production, however, land allocated to maize is relatively larger among households that are self-sufficient at baseline (81%) than among households that are net sellers (76%) and net buyers (35%) at baseline, on average. This means households that are net sellers and self-sufficient at baseline cultivate more maize than households that are net buyers at baseline, on average. On average, per capita consumption among households that are net buyers at baseline (MWK 144,106) is larger than among households that are net sellers (MWK 143,965) and self-sufficient (MWK 135,595) at baseline, on average. However, the difference in the mean is not significant at the 5% level. Thus, the standard of living is not different across the household groups at baseline. Similarly, per capita non-food consumption is larger among households that are net buyers at baseline (MWK 59,610) than among households that are net sellers (MWK 47,502) and self-sufficient (MWK 44,854) at baseline, on average. Further, per capita non-maize food consumption is larger among households that are net buyers at baseline (MWK 86,903) than among households that are self-sufficient (MWK 63,625) and net sellers (MWK 58,912) at baseline, on average. This indicates that households that are net buyers at baseline spend more per person on non-maize food than households that are self-sufficient and net sellers at baseline. Per capita quantity of maize consumption is larger among households that are net sellers at baseline (3.96 kgs) than among households that are self-sufficient (3.82 kgs) and net buyers (3.42 kgs) at baseline, on average. This indicates that households that are net sellers at baseline consume more maize per person than households that are self-sufficient and net buyers at baseline.

Table 3.2: Summary statistics of outcome indicators by household category at baseline

Variables	Net sellers	Net buyers	Self-sufficient	Mean Diff. (Wald chi (2) statistic)
Share of land allocated to maize	0.764	0.352	0.807	1058.32***
	(0.234)	(0.446)	(0.246)	
Annualized per capita consumption in MWK [a]	140,106	143,965	135,595	5.95*
	(74,257)	(95,231)	(80,944)	
Annualized per capita non-food consumption in MWK [a]	47,502	59,610	44,854	76.06***
	(36,130)	(48,318)	(36,328)	
Annualized per capita non-maize food consumption in MWK [a]	58,912	86,903	63,625	8.05**
	(49,542)	(182,918)	(76,242)	
Weekly quantity of maize consumed (kg) [a]	3.964	3.420	3.819	31.94***
	(2.126)	(2.089)	(1.946)	
Annualized share of non-maize food consumption to total consumption [b]	0.385	0.420	0.393	25.35***
	(0.153)	(0.156)	(0.154)	
Annualized share of maize consumption to total consumption [c]	0.264	0.136	0.261	664.61***
	(0.152)	(0.118)	(0.158)	
HH dietary diversity index [d]	8.478	8.971	8.280	66.10***
	(1.956)	(2.331)	(2.180)	
Observations	337	1235	1673	

Note: Numbers shown are averages and their corresponding standard deviations are presented in parenthesis. [a] trimmed at the top 5%, [b] trimmed at the top and bottom 1%, [c] trimmed at the top 1%, and [d] trimmed at the bottom 1%. * $p<0.10$, ** $p<0.05$, *** $p<0.010$

Non-maize food consumption to total household expenditure is larger among households that are net buyers (42%) than among households that are self-sufficient and net sellers at baseline (39%), on average. This suggests that households that are net buyers at baseline spend a larger share of their income on non-maize food (*i.e.,* monetary value of food purchased, consumed from own production and in-kind transfers) than households that are net sellers and self-sufficient at baseline. Given that non-maize food consumption to total household expenditure shares are less than 50 percent, the risk to non-maize food insecurity for these households is considered to be low (Smith & Subandoro, 2007). Similarly, maize consumption (*i.e.,* monetary value of food consumed from own production and in-kind transfers) to total household expenditure is larger among households that are net seller and self-sufficient at baseline (26%) than among households that are net buyers at baseline (14%), on average. Households that are net buyers at baseline have a better score of dietary diversity (9.0) compared to households that are net sellers (8.5) and self-sufficient (8.3) at baseline, on average. This suggests that households that are net buyers at baseline consume most foods from the 12 food groups compared to households that are net sellers and self-sufficient at baseline.

Turning to the independent variables used in the analysis, socio-economic characteristics differ across household groups at baseline (table B.3 in the appendix provides summary statistics of household characteristics by wave). Table 3.3 shows that households that are self-sufficient at baseline have more both male and female elders greater than 65 years old than households that are net sellers and net buyers at baseline, on average. However, the number of both male and female adults aged between 12 and 65 years old is not significantly different across household groups at 95% confidence interval, on average. This may suggest that endowment of family labour is not different across household groups. However, households that are net sellers and self-sufficient at baseline have more children less than 12 years old than households that are net buyers at baseline, on average. Over 70 percent of the household heads have a spouse, are males and young adults, while over 55 percent of the heads do not have a formal qualification, on average. Further, most of the household heads among households that are net buyers at baseline have attended secondary education (26%) and tertiary education (7%) compared to households that are net sellers and self-sufficient at baseline, on average. Surprisingly, a higher proportion of households that are net sellers at baseline (45%) participate in wage labour compared to households that are self-sufficient (39%) and net buyers at baseline (36%), on average. This may suggest that most households that are net sellers at baseline depend on wage labour to complement their crop income compared to households that are net buyers and self-sufficient at baseline.

Table 3.3: Characteristics by household groups at baseline

Variables	Net sellers	Net buyers	Self-sufficient	Mean Diff. (Wald chi (2) statistic)
# of males greater than 65 years old	0.062	0.056	0.093	15.37***
	(0.242)	(0.233)	(0.291)	
# of male adults 12-64 years old	1.415	1.364	1.338	1.64
	(1.058)	(1.030)	(1.024)	
# of females greater than 65 years old	0.101	0.048	0.120	54.05***
	(0.302)	(0.213)	(0.329)	
# of female adults 12-64 years old	1.407	1.364	1.445	5.47*
	(0.937)	(0.920)	(0.949)	
# of children less than 12 years old	2.027	1.728	1.947	20.06***
	(1.440)	(1.464)	(1.489)	
Marital status of head, =1 if has a spouse	0.804	0.713	0.747	13.65***
	(0.397)	(0.453)	(0.435)	
Gender of household head, =1 if male	0.807	0.792	0.754	8.22**
	(0.395)	(0.406)	(0.431)	
Age of head in years	41.543	38.579	44.507	103.84***
	(15.536)	(14.602)	(16.658)	
Highest qualification of household head, none	0.706	0.572	0.732	81.86***
	(0.456)	(0.495)	(0.443)	
Highest qualification of household head, primary	0.131	0.098	0.108	2.77
	(0.337)	(0.297)	(0.311)	
Highest qualification of household head, secondary	0.154	0.255	0.134	66.42***
	(0.362)	(0.436)	(0.341)	
Highest qualification of household head, tertiary	0.009	0.074	0.026	52.77***
	(0.094)	(0.263)	(0.160)	
If household member participates in wage labour	0.454	0.359	0.385	10.00***
	(0.499)	(0.480)	(0.487)	
If household owns a phone	0.359	0.586	0.369	153.82***
	(0.480)	(0.493)	(0.483)	
if household was on any social safety net program	0.131	0.199	0.154	14.21***
	(0.337)	(0.400)	(0.361)	
if household was hit by climatic shock	0.356	0.303	0.419	42.74***
	(0.480)	(0.460)	(0.494)	
Land area owned by household in ha	0.951	0.272	0.773	693.81***
	(0.703)	(0.455)	(0.709)	
Household distance in (KMs) to nearest road	7.248	5.106	9.440	185.40***
	(7.442)	(7.364)	(9.818)	
Household-specific distance to a market (KMs)	25.233	15.306	22.497	245.04***
	(13.524)	(13.411)	(13.433)	
Value of assets (MWK) [b]	284,129	225,446	272,802	17.74***
	(292,416)	(336,089)	(325,076)	
Observations	337	1235	1673	

Note: Numbers shown are averages and their corresponding standard deviations are presented in parenthesis. [a] Household heads without formal qualification is the reference category in the empirical analysis. [b] Value of assets winsorized at 5 percent both at the top and bottom. * p<0.10, ** p<0.05, *** p<0.010

A higher percentage of households that are net buyers at baseline (59%) own a phone compared to households that are self-sufficient and net sellers (36%) at baseline, on average.

This may suggest that households that are net buyers at baseline have a better access to marketing information than households that are net sellers and self-sufficient at baseline. As we expected, a higher proportion of households that are net buyers have been on social safety net programmes at baseline (20%) compared to those households that are self-sufficient (15%) and net sellers (13%) at baseline. A higher proportion of households that are self-sufficient have been affected by climatic shocks at baseline (42%) relative to those households that are net sellers (36%) and net buyers (30%) at baseline. This suggests that households that are self-sufficient at baseline are more vulnerable to the effects of climatic shocks than households that are net sellers and net buyers at baseline. Households that are net sellers at baseline (0.95 ha) own more land than households that are self-sufficient (0.77 ha) and net buyers at baseline (0.27 ha), on average. This suggests that access to land among households that are net buyers at baseline is limited than among households that are net sellers and self-sufficient at baseline (Chirwa, 2009). On average, households that are net buyers at baseline reside closer to the trunk and primary road (5.1 km) than households that are self-sufficient (9.4 km) and net sellers at baseline (7.2 km). Further, most households that are net buyers at baseline reside closer to the market or consumption centre (15.3 km) than households that are self-sufficient (22.5 km) and net sellers (25.2 km) at baseline, on average. These findings indicate that households that are net buyers at baseline have a better access to markets (*i.e.*, a local large market or a consumption centre) compared to households that are self-sufficient and net sellers at baseline, which suggests that most households that are net sellers and self-sufficient at baseline reside in remote areas. The value of assets owned is higher (MWK 284,129) among households that are net sellers at baseline than among households that are self-sufficient (MWK 272,802) and net buyers (MWK 225,446) at baseline, on average.

3.6 Empirical results

3.6.1 Impact by household heterogeneity

3.6.1.1 Production differential effects of fuel reform

Table 3.4 presents the results of the differential impact of the fuel policy reform on maize production by household status, equation [3.1], where immediate differential effects are in column 1 and persistent differential effects are in column 2. The table indicates that the direction of both immediate (column 1, row 1) and persistent (column 2, row 1) differential effect of the policy reform on share of land allocated for maize production is positive and significant at 95% confidence interval among households that are net buyers relative to households that are self-sufficient, on average. This indicates that the policy reform has increased the share of land allocated for maize production by 13.5 percentage points in the short run and by 23.1 percentage points in the long run among households that are net buyers compared to households that are self-sufficient, *ceteris paribus*. However, there are no significant differential impacts of the policy reform on maize production among households that are net sellers relative to households that are self-sufficient (row 2). Thus, the effect of the policy reform on maize production is similar for households that are net sellers and self-sufficient.[55]

Table 3.4: Impact of the fuel policy reform on share of land allocated for maize production by household status.

Variables	Immediate effects	Persistent effects
Post reform x net buyer	0.135***	0.231***
	(0.0162)	(0.0265)
Post reform x net seller	-0.000223	0.0366
	(0.0208)	(0.0342)
Other covariates	Yes	Yes
Year FE	Yes	Yes
Household FE	Yes	Yes
N	6204	3044

Note: Linear regression absorbing multiple levels of fixed effects (reghdfe) estimator results in columns 1 and 2. The dependent variable is share of land allocated for maize production. Standard errors are clustered at the household level in parentheses * p<0.10, ** p<0.05, *** p<0.010. We did not control for seasonality in maize production because most of the maize production is under rain-fed agriculture from November through to April.

This confirms that there are heterogeneous average differential effects of the policy reform that varies with household status as we expected. Overall, these findings may indicate that the policy reform has increased maize production among households that are net buyers

[55] We find similar results at the extensive margin of maize production where the likelihood of allocating all land for maize production has increased by 11% in the short run and 17% in the long run among net buyers compared to self-sufficient households (see table B.4 in the appendix). However, there are no significant differential impacts on net sellers.

relative to households that are in autarky, which has positive food security implications, on average. However, these findings mask important heterogeneous differential effects of the policy reform that varies with market access where we base our predictions because it involves transport costs. We present heterogeneous differential effects of the policy reform that varies with household status and market access in the next section.

3.6.1.2 Consumption differential effects of fuel reform

Table 3.5 presents the results of the differential impact of the fuel policy reform on per capita consumption by household status, equation [3.1], where immediate differential effects are in columns 1-4 and persistent differential effects are in columns 5-8. The table shows that the policy reform has increased per capita consumption by 20 percent (=100[exp(0.18)-1]) in the short run (column 1, row 1) and by 25 percent (=100[exp(0.22)-1]) in the long run (column 5, row 1) among households that are net buyers relative to households that are self-sufficient. Conversely, per capita consumption has decreased by -16 percent (=100[exp(-0.18)-1]) in the long run (column 5, row 2) among households that are net sellers relative to households that are self-sufficient, on average. However, there is no significant impact on households that are net sellers in the short run at 95% confidence interval (column 1, row 2).

Turning on now to per capita non-food consumption, the table shows that the policy reform has reduced per capita non-food consumption by -10 percent (=100[exp(-0.11)-1]) in the short run (column 2, row 2) and by -16 percent (=100[exp(-0.17)-1]) in the long run (column 6, row 2) among households that are net sellers relative to households that are self-sufficient, *ceteris paribus.* However, there are no significant impacts on households that are net buyers (columns 2 and 6, row 1). The finding on households that are net buyers is consistent with Hasan (2016) who found that the 2007/08 food price shock did not affect non-food consumption among net buyers of rice relative to self-sufficient households in Bangladesh, on average. However, the finding on net sellers is in contrast with Hasan (2016) who found that the food price shock did not affect non-food consumption on net sellers, on average.

Moving on now to per capita non-maize food consumption, the results indicate that per capita non-maize food consumption (columns 3 and 7, row 1) has improved in the short run by 12 percent (=100[exp(0.11)-1]) as well as in the long run by 17 percent (=100[exp(0.16)-1]) among households that are net buyers relative to households that are self-sufficient after the policy reform, *ceteris paribus*. This suggests that households that are net buyers consume more non-maize food per capita than households that are self-sufficient after the policy reform in both the short-run and long-run. Conversely, households that are net sellers have only experienced a reduction in per capita non-maize food consumption in the long run by -25 percent (=100[exp(-0.22)-1]) (column 7, row 2) relative to households that are self-sufficient, on average. These findings are in contrast with Hasan (2016) who found that the 2007/08 food price shock reduced non-rice food consumption among net buyers of rice, while net sellers of rice increased their non-rice food consumption relative to self-sufficient households in Bangladesh, on average. Although this related study did not investigate the effects of higher rice prices on rice production across household groups, the author concluded that net buyers lost in non-rice food consumption to maintain rice consumption, while net sellers increased non-rice food consumption because they gained in income from higher rice prices.

Turning to per capita quantity of maize consumption, the table indicate that households that are net sellers have only experienced a reduction in maize consumption in the short run by -9 percent (=100(-0.09)) (column 4, row 2) relative to households that are self-sufficient, on average. Thus, households that are net sellers consumed less quantity of maize per capita than households that are self-sufficient after the policy reform in the short run, *ceteris paribus*. However, there are no significant impacts on households that are net sellers in the long run (column 8, row 2), and on households that are net buyers in both the short-run and long-run (columns 4 and 8, row 1). These findings are consistent with Hasan (2016), except for net sellers in the short run. This finding suggests that net buyers increased maize production to maintain their maize consumption (*i.e.*, consume more maize from own production) in both the short-run and long-run, while net sellers only reduced maize consumption in the short run but maintained their maize production in both the short-run and long-run relative to households in autarky after the policy reform. Using the value of maize consumed, we find that the value of maize consumed per capita and unit cost of maize have increased in both the short-run and long-run among households that are net buyers relative to households that are self-sufficient after the policy reform (see table B.5 in the appendix). Although the direction of the differential effect of the policy on the value of maize consumed per capita remains negative in both the short-run and long-run among households that are net sellers relative to households that are self-sufficient, the differential effects are only marginally significant in the long run. This indicates that households that are net buyers consume the same quantity in both the short-run and long-run; but the value of their maize consumption has gone up because they face higher prices, while households that are net sellers consume less quantity of maize in the long run and face lower prices relative to households that are in autarky.[56,57]

[56] We find similar results on net buyers relative to households that are in autarky in both the short-run and long-run using the full sample (see table B.6 in the appendix). However, differential effects on per capita quantity of maize consumption becomes significant among households that are net sellers in the long run relative to self-sufficient households. This suggests that per capita quantity of maize consumption among net sellers is sensitive to the presence of outliers in the data.

[57] Although the sizes of the differential effects are different, we find similar results on per capita consumption, per capita non-food and non-maize food consumption among households that are net buyers and net sellers relative to households that are self-sufficient using the fixed effects Poisson estimator (see table B.8 in the appendix). This means that the results are not sensitive to the choice of a functional form.

Table 3.5: Impact of the fuel policy reform on per capita consumption by household status

Variable	Immediate effect				Persistent effect			
	Log of per capita consumption	Log of per capita consumption (non-food)	Log of per capita consumption (non-maize food)	Per capita consumption (maize)	Log of per capita consumption	Log of per capita consumption (non-food)	Log of per capita consumption (non-maize food)	Per capita consumption (maize)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post reform x net buyer	0.184***	0.0168	0.113***	-0.00387	0.216***	-0.0215	0.157***	0.0255
	(0.0254)	(0.0277)	(0.0386)	(0.0308)	(0.0406)	(0.0445)	(0.0589)	(0.0425)
Post reform x net seller	-0.0647*	-0.105***	-0.00150	-0.0901**	-0.178***	-0.166***	-0.219***	-0.0930
	(0.0348)	(0.0371)	(0.0589)	(0.0450)	(0.0499)	(0.0630)	(0.0790)	(0.0584)
Other covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Household FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	5752	5768	6202	5578	2822	2838	3042	2712

Note: Linear regression absorbing multiple levels of fixed effects (reghdfe) estimator results in columns 1, 2, 3, 5, 6, and 7, and Poisson pseudo-likelihood regression with multiple levels of fixed effects (ppmlhdfe) estimator results in columns 4 and 8. The dependent variable is log of per capita annual consumption (trimmed at the top 5%) in columns 1 and 5, log of per capita annual non-food consumption (trimmed at the top 5%) in columns 2 and 6, log of per capita annual non-maize food consumption (trimmed at the top 5%) in columns 3 and 7, per capita weekly quantity of maize consumption (trimmed at the top 5%) in columns 4 and 8. Standard errors are clustered at the household level in parentheses * p<0.10, ** p<0.05, *** p<0.010

Table 3.6 presents the results of the differential impact of the fuel policy reform on consumption shares and dietary quality across households, equation [3.1], where immediate differential effects are in columns 1-3 and persistent differential effects are in columns 4-6. The table shows that maize consumption share (columns 2 and 4, row 1) has increased among households that are net buyers relative to households that are self-sufficient in both the short run and long run, on average. The increase in maize consumption share is attributed to higher maize prices that net buyers face. This means that households that are net buyers spend relatively a larger share of their income on maize consumption compared to households that are self-sufficient after the policy reform, *ceteris paribus*. This indicates that households that are net buyers have become more prone to maize insecurity than households that are self-sufficient after the policy reform (Smith & Subandoro, 2007). However, differential effects are not significant on households that are net sellers (row 2). Considering now non-maize food consumption share (columns 1 and 4) and quality of diet (columns 3 and 6), the results indicate that there are no significant differential effects on households that are net buyers and net sellers relative to households that are self-sufficient in either the short-run or long-run, on average. This suggests that the effects of the policy reform on non-maize food consumption and quality of diet are similar across the household groups.[58]

[58] We find similar results using the full sample (see table B.7 in the appendix). This indicates that the results are not sensitive to the presence of outliers.

Table 3.6: Impact of the reform on consumption shares and dietary quality by household status

Variable	Immediate effect			Persistent effect		
	Non-maize food expenditure share	Maize expenditure share	DD	Non-maize food expenditure share	Maize expenditure share	DD
	(1)	(2)	(3)	(4)	(5)	(6)
Post reform x net buyer	-0.0122	0.0734***	0.0110	-0.0116	0.0993***	0.0109
	(0.00767)	(0.00573)	(0.0101)	(0.0110)	(0.00811)	(0.0157)
Post reform x net seller	0.0144	0.00306	-0.0178	-0.00434	-0.00673	-0.0124
	(0.0122)	(0.0103)	(0.0156)	(0.0163)	(0.0139)	(0.0244)
Other covariates	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Household FE	Yes	Yes	Yes	Yes	Yes	Yes
N	5962	6078	6128	2918	2974	3030

Note: Linear regression absorbing multiple levels of fixed effects (reghdfe) estimator results in columns 1, 2, 4, and 5, and Poisson pseudo-likelihood regression with multiple levels of fixed effects (ppmlhdfe) estimator results in columns 3 and 6. The dependent variable is share of consumption on non-maize food (trimmed at the bottom and top 1%) in columns 1 and 4, share of consumption on maize (trimmed at the top 1%) in columns 2 and 5, and dietary diversity score (trimmed at the bottom 1%) in columns 3 and 6. Standard errors are clustered at the household level in parentheses * p<0.10, ** p<0.05, *** p<0.010

Similarly, this confirms that there are heterogeneous average differential effects of the policy reform that varies with household status as we expected. The increase in per capita consumption among households that are net buyers is driven by the increase in maize production, which allow them to serve in maize consumption and increase non-maize food consumption in both the short-run and long-run. Conversely, the reduction in per capita consumption among households that are net sellers is driven by the decrease in per capita non-food consumption in both the short-run and long-run. Overall, this may suggest that the standard of living has improved among net buyers, while among net sellers it has deteriorated relative to households that are self-sufficient. However, these findings mask important heterogeneous differential effects of the policy reform that varies with market access where we base our predictions because it involves transport costs. We present heterogeneous differential effects of the policy reform that varies with household status and market access in the next section.

3.6.2 Impact by household and geographic heterogeneity

3.6.2.1 Production differential effects of fuel reform

Table 3.7 presents the results of estimating equation [3.2] to reveal immediate (column 1) and persistent (column 2) differential impacts of the policy reform on maize production by household status and distance. The table reveals that distance increases the differential effect of the policy reform on the share of land allocated to maize production in the short run among households that are in autarky after the reform (column 1, row 3). Conversely, distance reduces the differential effect of the policy reform on the share of land allocated for maize production in the short among households that are net buyers (column 1, row 4) relative to households that are self-sufficient, on average. However, the effects of distance on the effect of the reform on maize production are not different across the household groups in the long run at 95% confidence interval (column 2), on average.

Figure 3.2 presents the results of estimating equation [3.2] to show how the differential effect of the policy reform on maize production varies by levels of distance for households that are net buyers and households that are net sellers relative to households that are in autarky, where immediate differential effects are in the first row and persistent differential effects are in the second row. The figure shows that the differential effect of the policy reform on the share of land allocated for maize production is more positive for households that are in autarky in remote areas than for households that are in autarky closer to market in the short after the reform (Panel A). Thus, households that are in autarky have increased the share of land allocated for maize production for each increase in distance away from the market after the policy reform. However, the differential effects do not persist. As we expected, the differential effect of the policy reform is more positive closer to the market than in the remote areas among households that are net buyers relative to households that are in autarky in both the short-run and long-run. However, the differential effects are significant only for distance of less than 41 km in the short run (Panel B) and for distance of less than 56 km in the long run (Panel D) among households that are net buyers at 95% confidence interval. These findings suggest that households that are net buyers who reside closer to the market have increased the share of land allocated for maize production to avoid higher producer prices as transport costs of getting the maize grain from remote areas increases relative to households that are in autarky. These findings are consistent with Omamo (1998) in that production of staple maize declines as transport costs of accessing markets increases. However, differential effects closer to the remote areas are not significant at 95% confidence interval. Households that are net buyers that reside closer to the market have experienced an increase in maize production relative to households that are self-sufficient, while those that reside closer to the remote locations have experienced the same effect on maize production as households that are self-sufficient both in the short run and in the long run consistent with Fuje (2019). Turning to households that are net sellers (Panels B and D), the differential effect of the policy reform is more positive closer to the market than in the remote areas relative to households that are in autarky in both the short-run and long-run. However, the differential effects are not significant at 95% confidence interval. Thus, the effects of the policy reform on maize production are similar between households that are net sellers and self-sufficient as distance changes.[59,60]

[59] We find similar results at the extensive margin of maize production, except that the differential effect on households that are in autarky becomes significant in the long run (see figure B.3 in the appendix).

[60] We find similar results using a discrete variable of remoteness that takes on a value of one if the distance is above the mean, and zero otherwise (see figure B.4 in the appendix). This indicates that our results are not sensitive to how market access is measured as a continuous or a discrete variable.

Table 3.7: Impact of the policy reform on share of land allocated for maize production by household status and distance

Variables	Immediate effect	Persistent effect
Post reform x net buyer (γ_1)	0.224***	0.277***
	(0.0285)	(0.0448)
Post reform x net seller (γ_2)	0.0474	0.137*
	(0.0481)	(0.0817)
Post reform x minimum distance (γ_3)	0.00312***	0.00162
	(0.000714)	(0.00104)
Post reform x net buyer x minimum distance (γ_4)	-0.00433***	-0.00234
	(0.00117)	(0.00204)
Post reform x net seller x minimum distance (γ_5)	-0.00222	-0.00397
	(0.00172)	(0.00266)
Other covariates	Yes	Yes
Year FE	Yes	Yes
Household FE	Yes	Yes
N	6204	3044

Note: Linear regression absorbing multiple levels of fixed effects (reghdfe) estimator results in columns 1 and 2. The dependent variable is share of land allocated for maize production. Distance to market is minimum distance to either local large market or consumption centre. Standard errors are clustered at the household level in parentheses. * p<0.10, ** p<0.05, *** p<0.010. We did not control for seasonality in maize production because most of the maize production is under rain-fed agriculture from November through to April.

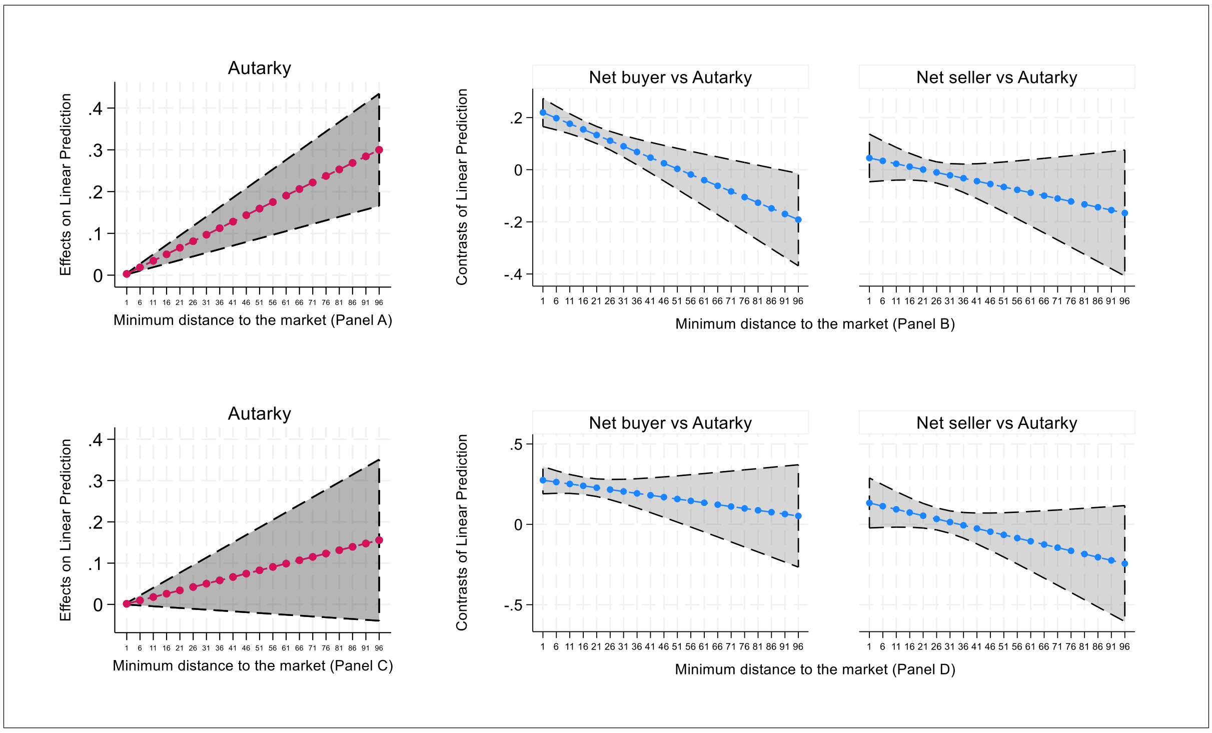

Figure 3.2: Impact of the policy reform on the share of land allocated for maize production that varies with distance to the market at 95% confidence interval.
Note: Panels for net buyers and net sellers use a common y-axis. Immediate differential effects are in the first row and persistent differential effects are in the second row.

We find similar results using distance to the district capital as an alternative measure of market access (see figure B.5 in the appendix). However, the long-run differential effect of the policy reform on the share of land allocated for maize production becomes significant among households that are in autarky in the long run. This finding indicates that the differential effects of the policy reform on maize production is less sensitive to measures of market access.

In summary, our findings confirms that differential effects of the policy reform on maize production varies with household status and market access, on average. We find that the differential effect of the policy reform on maize production increases with distance among households that are in autarky in the short run. Further, we find that households that are net buyers that reside closer to the market have increased maize production relative to households that are self-sufficient, however, the effects are similar in remote locations both in the short-run and long-run. The effects of the policy reform on maize production are similar between households that are net sellers and self-sufficient as distance changes both in the short-run and long-run. Contrary to our expectations, the differential effects of the reform persist over time among households that are net buyers that reside closer to the market relative to households that are self-sufficient. Further, the differential effects of the policy reform on maize production are less sensitive to how market access is measured.

3.6.2.2 Consumption differential effects of fuel reform

Table 3.8 presents the results of estimating equation [3.2] to reveal immediate (columns 1-4) and persistent (columns 5-8) differential impacts of the policy reform on consumption by household status and distance. The table shows that distance reduces the differential effect of the policy reform on per capita consumption (columns 1 and 5, row 3), and per capita non-maize food consumption (columns 3 and 7, row 3) among households that are in autarky after the reform. Turning to net buyers, the table shows that distance reduces the differential effect of the policy reform on per capita non-food consumption relative to households that are in autarky in the long run (column 6, row 4). Moving on to net sellers, the table shows that distance increases the differential effect of the policy reform on per capita non-maize food consumption relative to households that are in autarky in the long run (column 7, row 5). However, the differential effects of distance on net buyers and net sellers relative to households that are in autarky are not significant at 95% confidence interval in the short run (columns 1-4).

Figure 3.3 presents the results of estimating equation [3.2] to show how the differential effect of the policy reform on consumption varies by levels of distance for households that are net buyers and households that are net sellers relative to households that are in autarky, where immediate differential effects are in the first row and persistent differential effects are in the second row. The figure shows that the differential effect of the policy reform on per capita consumption is more negative in the remote areas than closer to the market both in the short run and in the long run among households that are in autarky (Panels A and C). Thus, households that are in autarky have reduced per capita consumption for each increase in distance away from the market after the policy reform. Turning to net buyers, the figure shows that the differential effect of the policy reform on per capita consumption is more positive closer to the market than in the remote areas (*i.e.*, contrary to our expectation) both in the short run and in the long run relative to households that are in autarky. However, the differential effects are significant only for distance of less than 41 km in the short run (Panel B) and for distance of less than 31 km in the long run (Panel D) at 95% confidence interval.

These differential effects decrease as distance away from the market increases in the short run and in the long run. Households that are net buyers that reside closer to the market have experienced an increase in per capita consumption relative to households that are self-sufficient, while those that reside closer to the remote locations have experienced the same effect on per capita consumption as households that are self-sufficient both in the short run and in the long run. This finding is consistent with Stifel & Minten (2017) who found that per capita consumption reduces as transport costs of accessing markets increases in Ethiopia. Moving on to net sellers, the differential effect of the policy reform on per capita consumption is more negative closer to the market than in remote areas (*i.e.*, contrary to our expectation) in the long run. However, the differential effects are significant only for distance of less than 31 km in the long run at 95% confidence interval. Thus, households that are net sellers that reside closer to the market have experienced a reduction in per capita consumption relative to households that are in autarky, while those that reside closer to the remote locations have experienced the same effect on per capita consumption as households that are self-sufficient in the long run, *ceteris paribus.*

Turning on now to per capita non-food consumption, the figure shows that the differential effect of the policy reform is more negative in remote areas than closer to the market among households that are net buyers relative to households that are in autarky in the long run (Panel H). However, the differential effect of the policy reform is significant only for distance of greater than 51 km at 95% confidence interval. Thus, non-food consumption for households that are net buyers in remote areas has declined relative to households that are in autarky, while those closer to the market have experienced the same effect on per capita non-food consumption as households that are in autarky in the long run, *ceteris paribus*. Moving on to net sellers, the figure shows that the differential effect of the policy reform is more negative in remote areas than closer to the market in the short run (Panel F), while the differential effect of the policy reform is more negative closer to the market than in remote areas in the long run (Panel H) relative to households that are in autarky. However, the differential effect of the policy reform is significant only for distance of between 16 and 91 km in the short run, and for distance of between 11 and 36 km in the long run at 95% confidence interval. Thus, non-food consumption for households that are net sellers in remote areas has declined relative to households that are in autarky, while for those closer to the market or more remote have experienced the same effect on per capita non-food consumption as households that are in autarky in the short run. However, non-food consumption for households that are net sellers closer to the market has declined relative to households that are in autarky, while those closer to the market or more remote have experienced the same effect on per capita non-food consumption as households that are in autarky in the long run.

Moving on to per capita non-maize food consumption, figure 3.3 shows that the differential effect of the policy reform is more negative in remote areas than closer to the market among households that are in autarky in the short run and in the long run after the reform (Panels I and K). Thus, households that are in autarky in remote areas consume less non-maize food per capita than those closer to the market for each increase in distance away from the market after the reform, *ceteris paribus*. Moving on to net sellers, the differential effect of the policy reform is more negative closer to the market than in remote areas. However, the differential effect is significant only for distance of less than 26 km in the long run at 95% confidence interval (Panel L). Thus, per capita non-maize food consumption has declined for households that are net sellers that reside closer to the market relative to households that are in autarky, while those in remote locations have experienced the same effect on per capita non-maize food consumption as households that are in autarky in the long run.

Considering now per capita quantity of maize consumption, figure 3.3 shows that the differential effect of the policy reform is more negative in remote areas than closer to the market in the short run (*i.e.*, contrary to our expectation), while in the long run the differential effect of the policy is more negative closer to the market than in remote areas (*i.e.*, consistent with our expectation) among households that net sellers relative to households that are in autarky. However, the differential effects are significant only for distance of greater than 26 km in the short run (Panel N) relative to households that are in autarky at 95% confidence interval. Thus, households that are net sellers in remote areas consume less quantity of maize per capita relative to households that are in autarky, while those closer to the market have experienced the same effect on per capita quantity of maize consumption as households that are in autarky in the short run. The switch in sign between the short run and long run suggests that households that are net sellers reduced maize consumption and sold more maize in the short run, but they increased maize consumption and sold less maize given lower producer prices in remote locations relative to households that are in autarky though it is not significant in the long run. However, there is no clear evidence of how the differential effect of the policy reform on quantity of maize consumption changes with distance among households that are net buyers relative to households that are in autarky. Using the value of maize consumed, we find that the value of maize consumed per capita and unit cost of maize are more positive closer to the market than in remote areas in the short run and in the long run among households that are net buyers relative to households that are self-sufficient (see figure B.6 in the appendix). These findings mean that households that are net buyers consume the same quantity of maize per capita in the short run and in the long run; but the value of their maize consumption has gone up because they face higher prices.[61,62,63]

[61] We find similar results on per capita consumption, per capita non-food and non-maize food consumption using the fixed effects Poisson estimator (see figure B.7 in the appendix). Thus, the results are not sensitive to the choice of the functional form.
[62] We find similar results using the full sample, except for per capita quantity of maize consumption among net buyers in the long run (see figure B.8 in the appendix). This indicates that the results are less sensitive to the presence of outliers in the data, except for maize consumption.
[63] We find similar results using a discrete variable of remoteness (see figure B.10 in the appendix). This indicates that our results are not sensitive to how market access is measured as a continuous or a discrete variable.

Table 3.8: Impact of the fuel policy reform on consumption by household status and distance

Variables	Immediate effect				Persistent effect			
	Log of per capita consumption	Log of per capita consumption (non-food)	Log of per capita consumption (non-maize food)	Per capita consumption (maize)	Log of per capita consumption	Log of per capita consumption (non-food)	Log of per capita consumption (non-maize food)	Per capita consumption (maize)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post reform x net buyer (γ_1)	0.210*** (0.0413)	0.0736 (0.0463)	0.0998* (0.0599)	0.0301 (0.0515)	0.239*** (0.0667)	0.0900 (0.0715)	0.110 (0.0921)	0.0481 (0.0740)
Post reform x net seller (γ_2)	-0.0243 (0.0725)	-0.00344 (0.0819)	0.00727 (0.113)	0.0430 (0.0940)	-0.343*** (0.106)	-0.205 (0.146)	-0.582*** (0.154)	-0.238 (0.148)
Post reform x minimum distance (γ_3)	-0.00304*** (0.00113)	0.000121 (0.00127)	-0.00427** (0.00181)	-0.00184 (0.00141)	-0.00471*** (0.00178)	0.00181 (0.00201)	-0.0102*** (0.00271)	-0.000124 (0.00208)
Post reform x net buyer x minimum distance (γ_4)	-0.00287 (0.00181)	-0.00342* (0.00203)	-0.00114 (0.00278)	-0.00341 (0.00242)	-0.00373 (0.00295)	-0.00663** (0.00324)	-0.00183 (0.00418)	-0.00178 (0.00395)
Post reform x net seller x minimum distance (γ_5)	-0.00128 (0.00246)	-0.00402 (0.00289)	0.0000858 (0.00410)	-0.00509 (0.00324)	0.00684* (0.00361)	0.00106 (0.00512)	0.0152*** (0.00524)	0.00541 (0.00508)
Other covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Household FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	5752	5768	6202	5578	2822	2838	3042	2712

Note: Linear regression absorbing multiple levels of fixed effects (reghdfe) estimator results in columns 1, 2, 3, 5, 6, and 7, and Poisson pseudo-likelihood regression with multiple levels of fixed effects (ppmlhdfe) estimator results in columns 4 and 8. The dependent variable is log of per capita annual consumption (trimmed at the top 5%) in columns 1 and 5, log of per capita annual non-food consumption (trimmed at the top 5%) in columns 2 and 6, log of per capita annual non-maize food consumption (trimmed at the top 5%) in columns 3 and 7, per capita weekly quantity maize consumption (trimmed at the top 5%) in columns 4 and 8. Standard errors are clustered at the household level in parentheses. * p<0.10, ** p<0.05, *** p<0.010

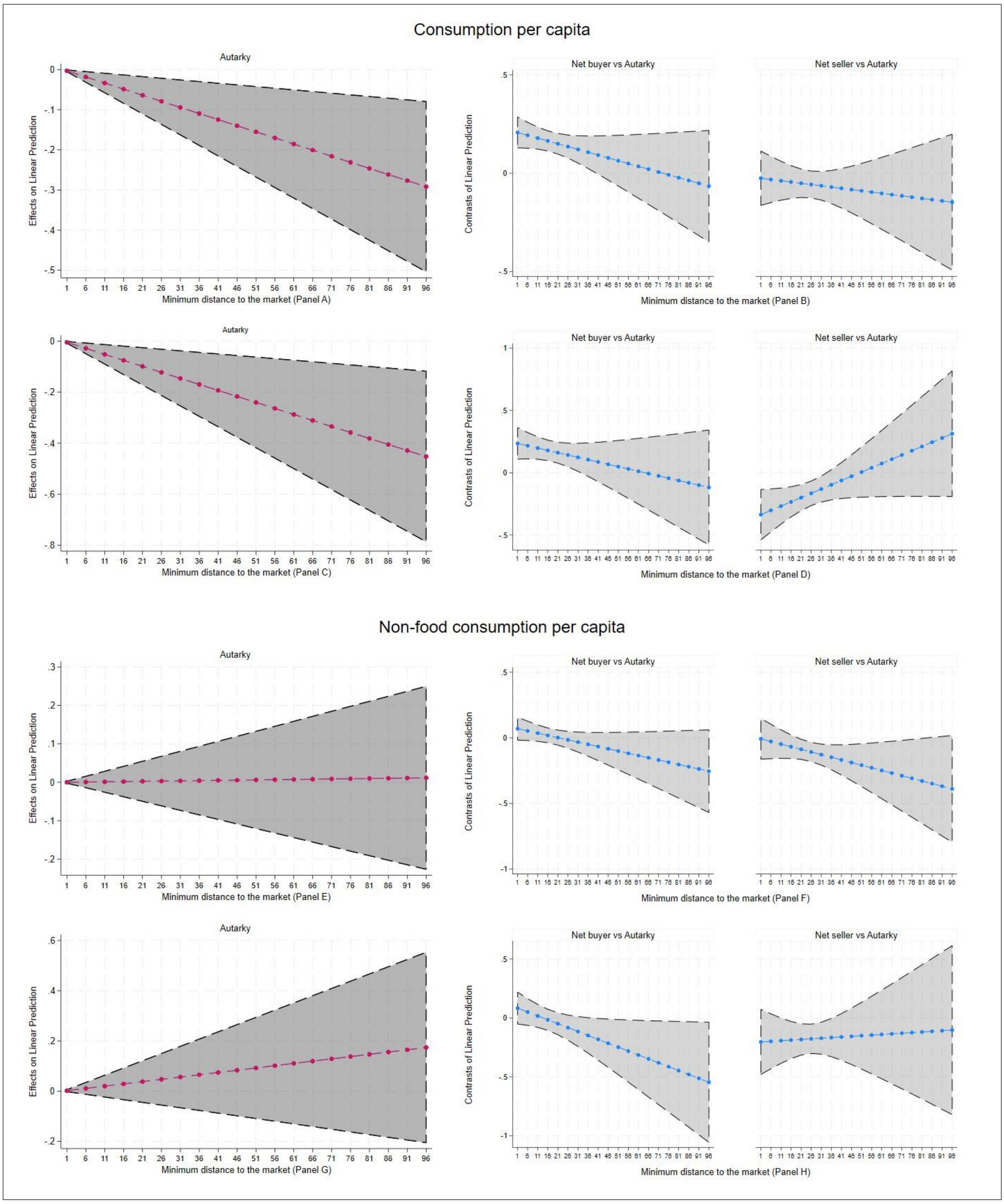

Figure 3.3: Impact of the policy reform on consumption-related outcomes that varies with distance to the market at 95% confidence interval.

Note: Panels for net buyers and net sellers use a common y-axis. Immediate differential effects are in the first row and persistent differential effects are in the second row.

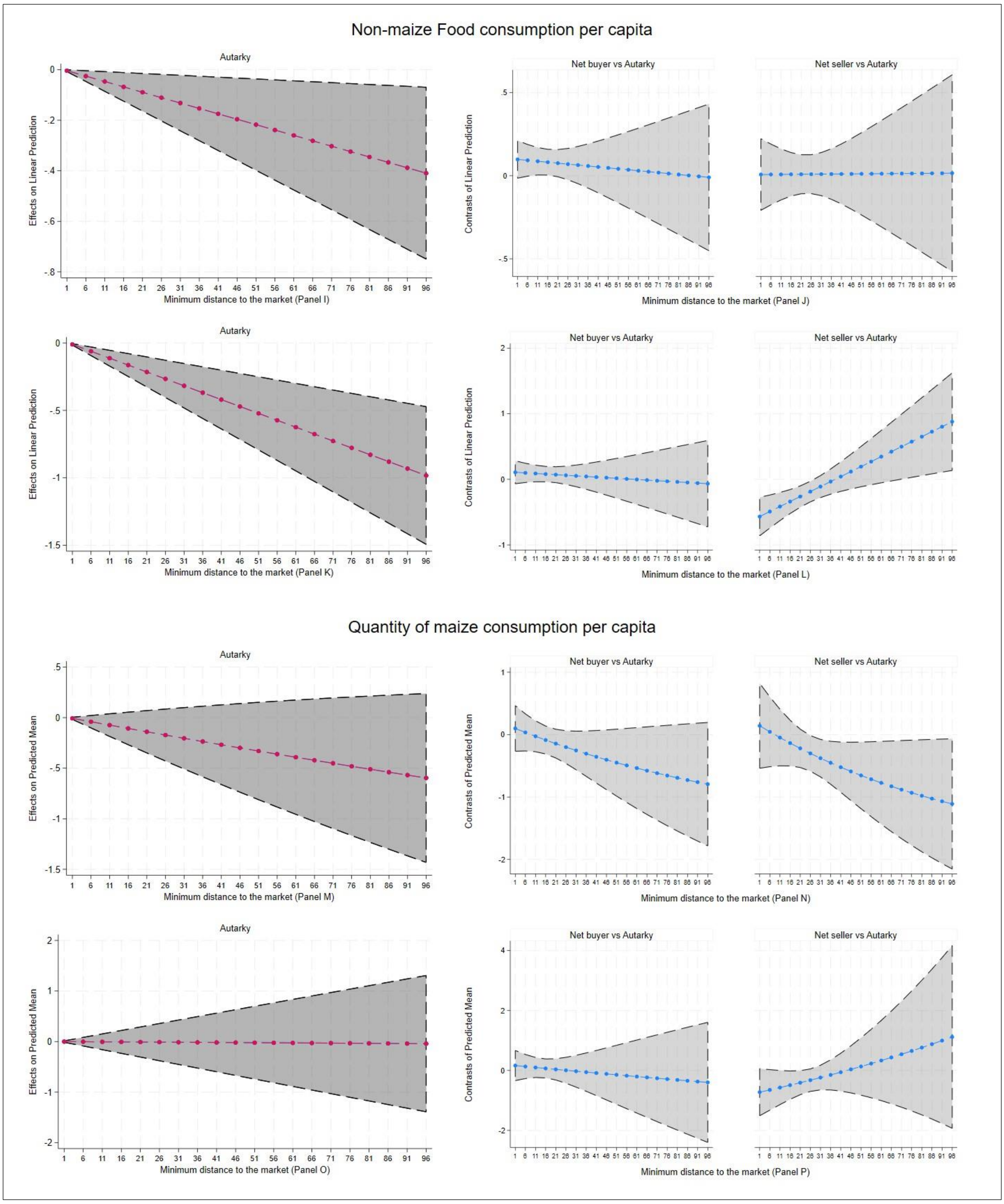

Figure 3.3 continued…

Table 3.9 presents the results of estimating equation [3.2] to reveal immediate (columns 1-3) and persistent (columns 4-6) differential impacts of the policy reform on consumption shares and dietary quality by household status and distance. The table shows that distance reduces the differential effect of the policy reform on maize expenditure share (column 2, row 3) in the short run, and non-maize food expenditure in the long run among households that are in autarky. The effects of distance on the differential impact of the policy reform on households that are net buyers and net sellers relative to households that are in autarky are not significant at 95% confidence interval.

Figure 3.4 presents the results of estimating equation [3.2] to show how the differential effect of the policy reform on consumption shares and dietary quality varies by levels of distance for households that are net buyers and households that are net sellers relative to households that are in autarky, where immediate differential effects are in the first row and persistent differential effects are in the second row. The figure shows that the differential effects of the policy reform on non-maize food expenditure share are not significant at 95% confidence interval. Thus, the effects of the policy reform on non-maize food consumption share are similar across household groups as distance changes.

Considering maize consumption share, figure 3.4 shows that the differential effect of the policy reform is more positive closer to the market than in remote areas (*i.e.*, as we expected) in the short run, while in the long run the differential effect of the policy is more positive in rural areas than closer to the market among households that are net buyers. However, the differential effects are significant only for distance of less than 61 km in the short run (Panel F) and for each increase in distance away from the market in the long run (Panel H) relative to households that are in autarky at 95% confidence interval. While some households that are net buyers closer to the market became more prone to maize insecurity than those households in rural areas in the short run, all households that are net buyers have become more prone to maize insecurity in the long run than households that are in autarky as distance away from the market increases. These differential effects decrease as distance away from the market increases in the short run, while in the long run the differential effects increase with distance. Although we did not find clear evidence of how the differential effect of the policy reform on quantity of maize consumption changes with distance among households that are net buyers relative to households that are in autarky, the switch in sign between the short-run and long-run suggests that households that are net buyers spend relatively a larger share of their income on maize consumption compared to households that are self-sufficient after the policy reform, *ceteris paribus*. Turning to net sellers, the differential effects of the policy on maize expenditure share are not significant at 95% confidence interval. Moving on to dietary quality, the differential effects of the policy reform are not significant in both the short-run and long-run as distance increases away from the market, on average.[64,65]

Table 3.9: Impact of the reform on consumption shares and dietary quality by household status and distance

	Immediate effect			Persistent effect		
Variables	**Non-maize food expenditure share (1)**	**Maize expenditure share (2)**	**DD (3)**	**Non-maize food expenditure share (4)**	**Maize expenditure share (5)**	**DD (6)**
Post reform x net buyer (γ_1)	-0.0145	0.0761***	-0.0113	-0.0172	0.0870***	0.0133
	(0.0122)	(0.00922)	(0.0155)	(0.0173)	(0.0134)	(0.0239)
Post reform x net seller (γ_2)	0.00998	-0.000451	-0.0222	-0.0423	-0.00860	-0.0221
	(0.0248)	(0.0206)	(0.0307)	(0.0319)	(0.0346)	(0.0557)
Post reform x minimum distance (γ_3)	0.000128	-0.000679**	-0.000377	-0.00101**	-0.000752*	-0.000270
	(0.000369)	(0.000314)	(0.000515)	(0.000504)	(0.000436)	(0.000814)
Post reform x net buyer x minimum distance (γ_4)	0.000210	-0.000501	0.00137*	-0.0000952	0.000496	-0.000315
	(0.000553)	(0.000438)	(0.000792)	(0.000824)	(0.000716)	(0.00129)
Post reform x net seller x minimum distance (γ_5)	0.000164	0.000207	0.000226	0.00157	0.000219	0.000407
	(0.000910)	(0.000750)	(0.00116)	(0.00107)	(0.00120)	(0.00184)
Other covariates	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Household FE	Yes	Yes	Yes	Yes	Yes	Yes
N	5962	6078	6128	2918	2974	3030

Note: Linear regression absorbing multiple levels of fixed effects (reghdfe) estimator results in columns 1, 2, 4, and 5, and Poisson pseudo-likelihood regression with multiple levels of fixed effects (ppmlhdfe) estimators results in columns 3 and 6. The dependent variable is share of consumption on non-maize food (trimmed at the bottom 5%) in columns 1 and 4, share of consumption on maize (trimmed at the top and bottom 1%) in columns 2 and 5, and dietary diversity score (trimmed at the top and bottom 1%) in columns 3 and 6. Standard errors are clustered at the household level in parentheses. * p<0.10, ** p<0.05, *** p<0.010

[64] We find similar results using the full sample, except for maize expenditure share among net buyers in the long run (see figure B.9 in the appendix). This indicates that the results are not sensitive to the presence of outliers in the data, except for maize expenditure shares.

[65] Overall, we find similar results using a discrete variable of remoteness (see figure B.11 in the appendix). This indicates that findings are not sensitive to how market access is measured as a continuous or a discrete variable.

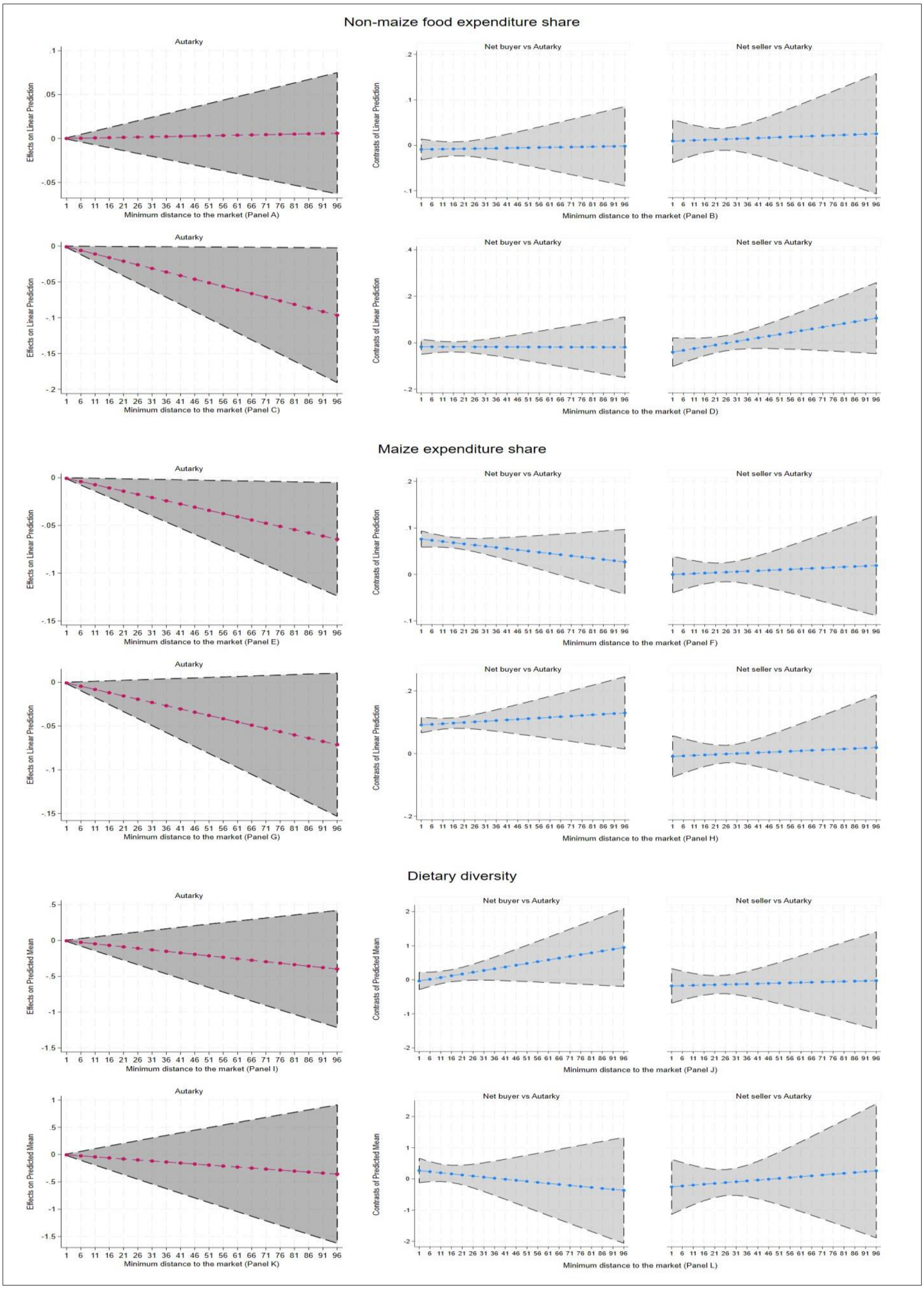

Figure 3.4: Impact of the policy reform on consumption shares and dietary quality that varies with distance to the market at 95% confidence interval.

Note: Panels for net buyers and net sellers use a common y-axis. Immediate differential effects are in the first row and persistent differential effects are in the second row.

Using alternative measures of market access, distance to the district capital, we find similar results of the differential impact of the policy reform on welfare-related outcomes except for per capita consumption (see figures B.12 and B.13 in the appendix). For instance, we find that the differential effects on per capita consumption among households that are in autarky are not significant in the short run and in the long run after the reform (figure B.12, Panels A and C). Turning to net buyers, the differential effect of the policy reform on per capita consumption is more positive closer to the district capital than in remote areas in the short run, while in the long run the differential effect is more positive in the remote area than closer to the district capital (figure B.12, Panels B and D) relative to households that are in autarky. Moving on to net sellers, the differential effect of the policy reform on per capita consumption becomes significant for distance of greater than 30 km in the short run relative to autarkic households (figure B.12, Panel B), while in the long run the effects are similar between net sellers and autarkic households at 95% confidence interval. Overall, these findings provide evidence that the differential impact of the policy reform by household status and market access is less sensitive to how market access is measured.

In summary, our findings confirms that the differential effects of the policy reform on consumption vary with household status and market access, on average. We find that households that are in autarky in remote areas have reduced per capita consumption and consume less non-maize food per capita than those closer to the market for each increase in distance away from the market after the reform in the short and the long run. Further, we find that households that are net buyers that reside closer to the market have experienced an increase in per capita consumption in the short run and the long run, but they have become more prone to maize insecurity in the short run and less prone to maize insecurity in the long run than households that are self-sufficient. Conversely, households that are net buyers that reside in remote locations have experienced a decrease in non-food consumption in the long run, became less prone to maize insecurity in the short run, but they have become more prone to maize insecurity in the long run relative to households that are in autarky. Turning to net sellers, we find that households that reside in remote locations have experienced a reduction in per capita non-food consumption and per capita quantity of maize consumption in the short run, while those that reside closer to the market have experienced a reduction in per capita consumption, per capita non-food consumption and per capita non-maize food consumption relative to households that are in autarky in the long run. Contrary to our expectations, the differential effects of the reform persist over time, *ceteris paribus*. Further, the differential effects of the policy reform on households' consumption are less sensitive to how market access is measured. While classification of households is based on consumption and production of maize, households also cultivate other crops in addition to maize (Jones et al., 2014; Snapp & Fisher, 2015). Therefore, we would also expect other crops to be affected by the fuel price reform.

3.7 Conclusions and policy implications

This chapter investigates the extent to which the reform to Malawi's fuel policy adopted in 2012 increase or decrease agricultural production and consumption of staple maize grain among households. We estimate both the immediate and persistent differential effects of the policy reform. We hypothesise that the policy reform has immediate differential effects on agricultural production and consumption of staple maize grain, but the differential effects do not persist over time once the policy is adopted. Consistent with the previous literature, we anticipate a heterogeneous differential impact of the fuel price reform on households that vary with household status (*i.e.*, net-seller, net-buyer or self-sufficient in staple maize grain), and market access. Our approach is to examine whether there are differences in the effect of the policy reform on welfare related outcomes by comparing households to which the policy reform should have a larger effect (*i.e.*, net sellers and net buyers) to households to which it should have a smaller effect (*i.e.*, self-sufficient households). To estimate the immediate and persistent differential impacts of the fuel policy reform on households, we use three waves of nationally representative panel data from IHPS, which were implemented in 2010, 2013, and 2016 as part of LSMS-ISA for Malawi and apply the fixed effects estimator.

Our analysis confirms that there are heterogeneous differential impacts of the fuel price reform on households' maize production and consumption that vary with household status and market access. For instance, we find that households that are in autarky in remote areas have increased maize production more than those closer to the market, but they have experienced a reduction in per capita consumption and per capita maize consumption in the short and the long run after the reform. Conversely, households that are net buyers that reside closer to the market have experienced an increase in maize production and per capita consumption and have become less prone to maize insecurity, while those that reside in remote locations have experienced a reduction in non-food consumption and have become more prone to maize insecurity relative to households that are in autarky in the long run. Turning to net sellers, we find that households that reside in remote locations have experienced a reduction in per capita non-food consumption and per capita maize consumption in the short run, while those that reside closer to the market have experienced a reduction in per capita consumption, per capita non-food consumption and per capita non-maize food consumption relative to households that are in autarky in the long run. Overall, these findings are robust to alternative measures of market access. Contrary to our expectations, the differential effects of the reform persist over time, *ceteris paribus*.

One of the limitations of our identification strategy is that households might switch across the groups after the policy reform whereby those that did not sell any maize grain before the policy might start selling maize grain after the policy or those that sold maize grain before the policy might stop selling maize grain after the policy likely leading to biased results, our estimation does not capture any differential effect of the policy reform on households' entry and exit across the groups. Further, the country experienced a few crises over the period under investigation such as intensification of the border dispute with neighbouring Tanzania in 2011 (Mahony et al., 2014), a constitutional crisis in 2011/2012 (Cammack, 2012), flood shock in 2014/2015, and drought shock in 2015/2016 agricultural seasons (Floodlist, 2015; Government of Malawi, 2016c), which may lead to an overestimation of the effects of the fuel price reform on our outcomes of interest. Despite these limitations, our study has demonstrated that fuel subsidy removal, which increases transport costs has both short- and long-term consequences on households' agricultural production and consumption.

Unlike in Ethiopia where the government implemented a food subsidy scheme to reduce welfare effects of higher food prices in consumption centres at the time of fuel subsidy removal, the government of Malawi removed the fuel subsidy without a safety net programme to protect households from the effects of the policy reform. Fuje (2019) finds that ultra-poor households that were part of the food subsidy scheme gained from the fuel policy reform compared to those that did not participate in the scheme. Overall, our results indicate that households that are in autarky in remote areas increased maize production more than those closer to the market but lost in consumption due to the increase in transport costs of accessing markets. Further, households that are net buyers that reside closer to the market increased maize production, consumption, and became less prone to maize insecurity, while those that reside in remote locations lost in non-food consumption and became more prone to maize insecurity relative to households that are in autarky. Conversely, households that are net sellers that reside in remote locations lost in non-food consumption and maize consumption, while those that reside closer to the market lost in consumption, non-food consumption and non-maize food consumption relative to households that are in autarky. During the first 6 months of the subsidy removal, the government of Malawi saved about MWK36 billion (Kasalika, 2013). Therefore, the lesson for other countries that are considering rescaling or removing fuel subsidies is to contain welfare differential effects of the reform on households that reside both closer to markets or consumption centres and in remote areas. One way this can be achieved is to use the money that is saved to implement social safety net programmes at the time of subsidy removal like in Ethiopia. Such programs can target autarkic households, net buyers, and net sellers that reside in remote locations more than those that reside closer to the market or consumption centre to mitigate negative differential effects of the reform. For countries that are already implementing social safety net programmes such as input subsidies, expanding the coverage of these programmes can lead to an increase in cultivation of staple crops such as maize to increase food availability in the short term (Chibwana et al., 2010). Lowering structural barriers to trade in rural areas through investment in modern transport infrastructure such as railways and information communication technologies will be vital to reduce transport costs between the farm and consumption centre in the long-term.

3.8 Appendix B

Table B.1: Description of variables used in the analysis.

Variables	Description	Type
Dependent variables		
Share of land allocated to maize	The ratio of cultivated land under maize to total land cultivated by the household.	Continuous
Value of annual per capita expenditures	Annualized value of per capita expenditure divided by adult male equivalent scales (AME) in Malawian Kwacha.	Continuous
Annualized per capita non-food consumption	Annualized value of per capita non-food expenditure divided by adult male equivalent scales (AME) in Malawian Kwacha.	Continuous
Annualized per capita non-maize food consumption	Annualized value of per capita non-maize food expenditure divided by adult male equivalent scales (AME) in Malawian Kwacha.	Continuous
Per capita quantity of maize consumption	Weekly quantity of maize consumption divided by adult male equivalent scales (AME) in kilograms.	Continuous
Annualized share of non-maize food consumption to total consumption	The ratio of annualized share of non-maize food consumption expenditures to total household consumption expenditures.	Continuous
Annualized share of maize consumption to total consumption	The ratio of annualized share of maize consumption expenditures to total household consumption expenditures. Deaton (1989) refers this measure as consumption ratio.	Continuous
Household dietary diversity index	Household dietary diversity score measured as the number of food groups consumed.	Continuous
Independent variables		
# of males greater than 65 years old	The number of male members in the household greater than 65 years old.	Continuous
# of male adults 12-64 years old	The number of male adults in the household between the age of 12 and 65.	Continuous
# of females greater than 65 years old	The number of female members in the household greater than 65 years old.	Continuous
# of female adults 12-64 years old	The number of female adults in the household between the age of 12 and 65.	Continuous
# of children less than 12 years old	The number of children in the household below the age of 12.	Continuous
Marital status of head, =1 if has a spouse	Marital status of the household head.	Dummy
Gender of household head, =1 if male	Gender of the household head.	Dummy
Age of head in years	Age of the household head in years.	Continuous
Highest qualification of household head, none	If the household head does not have a formal qualification.	Dummy
Highest qualification of household head, primary	If the household head attended primary education.	Dummy
Highest qualification of household head, secondary	If the household head attended secondary education.	Dummy
Highest qualification of household head, tertiary	If the household head attended tertiary education.	Dummy
If household member participates in wage labour	If household has a member who participate in wage or casual labour.	Dummy
If household owns a phone	If the household has a phone.	Dummy
if household was on any social safety net program	If the household was a beneficiary of a social safety net program.	Dummy
if household was hit by climatic shock	If the household was hit by climatic shocks that affected its agricultural production.	Dummy
Land area owned by household in acres	Area of land owned by the household in acres	Continuous
Household distance in (KMs) to nearest road	Household distance to the nearest trunk road in kilometres	Continuous
Household-specific distance to a market	Minimum distance between the household and local large market or consumption centre in kilometres	Continuous
Value of assets (MWK)	Total value of assets for each household measured in Malawian Kwacha.	Continuous

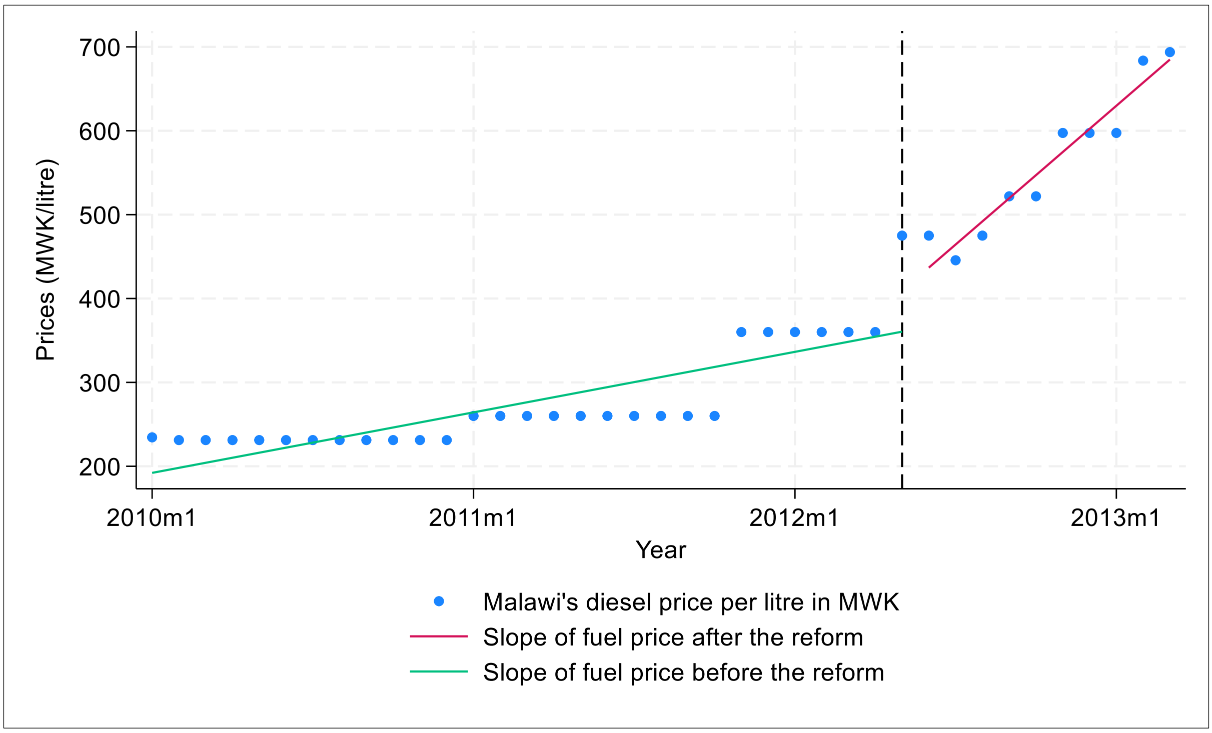

Figure B.1: Slope of diesel price before and immediately after the policy reform

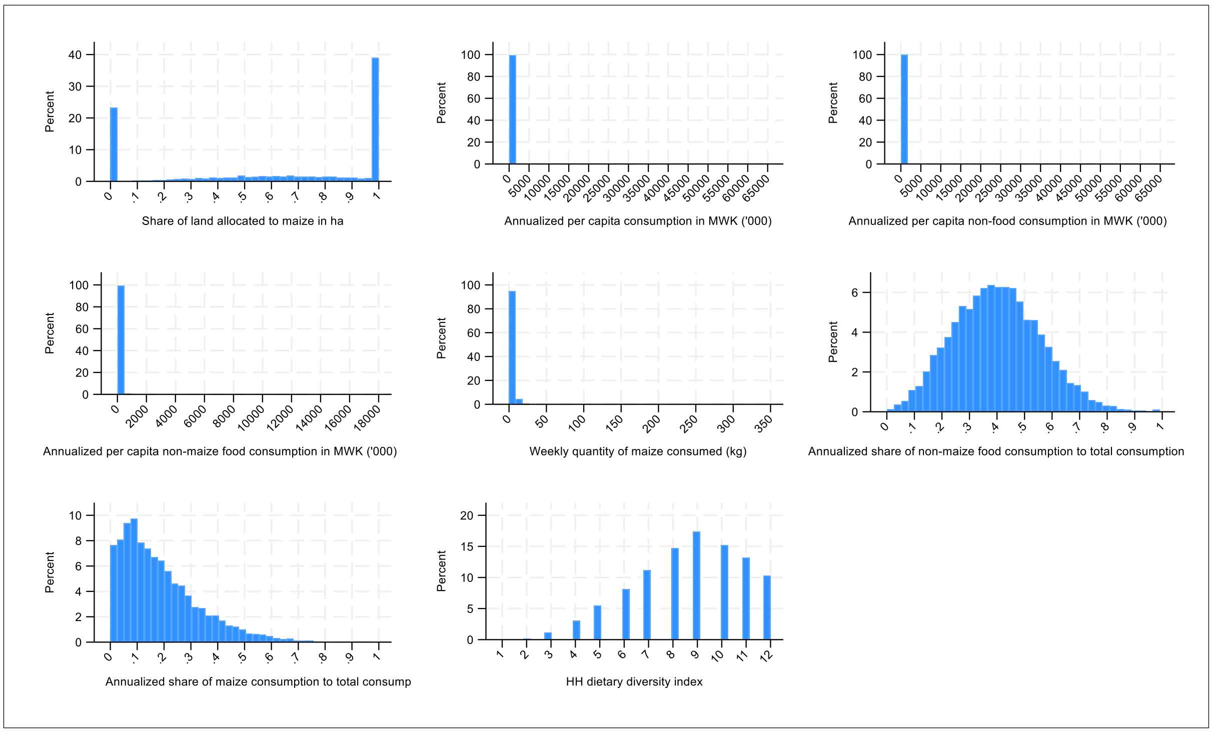

Figure B.2: Distribution of production and consumption indicators

Table B.2: Summary statistics of outcome indicators by survey wave

Variables	Wave 1	Wave 2	Wave 3	Mean Diff. (Wald chi (2) statistic)
Share of land allocated to maize	0.630	0.604	0.602	8.15**
	(0.400)	(0.400)	(0.421)	
Annualized per capita consumption in MWK [a]	139,143	143,820	117,351	1744.20***
	(85,851)	(99,793)	(81,017)	
Annualized per capita non-food consumption in MWK [a]	50,544	58,462	58,227	1678.49***
	(41,676)	(49,599)	(47,735)	
Annualized per capita food consumption in MWK [a]	83,528	81,252	56,738	1506.24***
	(47,516)	(51,420)	(36,016)	
Weekly quantity of maize consumed (kg) [a]	3.682	3.518	3.239	63.58***
	(2.031)	(1.961)	(1.598)	
Annualized share of non-maize food consumption to total consumption [b]	0.403	0.418	0.358	193.66***
	(0.155)	(0.145)	(0.133)	
Annualized share of maize consumption to total consumption [c]	0.213	0.163	0.145	350.23***
	(0.156)	(0.120)	(0.102)	
HH dietary diversity index [d]	8.564	8.864	8.579	35.12***
	(2.240)	(2.103)	(2.248)	
Observations	3245	3103	1524	

Note: Numbers shown are averages and their corresponding standard deviations are presented in parenthesis. [a] trimmed at the top 5%, [b] trimmed at the top and bottom 1%, [c] trimmed at the top 1%, and [d] trimmed at the bottom 1%. * p<0.10, ** p<0.05, *** p<0.010

Table B.3: Characteristics of households by survey wave

Variables	Wave 1	Wave 2	Wave 3	Mean Diff. (Wald chi (2) statistic)
# of males greater than 65 years old	0.076	0.094	0.219	77.05***
	(0.266)	(0.296)	(0.612)	
# of male adults 12-64 years old	1.356	1.552	1.509	56.51***
	(1.029)	(1.124)	(1.139)	
# of females greater than 65 years old	0.091	0.119	0.254	97.19***
	(0.289)	(0.338)	(0.627)	
# of female adults 12-64 years old	1.410	1.595	1.602	68.63***
	(0.937)	(1.048)	(1.066)	
# of children less than 12 years old	1.872	1.972	1.801	15.54***
	(1.479)	(1.478)	(1.457)	
Marital status of head, =1 if has a spouse	0.740	0.739	0.724	1.52
	(0.439)	(0.439)	(0.447)	
Gender of household head, =1 if male	0.774	0.770	0.739	7.27**
	(0.418)	(0.421)	(0.439)	
Age of head in years	41.943	44.914	46.946	124.75***
	(16.026)	(15.524)	(14.607)	
Highest qualification of household head, none	0.668	0.670	0.883	442.82***
	(0.471)	(0.470)	(0.322)	
Highest qualification of household head, primary	0.107	0.104	0.045	83.64***
	(0.309)	(0.305)	(0.208)	
Highest qualification of household head, secondary	0.182	0.178	0.061	232.83***
	(0.386)	(0.382)	(0.239)	
Highest qualification of household head, tertiary	0.043	0.048	0.011	84.32***
	(0.203)	(0.215)	(0.105)	
If household member participates in wage labour	0.382	0.413	0.625	276.43***
	(0.486)	(0.493)	(0.484)	
If household owns a phone	0.451	0.530	0.602	104.58***
	(0.498)	(0.499)	(0.490)	
if household was on any social safety net program	0.169	0.366	0.369	421.36***
	(0.374)	(0.482)	(0.483)	
if household was hit by climatic shock	0.368	0.325	0.446	62.18***
	(0.482)	(0.469)	(0.497)	
Land area owned by household in ha	0.601	0.912	0.666	12.32***
	(0.677)	(7.759)	(0.781)	
Household distance in (KMs) to nearest road	7.563	7.661	7.747	0.45
	(8.954)	(9.126)	(9.443)	
Household-specific distance to a market	20.167	20.165	19.659	1.72
	(13.951)	(13.949)	(13.417)	
Value of assets (MWK)	255,955	435,030	526,604	371.85***
	(326,915)	(594,088)	(699,477)	
Observations	3245	3103	1524	

Note: Numbers shown are averages and their corresponding standard deviations are presented in parenthesis. * p<0.10, ** p<0.05, *** p<0.010

Table B.4: Impact of the fuel policy reform on allocating all land for maize production by household status

Variables	Immediate effect	Persistent effect
Post reform x net buyer	0.113***	0.173***
	(0.0210)	(0.0333)
Post reform x net seller	0.0386	0.0783
	(0.0324)	(0.0501)
Other covariates	Yes	Yes
Year FE	Yes	Yes
Household FE	Yes	Yes
N	6206	3044

Note: Linear regression absorbing multiple levels of fixed effects (reghdfe) estimator results in columns 1 and 2. The dependent variable is a dummy variable equal to 1 if all the land is allocated for maize production. Standard errors are clustered at the household level in parentheses * p<0.10, ** p<0.05, *** p<0.010

Table B.5: Impact of the reform on value of maize consumed and cost of maize consumed by household status.

Variable	Immediate effect		Persistent effect	
	Per capita maize consumed (1)	Unit cost of maize consumed (2)	Per capita maize consumed (3)	Unit cost of maize consumed (4)
Post reform x net buyer	0.436*** (0.0361)	0.395*** (0.0334)	0.655*** (0.0555)	0.550*** (0.0528)
Post reform x net seller	-0.0692 (0.0507)	0.0204 (0.0366)	-0.130* (0.0668)	-0.102 (0.0634)
Other covariates	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
Household FE	Yes	Yes	Yes	Yes
N	5574	5404	2726	2672

Note: Poisson pseudo-likelihood regression with multiple levels of fixed effects (ppmlhdfe) estimator results in columns 1 – 4. The dependent variable is value of maize consumed per capita (trimmed at the top 5%) in columns 1 and 3, and cost of maize consumed per kg (trimmed at the top 5%) in columns 2 and 4. Standard errors are clustered at the household level in parentheses * p<0.10, ** p<0.05, *** p<0.010

Table B.6: Impact of the fuel policy reform on consumption by household status using the full sample

Variables	Immediate effect				Persistent effect			
	Log of per capita consumption	Log of per capita consumption (non-food)	Log per capita consumption (food)	Per capita consumption (maize)	Log of per capita consumption	Log of per capita consumption (non-food)	Log per capita consumption (food)	Per capita consumption (maize)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post reform x net buyer	0.160***	-0.00947	0.113***	-0.0217	0.194***	-0.0595	0.157***	-0.0605
	(0.0273)	(0.0288)	(0.0386)	(0.0511)	(0.0415)	(0.0444)	(0.0589)	(0.0593)
Post reform x net seller	-0.0435	-0.0974**	-0.00150	-0.0868	-0.148**	-0.113	-0.219***	-0.161**
	(0.0375)	(0.0384)	(0.0589)	(0.0586)	(0.0592)	(0.0728)	(0.0790)	(0.0719)
Other covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Household FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	6202	6202	6202	6160	3042	3042	3042	3022

Note: Linear regression absorbing multiple levels of fixed effects (reghdfe) estimator results in columns 1, 2, 3, 5, 6, and 7, and Poisson pseudo-likelihood regression with multiple levels of fixed effects (ppmlhdfe) estimator results in columns 4 and 8. The dependent variable is log of per capita annual consumption (trimmed at the top 5%) in columns 1 and 5, log of per capita annual non-food consumption (trimmed at the top 5%) in columns 2 and 6, log of per capita annual food consumption (trimmed at the top 5%) in columns 3 and 7, per capita annual maize consumption (trimmed at the top 5%) in columns 4 and 8. Standard errors are clustered at the household level in parentheses * p<0.10, ** p<0.05, *** p<0.010

Table B.7: Impact of the fuel policy reform on consumption shares and dietary quality by household status using the full sample

Variables	Immediate effect				Persistent effect	
	Food expenditure share	Maize expenditure share	DD	Food expenditure share	Maize expenditure share	DD
	(1)	(2)	(3)	(4)	(5)	(6)
Post reform x net buyer	-0.0144*	0.0729***	0.0113	-0.0124	0.0988***	0.00953
	(0.00793)	(0.00613)	(0.0103)	(0.0115)	(0.00876)	(0.0157)
Post reform x net seller	0.0181	0.000548	-0.0188	-0.0103	-0.00966	-0.0177
	(0.0129)	(0.0108)	(0.0159)	(0.0165)	(0.0143)	(0.0247)
Other covariates	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Household FE	Yes	Yes	Yes	Yes	Yes	Yes
N	6202	6202	6202	3042	3042	3042

Note: Linear regression absorbing multiple levels of fixed effects (reghdfe) estimator results in columns 1, 2, 4, and 5, and Poisson pseudo-likelihood regression with multiple levels of fixed effects (ppmlhdfe) estimator results in columns 3 and 6. The dependent variable is share of consumption on food (trimmed at the bottom 5%) in columns 1 and 4, share of consumption on maize (trimmed at the top and bottom 1%) in columns 2 and 5, and dietary diversity score (trimmed at the top and bottom 1%) in columns 3 and 6. Standard errors are clustered at the household level in parentheses * p<0.10, ** p<0.05, *** p<0.010

Table B.8: Impact of the fuel policy reform on consumption by household status

Variables	Immediate effect			Persistent effect		
	Per capita consumption	**Per capita consumption (non-food)**	**Per capita consumption (non-maize food)**	**Per capita consumption**	**Per capita consumption (non-food)**	**Per capita consumption (non-maize food)**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Post reform x net buyer	0.141***	0.00934	0.105***	0.169***	-0.0394	0.147***
	(0.0262)	(0.0310)	(0.0331)	(0.0413)	(0.0473)	(0.0482)
Post reform x net seller	-0.0622	-0.111**	-0.00425	-0.186***	-0.166**	-0.225***
	(0.0381)	(0.0439)	(0.0522)	(0.0539)	(0.0691)	(0.0716)
Other covariates	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Household FE	Yes	Yes	Yes	Yes	Yes	Yes
N	5752	5768	5700	2822	2838	2800

Note: Poisson pseudo-likelihood regression with multiple levels of fixed effects (ppmlhdfe) estimator results in columns 1- 6. The dependent variable is per capita annual consumption (trimmed at the top 5%) in columns 1 and 4, per capita annual non-food consumption (trimmed at the top 5%) in columns 2 and 5, per capita annual non-maize food consumption (trimmed at the top 5%) in columns 3 and 6. Standard errors are clustered at the household level in parentheses * p<0.10, ** p<0.05, *** p<0.010

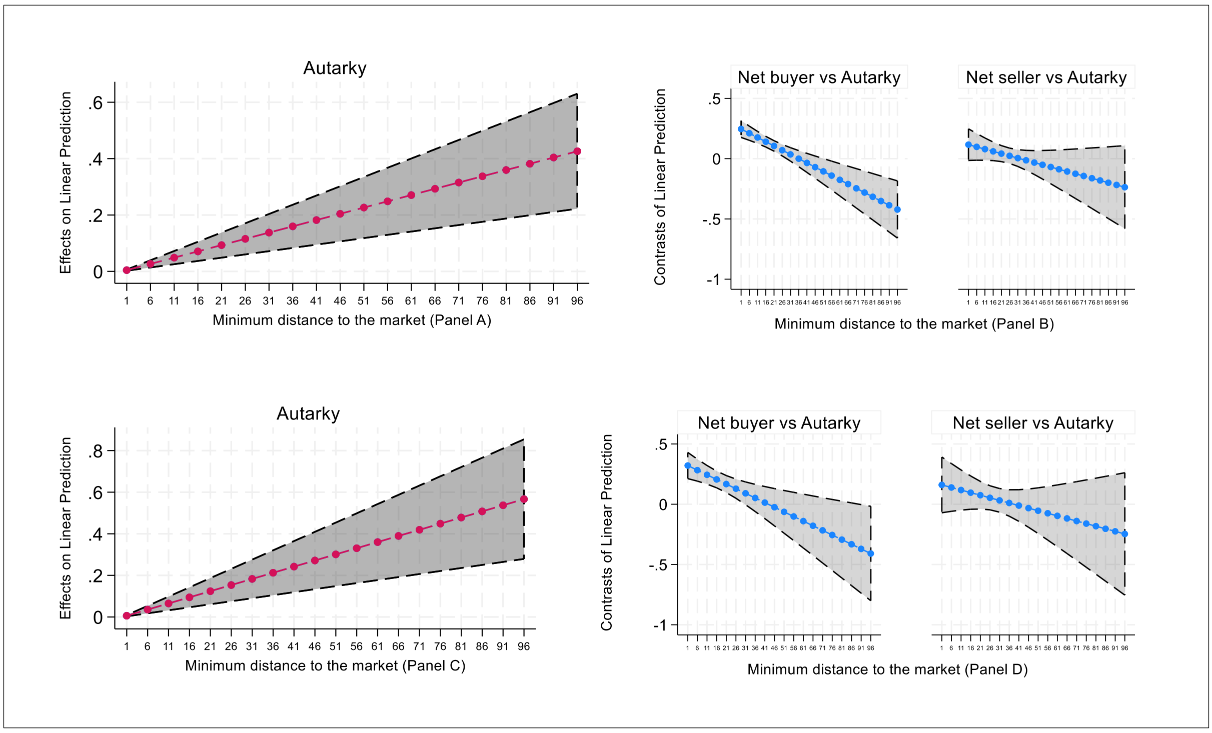

Figure B.3: Impact of the policy reform on allocating all land for maize production that varies with distance to the market at 95% confidence interval.

Note: Panels for net buyers and net sellers use a common y-axis. Immediate differential effects are in the first row and persistent differential effects are in the second row.

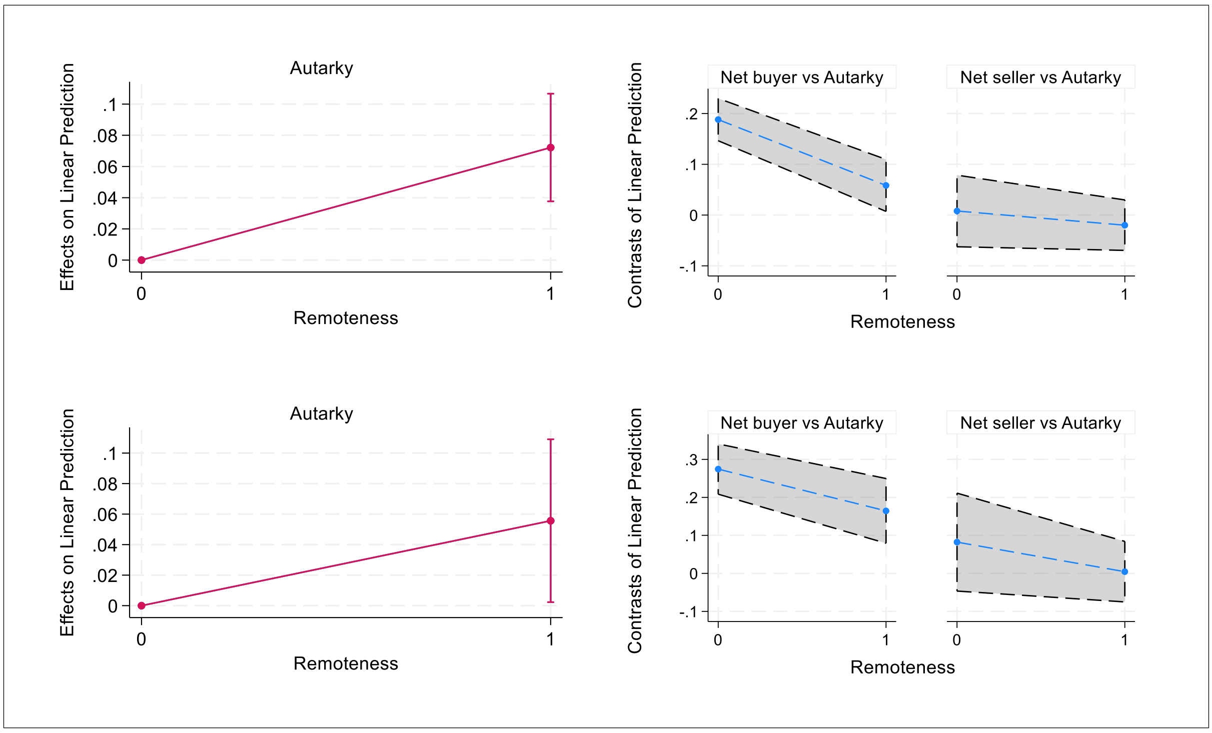

Figure B.4: Impact of the policy reform on share of land allocated for maize production that varies with remoteness at 95% confidence interval.

Note: Panels for net buyers and net sellers use a common y-axis. Immediate differential effects are in the first row and persistent differential effects are in the second row.

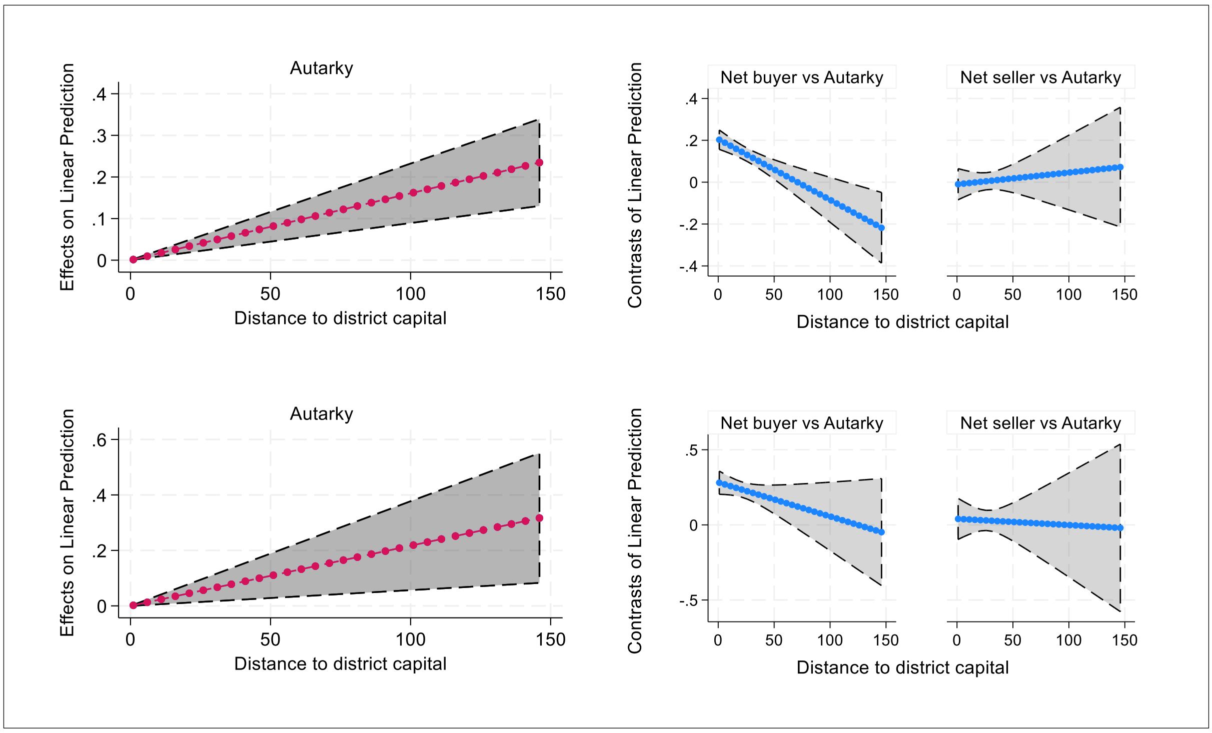

Figure B.5: Impact of the policy reform on the share of land allocated for maize production that varies with distance to the district capital at 95% confidence interval.
Note: Panels for net buyers and net sellers use a common y-axis. Immediate differential effects are in the first row and persistent differential effects are in the second row.

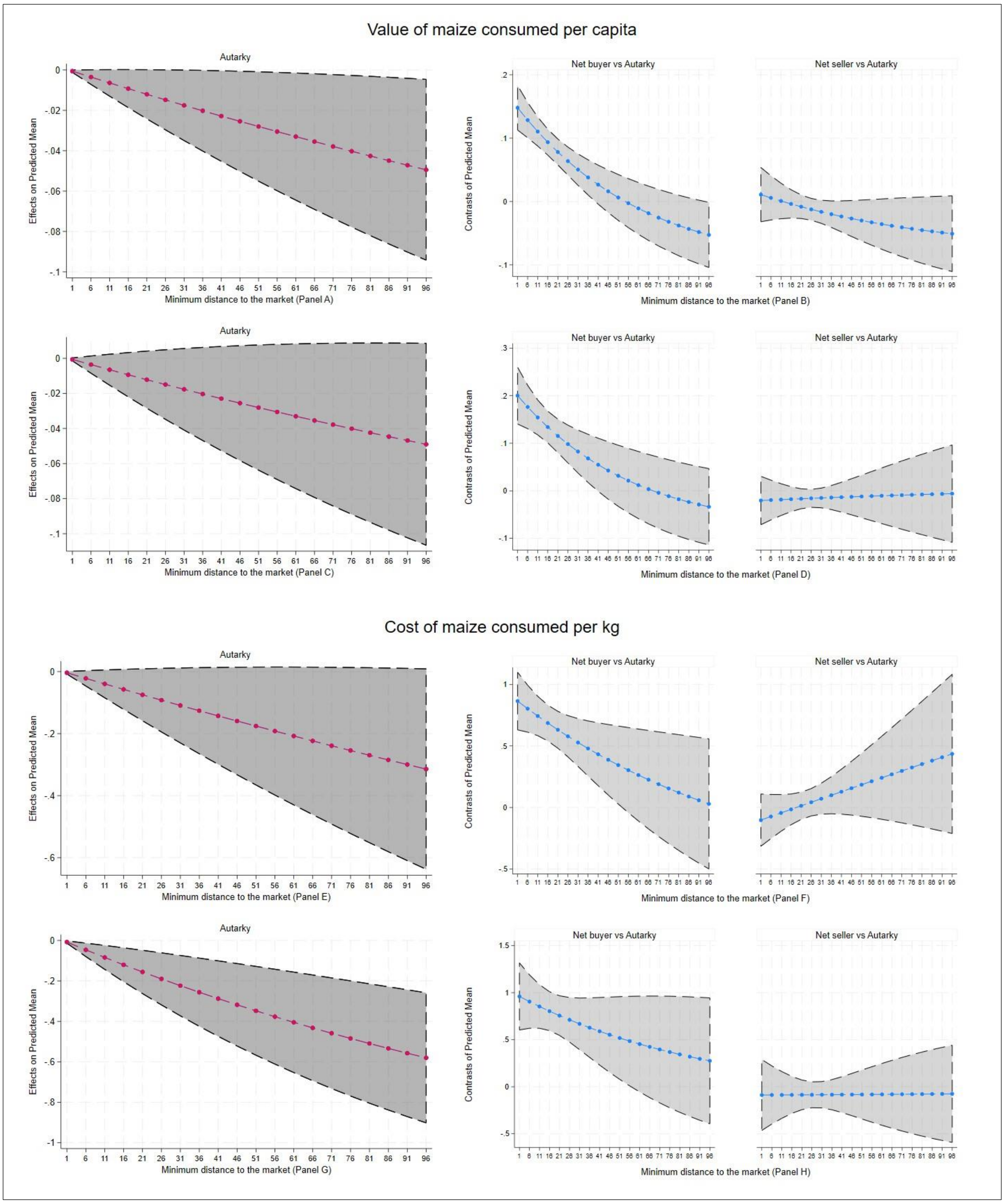

Figure B.6: Impact of the policy reform on quantity of maize consumed and cost of maize consumed that varies with distance to the market at 95% confidence interval.

Note: Panels for net buyers and net sellers use a common y-axis. Immediate differential effects are in the first row and persistent differential effects are in the second row.

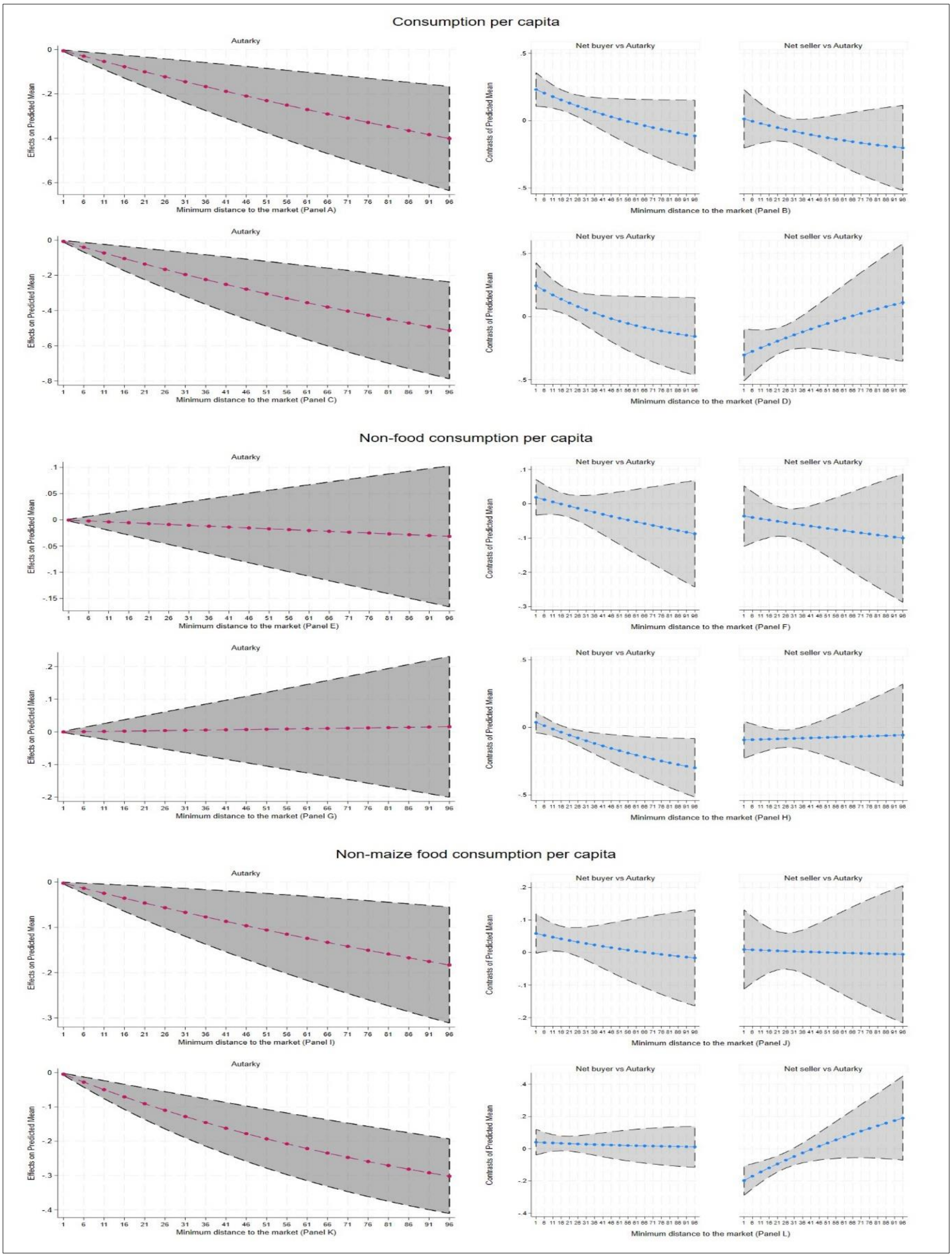

Figure B.7: Impact of the fuel policy reform on consumption that varies with distance to the market using Poisson pseudo-likelihood regression with multiple levels of fixed effects (ppmlhdfe) estimator at 95% confidence interval.

Note: Panels for net buyers and net sellers use a common y-axis. Immediate differential effects are in the first row and persistent differential effects are in the second row.

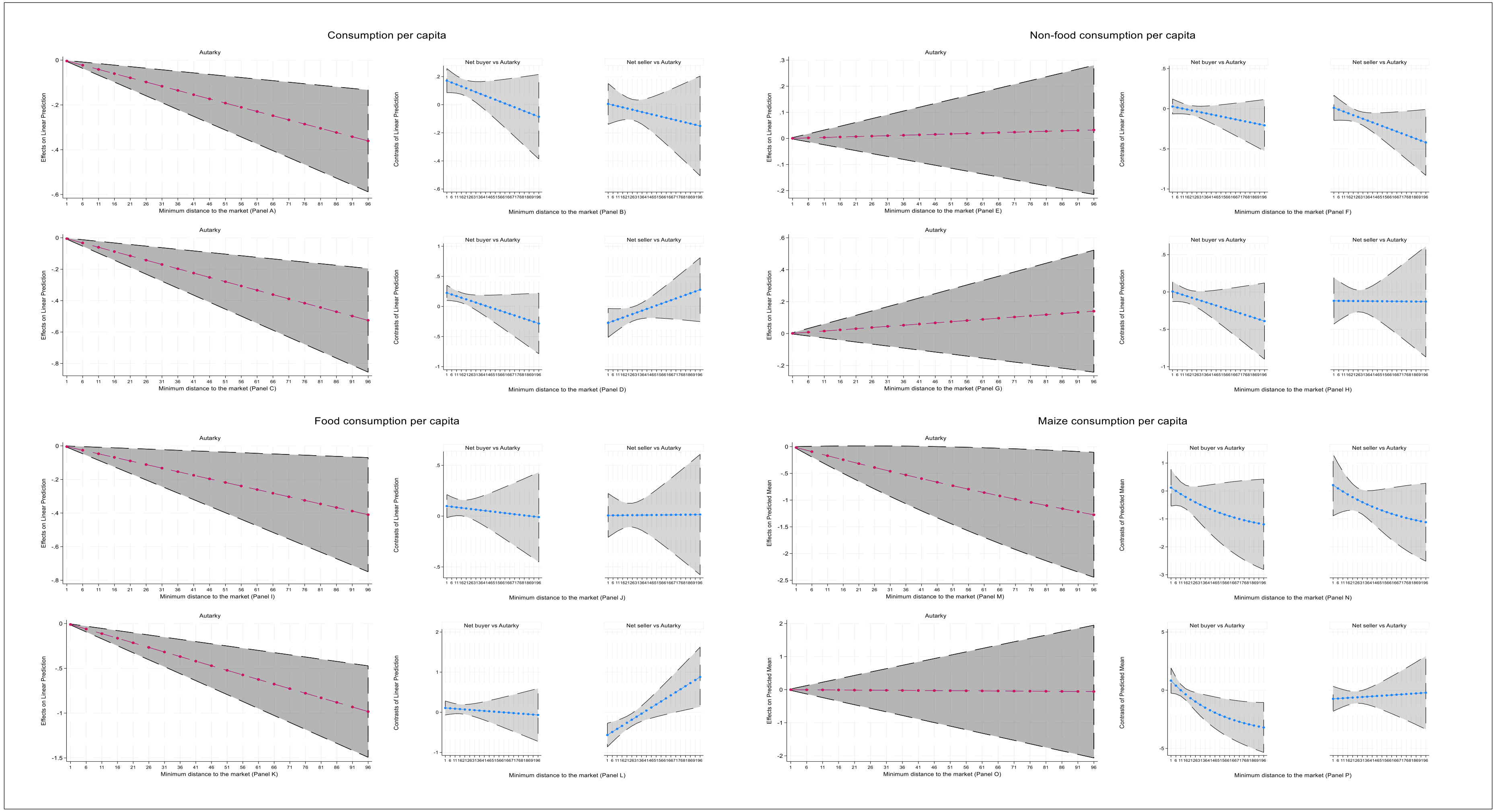

Figure B.8: Impact of the fuel policy reform on consumption that varies with distance to the market using the full sample at 95% confidence interval.

Note: Panels for net buyers and net sellers use a common y-axis. Immediate differential effects are in the first row and persistent differential effects are in the second row.

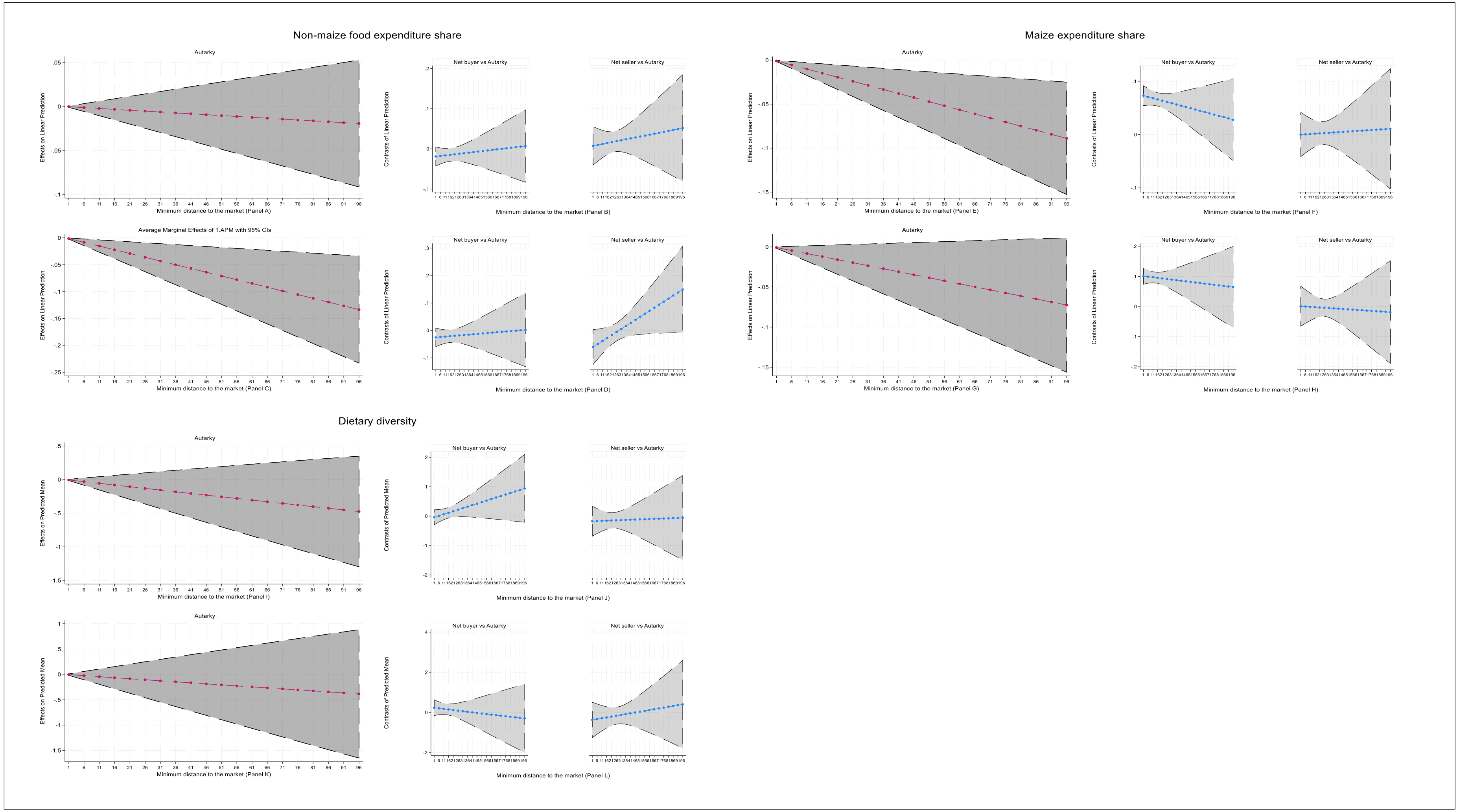

Figure B.9: Impact of the fuel policy reform on consumption that varies with distance to the market using the full sample at 95% confidence interval.

Note: Panels for net buyers and net sellers use a common y-axis. Immediate differential effects are in the first row and persistent differential effects are in the second row.

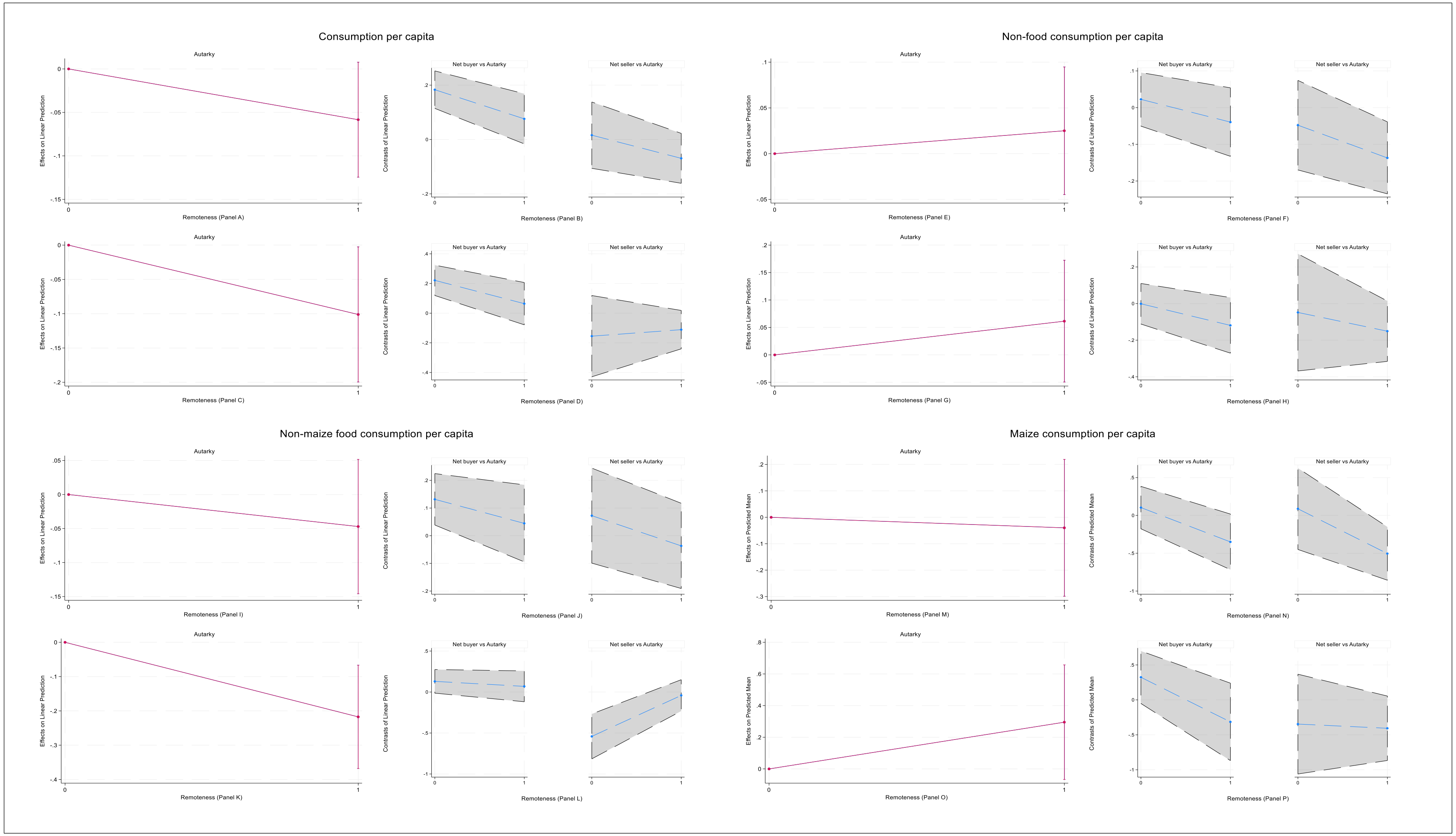

Figure B.10: Impact of the fuel policy reform on consumption that varies with remoteness to the market at 95% confidence interval. Note: Panels for net buyers and net sellers use a common y-axis. Immediate differential effects are in the first row and persistent differential effects are in the second row.

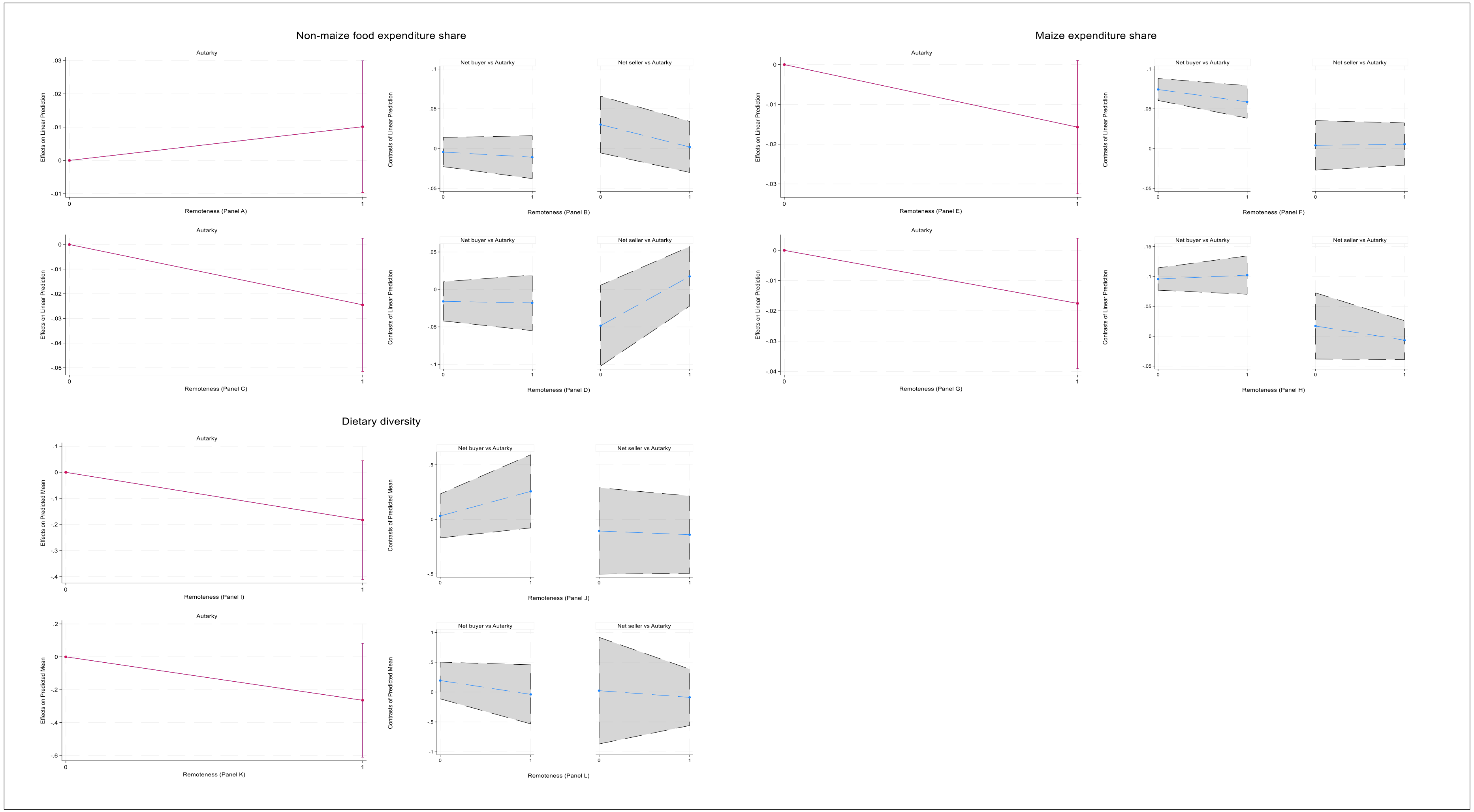

Figure B.11: Impact of the fuel policy reform on consumption share and dietary quality that varies with remoteness to the market at 95% confidence interval.

Note: Panels for net buyers and net sellers use a common y-axis. Immediate differential effects are in the first row and persistent differential effects are in the second row.

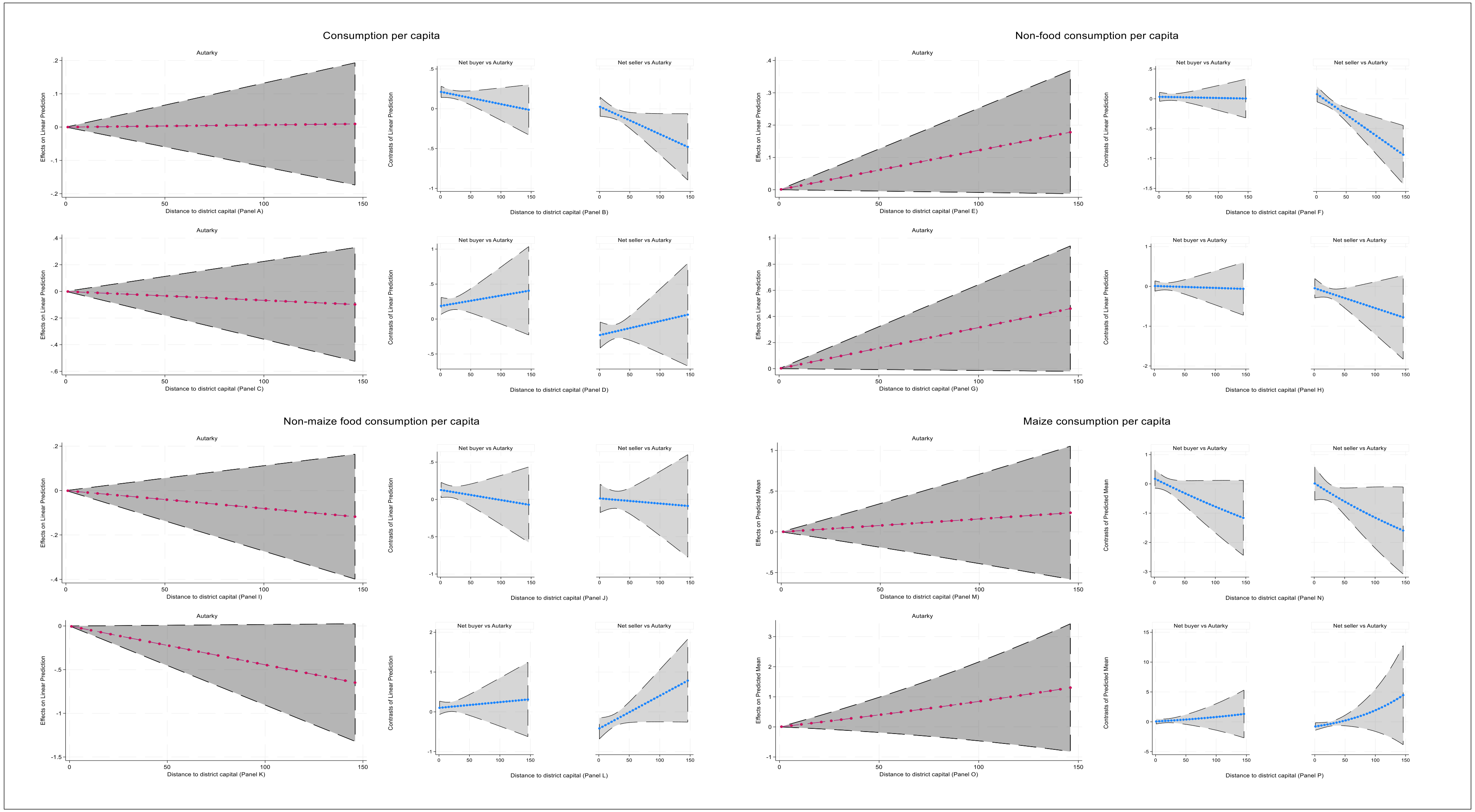

Figure B.12: Impact of the policy reform on consumption related outcomes that varies with distance to the district capital at 95% confidence interval.

Note: Panels for net buyers and net sellers use a common y-axis. Immediate differential effects are in the first row and persistent differential effects are in the second row.

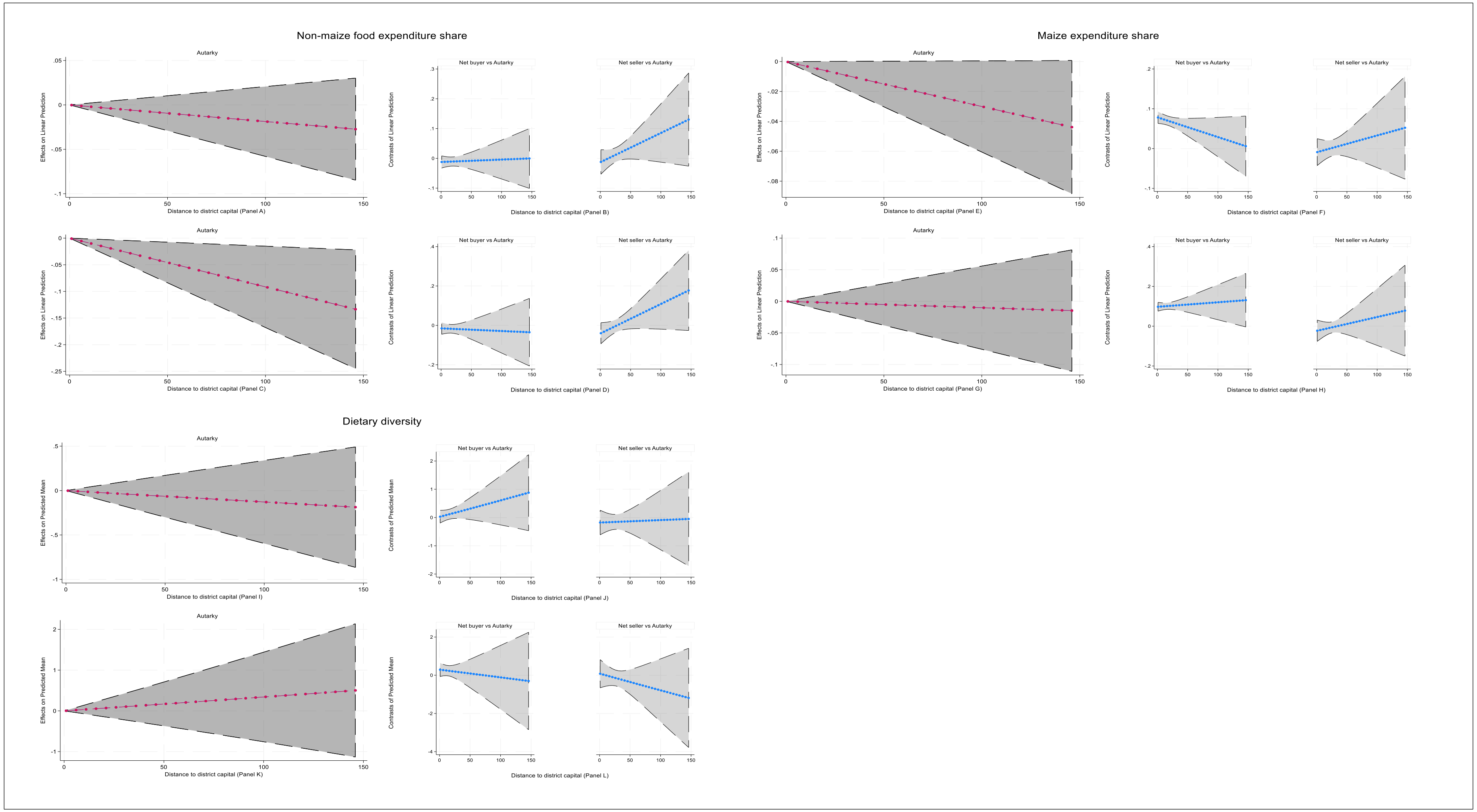

Figure B.13: Impact of the policy reform on consumption shares and dietary quality that varies with distance to the district capital at 95% confidence interval.

Note: Panels for net buyers and net sellers use a common y-axis. Immediate differential effects are in the first row and persistent differential effects are in the second row.

Chapter 4

Do rainfall shocks affect schooling outcomes? Evidence from Malawi

4.1 Introduction

Eliminating gender inequality in schooling remains a challenge in sub-Saharan Africa (SSA) despite an increase in school enrolment among school-aged children (Barro & Lee, 2015). Inequality in schooling arises because households invest differently in girls and boys, where boys' education is more preferred than girls' education because of expected income gains from future employment as grown-ups (Björkman-Nyqvist, 2013; Psaki et al., 2018; Rosenzweig & Schultz, 1982). Although factors such as poverty, inequality, and conflicts worsen schooling outcomes, recent studies have established that economic shocks that are induced by natural disasters such as droughts and floods increase differential treatment of boys and girls in most developing countries where most households depend on income from rain-fed agriculture (Björkman-Nyqvist, 2013; Zimmermann, 2020). As a coping mechanism to weather shocks, households may either reduce expenditures on education or increase child labour differentially among boys and girls, which may lead to differential effects on schooling outcomes among boys and girls (Alvi & Dendir, 2011; Amin et al., 2006b; Baez et al., 2017; Behrman, 1988; Björkman-Nyqvist, 2013).

The objective of this chapter is to investigate how rainfall shocks differentially affect schooling outcomes in both primary and secondary education among boys and girls in Malawi. Rainfall shocks affect household income and consumption through its effects on agricultural production. However, a recent literature has established that rainfall shocks have contrasting effects on agricultural production where the drought shock reduces agricultural production, while the flood shock increases agricultural production (see, for example, Amare et al., 2018; Borgomeo et al., 2018; Damania et al., 2020). Similarly, there are fewer studies that find contrasting effects of drought and flood shocks on schooling outcomes (see, for example, Björkman-Nyqvist, 2013; Zimmermann, 2020). These studies find that a drought shock has a negative effect on schooling outcomes, whereas a flood shock has a positive effect on schooling outcomes. These differential effects arise because the drought shock is considered to be unfavourable for agricultural production, while the flood shock is considered to be beneficial for agricultural production (Björkman-Nyqvist, 2013; Zimmermann, 2020). Consistent with this previous literature, we expect the drought shock to have negative effects on schooling outcomes, while the flood shock to have positive effects on schooling outcomes.

To estimate gender differential effects of rainfall shocks on schooling outcomes, we use three waves of nationally representative panel data from the Integrated Household Panel Survey (IHPS), which were implemented in 2010, 2013, and 2016 as part of the Living Standards Measurement Study-Integrated Surveys of Agriculture (LSMS-ISA), and the school census administrative data from the Ministry of Education in Malawi. We model both the effects of drought and flood shocks. Our measure of the drought shock is a dummy indicator that takes on a value of one if the negative standardised deviation of rainfall from historical mean precipitation in the community is equal to or less than negative one, and zero otherwise, whereas the flood shock is a dummy indicator that takes on a value of one if the positive standardised deviation of rainfall from historical mean precipitation in the community is equal to or greater than positive one, and zero otherwise (Abiona, 2017; Asfaw & Maggio, 2018; Carrillo, 2020; McCarthy et al., 2017; McLaughlin et al., 2023; Nübler et al., 2021). We use the fixed effects estimator to examine how rainfall shocks affect schooling outcomes separately for boys and girls.

Björkman-Nyqvist (2013) and Zimmermann (2020) have shown that the effects of rainfall shocks on schooling outcomes vary with child age. Björkman-Nyqvist (2013) finds that percentage negative deviation in rainfall from the mean reduces primary school enrolment and achievements among older girls but the effects on boys and young girls are not significant in Uganda, while Zimmermann (2020) finds that the drought shock increases school enrolment, whereas the flood shock reduces school enrolment among older girls but the effects on boys and young girls are not significant in India. Consistent with these previous studies, we expect the effects of rainfall shocks on schooling outcomes to vary with child age. Based upon the existing findings in the literature, we anticipate older girls to be affected by rainfall shocks, while boys and young girls to be insulated from the effects of rainfall shocks.

There is a small but growing literature that investigates the effects of weather shocks or changing climate on child schooling outcomes in developing countries. However, empirical evidence on how weather shocks affect child schooling outcomes in SSA is limited (Björkman-Nyqvist, 2013; Randell & Gray, 2016). We add to this literature by examining how rainfall shocks differentially affect schooling outcomes in both primary and secondary schools among boys and girls in Malawi. Björkman-Nyqvist (2013) is the closest study to ours to examine gender differential effects of weather shocks on schooling outcomes. However, this closest study used pooled cross-sectional household survey data and administrative primary school census data aggregated at the district level. We build on Björkman-Nyqvist (2013) and the previous studies to examine how rainfall shocks differentially affect schooling outcomes in both primary and secondary schools among boys and girls using household panel survey data and administrative school-level census data for Malawi.

Our analysis shows that there is differential treatment in children's education whereby households allocate more resources in boys' education during the periods of the flood shock, while resource allocation in girls' education is similar during the periods of the rainfall shock and the normal rainfall. As we expected, we find that the effects of the rainfall shock on school attendance and progression vary with child age. However, the effects of rainfall shocks on school attendance are similar between boys and girls, while the effects on school progression are different among boys and girls. For instance, we find that the drought shock increases school attendance among younger boys and girls in lower primary school, but it reduces school attendance among older boys and girls in secondary school relative to the normal rainfall. Conversely, the flood shock increases school attendance among older boys and girls in upper primary and secondary school relative to a normal rainfall. Moving on to school progression, we find that the drought shock increases school progression among boys and younger girls in lower primary school, while the flood shock increases school progression among older boys in upper primary school and younger boys in lower secondary school, and among girls in secondary school relative to the normal rainfall. Overall, these findings are consistent at the school level.

The rest of the chapter is organised as follows. The next section presents an overview of schooling in Malawi. Section 3 provides an overview of rainfall shocks in Malawi, while section 4 reviews related literature. Section 5 describes the conceptual framework, empirical strategy, and data used in this chapter. Findings are presented in section 6 and section 7 concludes.

4.2 An overview of schooling in Malawi

Malawi has eight years of formal primary and four years of secondary school education, and the recommended starting age is six years (Government of Malawi, 2016b). Thus, a child that enters primary education at the age of six is expected to complete primary school at the age of thirteen years and complete secondary school at the age of eighteen years. Malawi's school calendar is from September through to July. To increase school enrolment, entry into primary education has been free since 1994 in public schools and wearing of school uniform is not enforced (Government of Malawi, 2016b). However, secondary school education is not free (UNICEF, 2018). While private institutions have been allowed to provide both primary and secondary school education since 2004, over 90 percent of the children enrol in public schools (Government of Malawi, 2016b). Although enrolment of children of primary-school age and school attendance are high in primary schools (Government of Malawi, 2020b), late school entry, and high rates of repetition and dropouts characterise the education system in Malawi (Chimombo, 2009; Government of Malawi, 2016b; United Nations Children's Fund-Malawi, 2022). To improve child schooling outcomes in areas that are at risk of food insecurity, the government in collaboration with development partners has been implementing school feeding programmes in the form of breakfast porridge in some primary schools in rural Malawi since 1999 (Manea, 2020; World Food Programme, 2021). However, enrolment in secondary school is very low (UNICEF, 2018)

While limited access to teaching and learning materials, inadequate qualified teachers and school facilities affect learning outcomes, studies have established that sending children to work (*i.e.*, child labour) worsens school learning outcomes (Nankhuni & Findeis, 2004; United Nations Children's Fund-Malawi, 2022; Xia & Deininger, 2019).[66] For instance, Nankhuni & Findeis (2004) find that hours spent in collecting firewood and water reduce the probability of attending school among children aged between six and fourteen years using the first round (*i.e.*, 1997 – 1998) of Malawi's Integrated Household Survey data. Xia & Deininger (2019) find that the increase in the share of tobacco farmers in the village reduces the probability that the child progress to the next grade level using three rounds (*i.e.*, 2004 – 2005, 2010 – 2011, and 2012 – 2013) of Malawi's Integrated Household Survey data. In addition, the United

[66] Parents' may send children to work either when the school is in session, which reduces school attendance or when the school is not in session, which reduces hours of study.

Nations Children's Fund-Malawi (2022) finds children who are involved in child labour have lower reading and numeracy skills compared to those that are not involved in child labour using the sixth round of Malawi's Multiple Indicator Cluster Survey (MICS) data.

4.3 Rainfall shocks in Malawi

Occurrence of natural disasters such as drought and flood shocks impact Malawi's agricultural production leading to crop failure, increased cases of food insecurity, hunger-related diseases, and loss of lives. The main agricultural season is from November through to April and about 90 percent of the land cultivated is under rain-fed agriculture (Government of Malawi, 2016a). Extreme weather events are mostly induced by El Niño and La Niña (Botha et al., 2018; Devereux, 2002). Although moderate natural disasters and seasonal hunger or food shortages are more common across the country, Malawi has experienced severe drought and flood shocks in some agricultural seasons since the 1860s. However, the occurrence of natural disasters is more frequent in post-independence than during the colonial era. During the colonial era, Malawi experienced two extreme weather events in 1861-1863 and 1949. After independence in 1964, Malawi has experienced seven extreme weather events, which occurred in 1991-1992, 2001-2002, 2004-2005, 2014-2015, 2015-2016, 2018-2019, and 2022-2023 agricultural seasons. Extreme weather events are characterised by government declaration of state of emergency and an extensive humanitarian response programme in affected districts.

4.4 Literature review

This section reviews approaches that are often used in the literature to identify locations affected by weather shocks (*i.e.*, a drought or flood shock), and empirical methods used to identify a causal relationship between the weather shock and outcome of interest. Further, the section reviews some of the previous studies that examine effects of weather shocks on schooling outcomes in developing countries.

4.4.1 Approaches to measure weather shock exposure

Usually, household surveys collect information on shocks including weather induced shocks that affect household well-being, and governments conduct rapid assessments in the event of an extreme weather to identify areas that are severely affected. However, the use of this information to identify households or areas that are affected by weather shocks when estimating the impact of the shock on welfare is subject to endogeneity bias (Baquie & Fuje, 2020; G. Nguyen & Nguyen, 2020). This is the case because weather shocks affect many households within the same geographic area or community (*i.e.*, a covariate shock) but the extent of the shock expressed by the household depends on household's coping capability, level of adaptation to the environment, experience with the shock, and risk perception (see, Nguyen & Nguyen, 2020). As a result, recent studies use meteorological station or satellite climatic data to measure community exposure to weather shocks, which is arguably exogenous (Baquie & Fuje, 2020; Dell et al., 2014; G. Nguyen & Nguyen, 2020).[67] This approach requires historical weather data on weather indicators such as temperature and precipitation closer to the location of households if meteorological station data are used or within a geographical or spatial area if satellite climatic data are used.

Several indices have been developed to identify periods of drought and flood shocks using historical weather data. Heim (2002) reviews drought indices that have been developed that allow comparability of the intensity of drought shock across time and spatial area. These drought indices include Palmer's index, crop moisture index, Keetch-Byram drought index, Standardized Precipitation index, vegetation condition index, and drought monitor (see, Heim (2002), for a detailed discussion on how each index is calculated and data requirements). These indices have been applied in empirical studies to identify periods of drought and flood shocks across space and time (see, for example, Aguilar & Vicarelli, 2022; Baquie & Fuje, 2020; Pauw et al., 2011; Salazar-Espinoza et al., 2015). Conversely, other studies use historical weather data to calculate standardised deviations or percentage deviations of current rainfall from historical mean precipitation or temperature to identify periods of weather shocks in a specific growing season within a spatial area (note that this is also known as Rainfall Anomaly Index (RAI) in this literature). One or more standardised deviations above the historical mean precipitation or temperature (or percentage deviations above the 80th percentile) represents the flood (or a positive rainfall) shock and, zero otherwise, while a standardised deviation

[67] Satellite climatic data is sourced from various institutions such as the National Oceanic and Atmospheric Administration (NOAA), the European Centre for Medium Range Weather Forecasts (ECMWF), and the National Aeronautics and Space Administration's (NASA) Modern-Era Retrospective analysis for Research and Applications, version 2 (MERRA-2). Conversely, meteorological station data can be obtained from institutions such as the Climatic Research Unit (CRU) at the University of East Anglia, the Centre for Climatic Research at the University of Delaware, and the department of meteorological services in the specific country of interest.

equal to or less than negative one of historical mean (or percentage deviations below the 20th percentile) represents the drought (or a negative rainfall) shock, and zero otherwise (see, for example, Abiona, 2017; Asfaw & Maggio, 2018; Ba & Mughal, 2022; Carrillo, 2020; Maccini & Yang, 2009; McCarthy et al., 2017; McLaughlin et al., 2023; Nübler et al., 2021; Shah & Steinberg, 2017; Zimmermann, 2020).

4.4.2 Empirical approach

Availability of panel data at the household level allows estimation of the impact of weather shocks on welfare using econometric methods. However, household-level data collected immediately after the occurrence of an extreme weather event are scarce (Baquie & Fuje, 2020; McCarthy et al., 2017). As a result, there are fewer studies that explicitly estimate the impact of a single extreme weather (*i.e.*, the drought or flood shock) event on household welfare outcomes (see, for example, Aguilar & Vicarelli, 2022; Alvi & Dendir, 2011; Baez et al., 2017; Rosales-Rueda, 2018). Instead, most studies estimate the impact of continuous variation in weather indicators (*i.e.*, precipitation and temperature) or moderate weather shocks on household welfare outcomes over the period the household-level data are available (Abiona, 2017; Asfaw & Maggio, 2018; Baquie & Fuje, 2020; Dell et al., 2014; McLaughlin et al., 2023). The fixed effects estimator is the most widely used technique to identify the impact of weather shocks on household welfare outcomes because location-specific fixed effects (*i.e.*, a spatial area where a shock is identified) absorb geographic common attributes that are time-invariant, whereas time-specific fixed effects absorb time-varying common shocks (Dell et al., 2014). The next section reviews papers that have used both the fixed effects estimator and quasi-experimental approach to investigate the effect of weather shocks on schooling outcomes in developing countries.

4.4.3 Measures of climatic shocks and child schooling

There is a large literature that examines the effects of climatic or weather shocks on child labour supply in developing countries (see, for example, Alvi & Dendir, 2011; Boutin, 2014; Dumas, 2020; Weidinger, 2021). This sub-section reviews some of the previous studies that investigate the effects of weather shocks on child schooling outcomes using various measures of rainfall shocks, which is the focus of this chapter.

4.4.3.1 Studies that find negative effects of rainfall shocks on schooling outcomes

Some of the studies in this literature find negative effects of rainfall shocks on schooling related outcomes using various measures of rainfall shocks. For instance, Shah & Steinberg (2017) use dummy indicators constructed as deviations below the 20th percentile of the norm for the drought shock, and zero otherwise, and deviations above the 80th percentile of the norm for the flood shock, and zero otherwise as measures of rainfall shocks to show that the flood shock reduces scores in math, school attendance, and enrolment but it increases child labour in rural India, while Nordman et al. (2022) use percentage deviation of rainfall from the long-term average rainfall as the measure of rainfall shock to show that percentage increase in rainfall deviations reduces expenditure on school fees in rural India. Further, Baez et al. (2017) use a dummy indicator constructed as two or more standardised deviations above the long-term average rainfall, and zero otherwise as the measure of the flood shock to show that the 2010 tropical storm that led to excessive rainfall reduced child school attendance among households that were affected by the rainfall shock relative to unaffected households in Guatemala.

4.4.3.2 Studies that find positive effects of rainfall shocks on schooling outcomes

There are other studies that find positive effects of rainfall shocks on schooling related outcomes using various measures of rainfall shocks. Björkman-Nyqvist (2013) use percentage deviations of current rainfall from the historical mean rainfall as the measure of rainfall shock to show that a percentage increase in deviation of rainfall increases enrolment of older girls and their school performance when school is free but has no effects on boys and young girls in Uganda, while Randell & Gray (2016) use standardised deviation of rainfall as the measure of rainfall shock to show that the increase in standardised deviation of rainfall and mild temperatures increase the probability of children to complete the grade and attend school in rural Ethiopia.

4.4.3.3 Studies that find contrasting effects of rainfall shocks on schooling outcomes

There are fewer studies that find contrasting effects of rainfall shocks on schooling outcomes. For example, Zimmermann (2020) uses the same measures of rainfall shocks as in Shah &

Steinberg (2017) to show that rainfall shocks have long-term effects on schooling outcomes where the drought shock increases school enrolment, while the flood shock reduces school enrolment over time in India. As part of additional analysis, Björkman-Nyqvist (2013) uses alternative measures of rainfall shocks constructed as percentage negative deviations of current rainfall from the historical mean rainfall for the drought shock and percentage positive deviations of current rainfall from the historical mean rainfall for the flood shock to show that the direction of the effect of percentage negative deviation in rainfall on enrolment of older girls is negative and significant, while the direction of the effect of percentage positive deviation in rainfall on enrolment of older girls is positive but it is not significant in Uganda.

4.4.3.4 Heterogeneity of the effects of rainfall shocks on schooling outcomes

Child gender and age are the main characteristics through which heterogenous effects of rainfall shocks on schooling outcomes are explored. Björkman-Nyqvist (2013) finds that negative deviation in rainfall from the mean reduces primary school enrolment and achievements among older girls but the effects on boys and young girls are not significant in Uganda, while Zimmermann (2020) finds that the drought shock increases school enrolment, whereas the flood shock reduces school enrolment among older girls but the effects on boys and young girls are not significant in India.

In conclusion, these previous studies have established that rainfall shocks using various measures do affect schooling outcomes in developing countries. However, the major limitation of these previous studies relates to the quality of the data that were used. At the household level, Björkman-Nyqvist (2013), Zimmermann (2020), and Baez et al. (2017) used pooled cross-sectional household survey data, while Randell & Gray (2016) and Nordman et al. (2022) used household panel survey data from 15 villages from rural Ethiopia and nationally representative data from rural India, respectively. Further, Shah & Steinberg (2017) used census data on children from the Annual Status of Education Report survey data in rural India. The estimates from Randell & Gray (2016) and Nordman et al. (2022) are more efficient than those from Björkman-Nyqvist (2013), Zimmermann (2020), and Baez et al. (2017) because the authors were able to control for characteristics that are either observed or unobserved that would have biased their estimates (*i.e.*, the authors in the panel case were able to control for individual heterogeneity in the unit of observation compared to the repeated cross-sectional case). Conversely, the estimates from Nordman et al. (2022) are more nationally presentative and can be generalised to other areas without further research than those from Randell & Gray (2016), while those from Shah & Steinberg (2017) are nationally representative but they are less efficient. Turning to the school level, Björkman-Nyqvist (2013) is the only study to use administrative primary school census data that were aggregated at the district level. However, aggregated data suffers from aggregation bias, which may lead to loss of details and misleading conclusions relating to the effects of rainfall shocks at the school level (Cherry & List, 2002; Garrett, 2002). This chapter advances this literature by using household panel survey data and administrative school-level census data to examine how rainfall shocks differentially affect schooling outcomes among boys and girls in both rural and urban areas using the fixed effects estimator to establish the causal impact of rainfall shocks in Malawi. Further, we consider schooling outcomes among boys and girls in both primary and secondary schools as in Zimmermann (2020).

4.5 Methods

4.5.1 Conceptual framework

Households often experience income shocks, which can be household-specific (*i.e.*, idiosyncratic shock) such as a death of a household member, unemployment, and theft of assets or community-specific (*i.e.*, covariate shock) such as conflict and rainfall (*i.e.*, drought or flood) shocks. In the case of rainfall shocks, evidence shows that the drought and flood shocks impact agricultural productivity, which affects household income, food consumption, and diet quality (Aguilar & Vicarelli, 2022; Baez et al., 2017; Baquie & Fuje, 2020; McCarthy et al., 2017; McLaughlin et al., 2023; Rosales-Rueda, 2018). However, other studies have found that droughts and floods have contrasting effects on agricultural production where the drought shock reduces agricultural production, while the flood shock increases agricultural production (Amare et al., 2018; Borgomeo et al., 2018; Damania et al., 2020). To off-set negative effects of rainfall shocks, households may use savings or insurance, sell assets, access credit, increase off-farm labour supply or reduce consumption of food and non-food items such as housing, education, and health. However, the prevalence of poverty is high, and insurance and credit markets are limited in most developing countries (Kochar, 1999; Rose, 2001). Further, households may want to protect their assets because of the need to revive own-farm production after the shock (Jodha, 1978).

Evidence shows that relocation of labour between the farm and off-farm employment opportunities and reducing consumption are the most common livelihood strategies that most households use to cope with rainfall shocks in most developing countries (Ansah et al., 2020; Branco & Féres, 2021; Ito & Kurosaki, 2009; Jessoe et al., 2016; Jodha, 1978; Kochar, 1999; Mishra & Goodwin, 1997; T. Nguyen et al., 2020; Nikoloski et al., 2018; Rose, 2001; Saha, 1994; World Bank Group et al., 2016). However, the demand for agricultural labour reduces during the periods of the drought shock compared to the normal agricultural season (Sesmero et al., 2018; Zimmermann, 2020). When income and consumption fall below subsistence levels, households may either allow their children to work at home (*i.e.*, unpaid work) to reduce expenditure on hired labour or free up adults to increase their participation in the labour market (*i.e.*, substitution effect) or allow their children to work in the labour market (*i.e.*, paid work) to augment household income (*i.e.*, complementary effect) (Alvi & Dendir, 2011; Amin et al., 2006a, 2006b; Baez et al., 2017; Edmonds, 2007; Orazem & Gunnarsson, 2003; Ray, 2000a, 2000b). If households reduce consumption or increase child labour supply differentially among boys and girls because of differential expectations of returns to schooling, then rainfall shocks may have differential effects on schooling outcomes among boys and girls (Alvi & Dendir, 2011; Amin et al., 2006b; Baez et al., 2017; Behrman, 1988; Björkman-Nyqvist, 2013).

Although there is no consensus on the effects of rainfall shocks on schooling outcomes across the regions, Zimmermann (2020) discusses two channels through which the drought shock would affect child schooling outcomes, namely credit constraint and opportunity costs channels. The credit constraint channel occurs when the drought shock reduces school enrolment because households may not have adequate financial resources to keep their children in school (Zimmermann, 2020). For instance, Björkman-Nyqvist (2013) finds that the percentage negative deviation in rainfall from the mean reduces enrolment of older girls in Uganda, while Randell & Gray (2016) show that the increase in standardised deviation of rainfall and mild temperatures increases the probability of children to complete a grade and attend school in rural Ethiopia. Conversely, the opportunity cost channel occurs when the drought shock increases school enrolment because of limited employment opportunities compared to the normal rainfall in farm work (Zimmermann, 2020), which employs a larger proportion of children than service and industrial sectors (International Labour Office and United Nations Children's Fund, 2021; United States Department of Labor, 2022). For instance, Baez et al. (2017) find that the 2010 tropical storm that led to floods reduced the probability of children to attend school in affected households relative to unaffected households in Guatemala, while Shah & Steinberg (2017) find that the flood shock reduces scores in math, school attendance, enrolment, but it increases child labour in India. In a related study, Zimmermann (2020) shows that the drought shock increases school enrolment, while the flood shock reduces school enrolment more in older children and among girls than in younger children and among boys over time in India. Further, Nordman et al. (2022) show that increase in rainfall deviations reduces investment in children education in India. Zimmermann (2020) concludes that the opportunity costs channel dominates as the country develops than credit constraint channel.

Given that agriculture remains the main source of income for most households and access to credit is limited in Malawi (Diagne, 1998; Government of Malawi, 2016a; National Statistical Office, 2013b), we expect the credit constraint channel to dominate. Consistent with previous literature we are anticipating contrasting effects of drought and flood shocks on schooling outcomes, where the drought shock is expected to have negative effects on schooling outcomes, while the flood shock is expected to have positive effects on schooling outcomes. Further, given that it has been established that older girls are affected by rainfall shocks, while boys and young girls are shielded from the effects of rainfall shocks (Björkman-Nyqvist, 2013; Zimmermann, 2020), we expect the effects of rainfall shocks on schooling outcomes to vary with child age.

Ultimately, the extent to which rainfall shocks differentially affect child schooling is an empirical question that we describe and estimate in the following section.

4.5.2 Empirical strategy

4.5.2.1 Child-level impacts

This part of the analysis uses household level survey data. Before examining how the effects of rainfall shocks vary with child age (*i.e.*, age-specific effects), we first test whether the effects of rainfall shocks on schooling related outcomes are different between boys and girls. To identify average effects of rainfall shocks on schooling related outcomes for child i in Traditional Authority (TA) w at time t, we estimate the following equation:

$$C_{iwt} = \beta_0 + \beta_1 D_{wt} + \beta_2 F_{\omega t} + \beta_3 X_{hwt} + \theta_t + \pi_i + \varepsilon_{iwt} \qquad (4.1)$$

where C represents schooling related outcomes of interest for child i in TA ω at time t. Our schooling related outcomes of interest are: (i) a dummy indicator equal to one if the child attends school during the current session or attended school during the completed session and is planning to attend the next session if school is not in session, and zero otherwise, (ii) a dummy indicator equal to one if the child changed the grade level from the previous academic year, and zero otherwise, and (iii) share of expenditure on education (the next section describes in more detail how each schooling outcome measure was constructed).[68]

D_{wt} is a dummy variable that takes on a value of 1 if the TA experienced the drought shock in the current growing season of the survey wave, and 0 otherwise. F_{wt} is a dummy variable that takes on a value of 1 if the TA experienced the flood shock in the current growing season of the survey wave, and 0 otherwise. θ_t represents year dummies (*i.e.*, time-specific fixed effects) that capture common shocks to TAs such as economic conditions and conflict, and π_i represents child-specific fixed effects that capture systematic differences across TAs such as weather patterns and other characteristics. X_{hwt} is a vector of time varying household characteristics such as household head's age and years of education, household size, and value of assets (see table C.1 in the appendix for details). The parameter β_1 provides an estimate of the impact of the drought shock relative to a normal growing season, while the parameter β_2 provides an estimate of the impact of the flood shock relative to a normal growing season on schooling related outcomes. ε_{iwt} represents a child random error term, and β_0 and β_3 are additional parameters to estimate. The underlying assumption for identification of a causal impact of rainfall shocks on schooling related outcomes is that drought and flood shocks are not correlated with the error term. Thus, there are no other factors that affect drought and flood shocks that affect schooling related outcomes. Consistent with the previous studies (Björkman-Nyqvist, 2013; Randell & Gray, 2016), we expect the drought and flood shocks to

[68] The share of expenditure on education is estimated at the household level and we controlled for household-specific fixed effects.

have negative effects on schooling outcomes ($\beta_1 \text{ and } \beta_2 < 0$). We estimate equation [4.1] separately for boys and girls, and cluster the standard errors at the TA level.

Heterogenous effects by child age

To examine how the effects of rainfall shocks on schooling related outcomes vary with child age, we interact our rainfall shock indicators with child age in our baseline specification (*i.e.*, equation [4.1]) as follows:

$$C_{iwt} = \alpha_0 + \alpha_1 G_{iwt} + \alpha_2 D_{wt} + \alpha_3 (D_{wt} * G_{iwt}) + \alpha_4 F_{\omega t} + \alpha_5 (F_{\omega t} * G_{iwt}) + \alpha_6 X_{hwt} + \pi_i + \epsilon_{iwt} \qquad (4.2)$$

where G_{iwt} represents child's age in years. The parameter α_3 provides an estimate of differential effect of age on the impact of the drought shock on schooling outcomes for older children relative to younger children, while the parameter α_5 provides an estimate of differential effect of age on the impact of the flood shock on schooling outcomes for older children relative to younger children. A combination of $\alpha_2 + \alpha_3 G_{iwt}$ captures how the differential effect of the drought shock on schooling outcomes varies with child age, while $\alpha_4 + \alpha_5 G_{iwt}$ captures how the differential effect of the flood shock on schooling outcomes varies with child age. We plot how the effects of rainfall shocks on schooling outcomes vary with child age. In equation [4.2], we do not include time-specific fixed effects, θ_t, because child age in years, G_{iwt}, is collinear with the time-specific fixed effects. Thus, child age in years is effectively the same as the time-specific fixed effects. ϵ_{iwt} represents child random error term, and α_0, α_1 through to α_6 are parameters to estimate. The rest of the variables are the same as those in equation [4.1]. In accordance with the previous studies (Björkman-Nyqvist, 2013; Zimmermann, 2020), we expect the effects of the drought and flood shocks on schooling outcomes to be more negative for older children than younger children given that primary education is free compared to secondary education ($\alpha_2 + \alpha_3 G_{iwt}$ and $\alpha_4 + \alpha_5 G_{iwt} < 0$). Similarly, we estimate equation [4.2] separately for boys and girls, and cluster the standard errors at the TA level.

4.5.2.2 School-level impacts

This part of the analysis uses school census data to examine whether the findings from the household level data are consistent with those at the school level. Similarly, we first test whether the effects of rainfall shocks on schooling related outcomes are different between boys and girls before examining how the effects of rainfall shocks vary with school grade. To identify the impacts of rainfall shocks on schooling related outcomes for school s and grade g (*i.e.*, grade-specific, or age-specific effects) in Traditional Authority w at time t, we estimate the following equation:

$$C_{sgwt} = \gamma_0 + \gamma_1 D_{wt} + \gamma_2 F_{\omega t} + \emptyset_t + \sigma_{sg} + \mu_{sgwt} \qquad (4.3)$$

where C represents: (i) enrolment; (ii) the dropout rate, and; (iii) the repetition rate for school, s for grade g in TA w at time t. D_{wt} is a dummy variable that takes on a value of 1 if the TA experienced the drought shock in the current first term of the school calendar, and 0 otherwise. F_{wt} is a dummy variable that takes on a value of 1 if the TA experienced the flood shock in the current first term of the school calendar, and 0 otherwise. $\emptyset_t$ represents year dummies (*i.e.*, time-specific fixed effects) that capture common shocks to TAs such as economic conditions and conflict, and σ_{sg} represents school-grade specific fixed effects that capture systematic differences across TAs such as weather patterns and other characteristics. The parameter γ_1 provides an estimate of the impact of the drought shock, while the parameter γ_2 provides an estimate of the impact of the flood shock on schooling related outcomes. μ_{sgwt} represents a school random error term, and γ_0 is an additional parameter to estimate. Similarly, the underlying assumption for identification of a causal impact of rainfall shocks on schooling outcomes, is that γ_1 and γ_2 are not correlated with the error term. We estimate equation [4.3] separately for boys and girls, and cluster the standard errors at the TA level.[69]

[69] Not that there are there are no control variables at the school level in equation 4.3.

Heterogenous effects by school grade

To examine how the effects of rainfall shocks on schooling related outcomes vary with school grade, we interact our rainfall shock indicators with school grade in our baseline specification (*i.e.*, equation [4.3]) as follows:

$$C_{sgwt} = \delta_0 + \delta_1 D_{wt} + \delta_2 (D_{wt} * H_{sg}) + \delta_3 F_{\omega t} + \delta_4 (F_{\omega t} * H_{sg}) + \emptyset_t + \sigma_{sg} + \vartheta_{sgwt} \qquad (4.4)$$

where H_{sg} represents school grade. The parameter δ_2 provides an estimate of differential effect of school grade on the impact of the drought shock on schooling outcomes for children in upper grades relative to lower grades, while the parameter δ_4 provides an estimate of differential effect of school grade on the impact of the flood shock on schooling outcomes for children in upper grades relative to lower grades. A combination of $\delta_1 + \delta_2 H_{sg}$ captures how the differential effect of the drought shock on schooling outcomes varies with school grade, while $\delta_3 + \delta_4 H_{sg}$ captures how the differential effect of the flood shock on schooling outcomes varies with school grade. Similarly, we plot how the effects of rainfall shocks on schooling outcomes vary with school grade. In equation [4.4], we do not include H_{sg} because it is collinear with the school-grade fixed effects. Hence, it will be absorbed by school-grade fixed effects since it does not vary across schools. ϑ_{sgwt} represents school grade random error term, and δ_0, δ_1 and δ_3 are parameters to estimate. The rest of the variables are the same as those in equation [4.3].

Functional form

The next step is to determine the functional form and find an estimator for equations [4.1 – 4.2] for household data and equation [4.3 – 4.4] for school data where our schooling related outcomes, dependent variables, take different range of values. School attendance and progression are discrete variables, whereas share of expenditure on education is restricted between a zero and a one. Turning to school level data, school enrolment is a continuous variable, whereas the dropout and repetition rates are substantively restricted between a zero and a one. We need to include the time-specific fixed effects, θ_t in equation [4.1 – 4.2] (or $\emptyset_t$ in equation [4.3 – 4.4), and child-specific fixed effects, π_i in equation [4.1 – 4.2] (or school-grade specific fixed effects, σ_{sg} in equation [4.3 – 4.4]), for the impact of rainfall shocks to be identified. Therefore, we use a linear fixed effects estimator to examine the impact of rainfall shocks on schooling related outcomes that are discrete, take continuous values, and are substantively restricted between a zero and a one.

4.5.3 Data

4.5.3.1 Household data

We use panel data from the three waves of the nationally representative Integrated Household Panel Survey (IHPS) implemented in 2010, 2013, and 2016 as part of the Living Standards Measurement Study-Integrated Surveys of Agriculture (LSMS-ISA) for Malawi. The IHPS tracks a stratified two-stage random sub-sample of 3,246 households that were part of the third Integrated Household Survey (IHS3) in 2010/2011 across 204 enumeration areas (EAs) (*i.e.*, communities) in 27 districts and 4 cities (*i.e.*, Lilongwe, Blantyre, Mzuzu, and Zomba Municipality). This indicates that each community has about 16 households (=3,246/204) in the dataset, on average. IHS3 cross-sectional full survey was conducted from March 2010 through to March 2011 across 768 communities from a sample of 12,271 (National Statistical Office, 2014a). In panel surveys, individuals that branched off from the original household and formed a new household were included into the IHPS sample, which indicates that the sample increases over time (National Statistical Office, 2014b). In 2012/13 panel survey, about 4,000 households were successfully tracked of which 896 households (*i.e.*, 23.2 percent) branched off into 2 or more households whereas the remaining 3,104 baseline households did not branch off (National Statistical Office, 2014b).

In subsequent panel surveys after 2013, the number of communities to be tracked was reduced from 204 to 102 EAs due to financial and resource constraints (National Statistical Office of Malawi, 2017). In this chapter, we use community level data from 102 EAs that are in 90 Traditional Authorities ($w = 1,2, \ldots ,90$ in equation 1) with 1,619 original households from the 2010 baseline year. In subsequent surveys, 1,555 and 1,524 original households were successfully tracked in 2012/13 and 2015/16, respectively. We exclude splitting-off households from the baseline households that went on to form new households over time. The fieldwork for the 2012/13 survey took place from April to December 2013 (National Statistical Office, 2014a), while for the 2015/16 survey fieldwork took place from April 2016 to April 2017 (National Statistical Office, 2017).

To better understand the impacts of rainfall shocks on schooling outcomes, we consider children aged between six (*i.e.*, Malawi's school starting age) and eighteen years. Our sample is unbalanced panel data initially comprising 8,547 children aged between six and eighteen years from the three survey waves (*i.e.*, 2,511 in 2010; 2,976 in 2013; and 3,060 in 2016). As part of data cleaning, we dropped eleven observations with incomplete information (Alvi & Dendir, 2011; Xia & Deininger, 2019). Our final sample is unbalanced panel data with 8,536 children ($i = 1,2, \dots, 8,536$ in equation 1) from the three survey waves (*i.e.*, 2,507 in 2010; 2,973 in 2013; and 3,056 in 2016).[70]

Construction of schooling indicators

The household survey collects information from each household member aged five years and older on schooling outcomes that includes, but is not limited to, school attendance, grade level during the current academic year, and the grade level during the previous academic year. We use this information to create two dummy variables to examine the effects of rainfall shocks on schooling. The first dummy variable takes on a value of one if the child attends school during the current session or attended school during the completed session and is planning to attend the next session if school is not in session, and zero otherwise. The second dummy variable takes on a value of one if the child changed the grade level from the previous academic year, and zero otherwise (Xia & Deininger, 2019). Further, the household survey collects information on household school expenditures on each child on items such as books, tuition, and school uniform during the previous 12 months prior to the survey. We aggregate these expenditures and construct the share of expenditure on education as a proportion of annual expenditure on education to the total annual expenditures for each household by gender.[71]

Distribution of the data

We find that about 91 percent of the children attend school, which indicates that school attendance is high in Malawi. Further, we find that about 66 percent of the children change the grade level from the previous academic year. This may suggest that school progression is

[70] The increase in the number of children simply means that there were more children that became older than 5 years but less than 18 years in that wave relative to the previous wave.
[71] Expenditures are adjusted for inflation using consumer price index (CPI) from the Reserve Bank of Malawi. Annual expenditures are constructed as in chapter 2.

considerably high among children in Malawi. The distribution of the share of expenditure on education indicated presence of outliers. Therefore, we drop the top 5 percent of the share expenditure on education for boys and girls.[72] Figure C.1 in the appendix shows that the share of expenditure on education is similar between boys and girls.

4.5.3.2 School administrative data

Malawi's Ministry of Education carries out an Annual School Census at the beginning of each school calendar (*i.e.*, first-term) across the country. The Ministry of Education has an Education Management Information System (EMIS) for storing the data. The data are used to produce the annual Malawi Education Statistics Report. We obtained enrolment, dropout, and repetition data by gender and grade from the Ministry of Education from 2010 through to 2016. The data has about 6,186 primary schools (*i.e.*, both public and private), and 1,027 secondary schools (*i.e.*, both public and private) across the country.

Construction of school indicators

Enrolment is the number of children that register at a school during the first term of each school calendar. In accordance with the Ministry of Education (Government of Malawi, 2022), we construct the dropout rate as follows:

$$dropout\ rate_{sgt} = \frac{dropout_{sgt}}{enrolment_{sgt-1}} \qquad (4.5)$$

where $dropout\ rate_{sgt}$ is the dropout rate in school s for grade g at time t. $dropout_{sgt}$ is the number of children that drop out of school s for grade g at time t whereas $enrolment_{sgt-1}$ is the previous enrolment in school s for grade g at time t.

Similarly, we construct the repetition rate as follows:

$$repetition\ rate_{sgt} = \frac{repetition_{sgt}}{enrolment_{sgt-1}} \qquad (4.6)$$

[72] Although the results are not shown, we find that the finding on the share of expenditure on education among boys is sensitive to the presence of outliers using the full sample.

where $repetition\,rate_{sgt}$ is the repetition rate in school s for grade g at time t. $repetition_{sgt}$ is the number of students that repeat in school s for grade g at time t whereas $enrolment_{sgt-1}$ is the previous enrolment in school s for grade g at time t.

Distribution of the data

Figure C.2 in the appendix provides the distribution of schooling outcomes at the school level. The distribution of the data indicated the presence of outliers in school enrolment, dropout rate, and repetition rate. Therefore, we drop the top 5 percent of values of enrolment, the dropout rate, and the repetition rate in both primary and secondary schools.[73]

4.5.3.3 Climatic data

We obtain monthly precipitation data from the National Aeronautics and Space Administration (NASA) Langley Research Center (LaRC) Prediction of Worldwide Energy Resource (POWER) Project funded through the NASA Earth Science/Applied Science Program.[74] The climatic data are available from 1981 to 2024. According to Funk et al. (2014), the dataset are available on a 0.05^0 latitude by 0.05^0 longitude grid, equivalent to 4.8 km by 4.8 km cells. Then, we obtain shapefiles for Malawi's Subnational Administrative Boundaries from the Humanitarian Data Exchange.[75] The shapefiles are available at the district and Traditional Authority levels. The Traditional Authority shapefiles indicate that Malawi is divided into 433 Traditional Authorities. We use the Traditional Authority shapefiles to obtain monthly precipitation data. Thus, our rainfall shocks are identified at the Traditional Authority level.

Matching household data with climate data

Geographical coordinates for each community (*i.e.*, enumeration area) are published along with the IHPS data. We successfully matched 100 of the 102 IHPS communities to 89

[73] Although the results are not shown, our findings at the school level are not sensitive to the presence of outliers using the full sample.
[74] https://power.larc.nasa.gov/data-access-viewer/
[75] https://data.humdata.org/dataset/cod-ab-mwi

Traditional Authorities from the shapefiles.[76] This suggests that some communities are in the same TA.

Matching school data with climate data

The Ministry of Education also collects geographical coordinates for each school location, which are made available upon request. We use the geographical coordinates to match each school location to Traditional Authorities from the shapefiles. Primary schools are matched to 400 Traditional Authorities, while secondary schools are matched to 328 Traditional Authorities from the shapefiles.

Measuring exposure to rainfall shocks

We add our monthly precipitation data covering the period from January 1981 through to December 2016 (*i.e.*, 35 years) to obtain yearly precipitation for each Traditional Authority (Björkman-Nyqvist, 2013; Carpena, 2019; Zimmermann, 2020). To identify rainfall shocks in each Traditional Authority during the growing season (*i.e.*, November through to April), we follow the economic literature that use standardised deviations of annual precipitation from the historical average precipitation (Abiona, 2017; Asfaw & Maggio, 2018; Carrillo, 2020; McCarthy et al., 2017; McLaughlin et al., 2023; Nübler et al., 2021). The standardised deviation of precipitation during the growing season for Traditional Authority w at time t is calculated as follows:

$$SD_{wt} = \frac{Prc_{wt} - \overline{Prc_w}}{\delta_w} \qquad (4.7)$$

where SD_w represents the standardised deviation of precipitation during the growing season in Traditional Authority w at time t, Prc_{wt} is the current precipitation during the growing season in each Traditional Authority, $\overline{Prc_w}$ represents historical or long-term average precipitation during the growing season for each Traditional Authority, and δ_w is standard deviation of long-term precipitation during the growing season for each Traditional Authority. The SD takes on both negative and positive values, where positive values represent wet

[76] We believe this brings less noise in the analysis given that geographical coordinates for each community that are published together with IHPS data are offset or modified to preserve the confidentiality of sampled households and communities (National Statistical Office, 2014b).

events and negative values represent dry events.[77] Thus, larger positive SD values signal the flood shock while smaller negative SD values signal the drought shock. We construct dummy indicators for both negative as well as positive rainfall shocks. The dummy indicator for the drought shock takes on a value of one if the SD_{wt} at time t in the Traditional Authority is equal to or less than minus 1 (*i.e.*, $SD_{wt} \leq -1$), and zero otherwise, while the dummy indicator for the flood shock takes on a value of one if the SD_{wt} at time t in the Traditional Authority is equal to or greater than positive 1 (*i.e.*, $SD_{wt} \geq 1$), and zero otherwise.

Table 4.1 presents summary statistics of rainfall variables used in the analysis by survey wave. The table shows that about 49 percent of the communities experienced the drought shock during the 2015/2016 growing season (Botha et al., 2018; Government of Malawi, 2016c), while 25 percent experienced the flood shock during the 2009/10 growing season.

Table 4.1: Summary statistics of rainfall variables

Variable	**2009/10 growing season**	**2012/13 growing season**	**2015/16 growing season**
Drought	0.010	0.000	0.490
	(0.100)	(0.000)	(0.502)
Floods	0.250	0.060	0.020
	(0.435)	(0.239)	(0.141)
Observations	100	100	100

Note: Numbers shown are averages and their corresponding standard deviations are presented in parenthesis. 100 communities are matched to 89 TAs, indicating that some communities are in the same TA.

Robustness

As a robustness check, we re-estimate our models using alternative measures of rainfall shocks. Firstly, we experiment with a cut-off of 0.5 (*i.e.*, moderate shock), where the drought shock is a dummy indicator that takes on a value of one if the negative standardised deviation of rainfall from historical mean precipitation in the community is equal to or less than minus 0.5, and zero otherwise, whereas the flood shock is a dummy indicator that takes on a value of one if the positive standardised deviation of rainfall from historical mean precipitation in the community is equal to or greater than plus 0.5, and zero otherwise (Salazar-Espinoza et

[77] One of the limitations of using precipitation data alone to identify flooded areas across space and time is that there are other parameters such as slope, distance from a waterbody, elevation, land use, and land cover that needs to be considered to identify areas that are affected by the flood shock (Kaya & Derin, 2023; Kazakis et al., 2015; Pratik & Sar, 2020; Salazar-Espinoza et al., 2015). Hence, our measure of the flood shock is relatively basic.

al., 2015).[78] Secondly, we explicitly use the standardised deviation of rainfall as a continuous measure of the rainfall shock (Björkman-Nyqvist, 2013; Maccini & Yang, 2009; Nordman et al., 2022)

4.5.3.4 Descriptive statistics

Household data

Table 4.2 presents summary statistics of variables used in the analysis by survey wave. We use a multivariate test of means procedure that Moore (1998) developed to determine whether the means are different across the waves. Focusing initially upon schooling dependent variables adopted in the analysis, table 4.2 reveals that the differences in means are significant for school progression and share of expenditure on education at 95% significance level. This means that school progression has declined over time, while the share of expenditure on education has increased over time among boys and girls, on average.

[78] We are unable to experiment with the cut-off of 1.5 (*i.e.*, extreme shock) because of little variation in the standardised deviation of rainfall above this point.

Table 4.2: Characteristics of household data by survey wave

Variables	Wave 1	Wave 2	Wave 3	Mean Diff. (Wald chi (2) statistic)
Child schooling outcomes				
Child school attendance	0.909	0.916	0.904	2.73
	(0.287)	(0.277)	(0.295)	
Child changed grade level	0.742	0.633	0.619	92.95***
	(0.438)	(0.482)	(0.486)	
Share of expenditure on education for boys	0.002	0.003	0.005	326.60***
	(0.003)	(0.004)	(0.006)	
Share of expenditure on education for girls	0.003	0.004	0.006	446.17**
	(0.004)	(0.005)	(0.008)	
Independent variables				
# of males greater than 65 years old	0.069	0.063	0.121	57.43***
	(0.254)	(0.244)	(0.359)	
# of male adults 12-64 years old	1.744	1.908	1.970	54.10***
	(1.146)	(1.155)	(1.206)	
# of females greater than 65 years old	0.080	0.097	0.161	76.90***
	(0.275)	(0.300)	(0.419)	
# of female adults 12-64 years old	1.848	1.964	2.072	56.09***
	(1.082)	(1.131)	(1.143)	
# of children less than 12 years old	2.646	2.748	2.482	47.12***
	(1.638)	(1.588)	(1.459)	
Marital status of head, =1 if has a spouse	0.790	0.790	0.782	0.63
	(0.408)	(0.408)	(0.413)	
Gender of household head, =1 if male	0.783	0.781	0.772	1.11
	(0.413)	(0.414)	(0.420)	
Age of head in years	44.206	45.247	46.636	49.37***
	(13.464)	(12.817)	(12.546)	
Highest qualification of household head, none	0.723	0.712	0.911	592.61***
	(0.448)	(0.453)	(0.285)	
Highest qualification of household head, primary	0.094	0.094	0.041	98.07***
	(0.292)	(0.291)	(0.198)	
Highest qualification of household head, secondary	0.161	0.164	0.044	361.69***
	(0.367)	(0.370)	(0.206)	
Highest qualification of household head, tertiary	0.022	0.031	0.004	88.18***
	(0.148)	(0.173)	(0.063)	
if HH was on any social safety net program	0.253	0.472	0.413	325.66***
	(0.435)	(0.499)	(0.492)	
If HH head has salary/wage primary employment	0.240	0.220	0.020	1137.51***
	(0.427)	(0.414)	(0.139)	
If HH head participate in casual labour	0.286	0.300	0.129	364.30***
	(0.452)	(0.458)	(0.335)	
Value of assets (MWK'000) [a]	239.062	586.333	1292.567	1755.03***
	(296.360)	(715.634)	(1577.591)	
Land area owned by HH in ha	0.702	0.677	0.774	24.47***
	(0.684)	(0.696)	(0.847)	
HH Distance in (KMs) to Nearest Road	8.101	8.258	8.272	0.51
	(9.691)	(9.686)	(9.651)	
Observations	2507	2973	3056	

Note: Numbers shown are averages and their corresponding standard deviations are presented in parenthesis. [a] Household heads without formal qualification is the reference category in the empirical analysis. [b] Value of assets winsorized at 5 percent both at the top and bottom. * p<0.10, ** p<0.05, *** p<0.010

Turning to the independent variables used in the analysis, most socio-economic characteristics differ by survey wave. The results show that composition of household members has changed over time whereby the number of both male and female adults aged over 65 years and both male and female adults aged between 12 and 64 years has increased over time, while the number of children less than 12 years old has decreased over time, on average. This may suggest that the households are not constrained in adult labour supply. Further, the results indicate significant differences in means for household age and education level of household head at 95% confidence interval, on average. The participation in social safety net programs has increased, while both formal and informal employment opportunities have decreased over time, on average. This suggests that households have become more vulnerable to income shocks. The value of household assets and the size of agricultural land have increased over time, on average.

School data

Table 4.3 presents summary statistics of the dependent variables used in the analysis. The table shows that enrolment of boys is higher than enrolment of girls, on average. The difference in means is significant at 95% significance level. This may suggest that fewer girls enrol in school than boys. Moving on to school dropout, the table reveals that the dropout rate for boys is lower than for girls, on average. The difference in means is significant at 95% significance level, suggesting that girls drop out of school more than boys in schools. Turning to school repetition, the results indicate that the repetition rate for boys is higher than for girls, on average. Similarly, the difference in means is significant at 95% significance level, indicating that boys repeat more than girls.

Table 4.3: Summary statistics of school data

	Boys		Girls		Mean
Variables	**Mean**	**SD**	**Mean**	**SD**	**Diff.**
Enrolment	45.345	31.486	44.975	32.801	0.370***
Dropout rate	0.031	0.047	0.041	0.059	-0.010***
Repetition rate	0.175	0.133	0.167	0.131	0.007***
Observations	7,213		7,213		

Note: Numbers shown are averages along with their corresponding standard deviations. School enrolment, the dropout rate, and the repetition rate are trimmed at the top 5 percent.

4.6 Empirical results

4.6.1 Household level

Table 4.4 presents results of the effects of rainfall shocks on schooling outcomes, equation [4.1], where effects on boys are in columns 1 – 3 and effects on girls are in columns 4 – 6. As we expected, the table indicates that the direction of the effect of the drought shock on school attendance is negative (row 1, columns 1 and 4), however, the direction of the effect of the flood shock on school attendance is positive (row 2, columns 1 and 4). However, the effects are not significant among boys and girls at 95% confidence interval. Turning on now to school progression, the table shows that the drought shock increases school progression among boys by 12 percent relative to a normal rainfall, on average. Moving on to the share of expenditure on education, the table shows that the direction of the effects of the drought shock is negative, while the direction of the effect of the flood shock is positive among boys (column 3). Conversely, the direction of the effect of the drought shock is positive, while the direction of the effect of the flood shock is negative among girls (column 6). However, the effects are significant only for the flood shock among boys at 95% confidence interval. Thus, households differentially allocate more resources to boys than girls in the event of the flood shock relative to the normal rainfall by 0.0007 percentage points, on average.[79]

Table 4.4: Impact of rainfall shocks on child schooling outcomes

	Boys			Girls		
Variables	**School attendance (1)**	**Changed grade level (2)**	**Share of school expenditure (3)**	**School attendance (4)**	**Changed grade level (5)**	**Share of school expenditure (6)**
Drought	-0.00915	0.121**	-0.000132	-0.0280	0.00328	0.000194
	(0.0200)	(0.0493)	(0.000334)	(0.0222)	(0.0463)	(0.000477)
Floods	0.0122	0.0824*	0.000686***	0.0230	0.0196	-0.000374
	(0.0127)	(0.0480)	(0.000249)	(0.0156)	(0.0464)	(0.000313)
Other covariates	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Child FE	Yes	Yes	No	Yes	Yes	No
Household FE	No	No	Yes	No	No	Yes
N	2927	2448	3085	2981	2470	3084

Note: Linear regression absorbing multiple levels of fixed effects (reghdfe) estimator results in columns 1 through to 6. The dependent variable is an indicator for school attendance in columns 1 and 4, an indicator for school progression in columns 2 and 4 and share of expenditure on education in columns 3 and 6. The drought indicator takes on a value of one if the negative standardised deviation of rainfall is equal to or less than negative one, and zero otherwise, whereas the flood indicator takes on a value of one if the positive standardised deviation of rainfall is equal to or greater than positive one, and zero otherwise. Standard errors are clustered at the Traditional Authority level in parentheses. * p<0.10, ** p<0.05, *** p<0.010.

[79] Using a Chow test to examine whether the results are systematically different between boys and girls, we find that the results for school progression (F-statistic=3.98, p-value=0.02) and share of expenditure on education (F-statistic=4, p-value=0.022) are significantly different between boys and girls, while the results for school attendance (F-statistic=1.42, p-value=0.25) are similar between boys and girls at 95% confidence interval.

We now turn on to the alternative measures of rainfall shocks. Firstly, we re-estimate equation [4.1] with the drought indicator that takes on a value of one if the negative standardised deviation of rainfall from historical mean precipitation in the community is equal to or less than minus 0.5, and zero otherwise, and the flood indicator that takes on a value of one if the positive standardised deviation of rainfall from historical mean precipitation in the community is equal to or greater than plus 0.5, and zero otherwise. Overall, we find similar results of the effects of the drought and flood shocks on school attendance, school progression, and the share of expenditure on education among boys and girls relative to the normal rainfall, on average (see table C.2 in the appendix). Secondly, we re-estimate equation [4.1] with the standardised deviation of rainfall as a continuous measure of the rainfall shock. We find that there are no significant effects of rainfall shocks on our schooling outcomes at 95% confidence interval (see table C.3 in the appendix).

Overall, this confirms that the effects of rainfall shocks on schooling outcomes are different between boys and girls. We find that the drought shock increases school progression among boys and households differentially allocate more resources to boys than girls in the event of the flood shock relative to the normal rainfall, on average. These findings are sensitive to alternative measures of rainfall shocks, and they mask important differential effects of the rainfall shocks on schooling outcomes that vary with child age. We present differential effects of rainfall shocks that vary with child age in the next sub-section focusing on school attendance and progression.

Heterogenous effects by age

Table 4.5 presents results of how the impacts of rainfall shocks on schooling outcomes vary with child age, equation [4.2].[80] The effects on boys are in columns 1 – 2 and effects on girls are in columns 3 – 4. The table reveals that age reduces the differential effect of the drought shock on school attendance among boys (column 1, row 3) and girls (column 3, row 3) relative to the normal rainfall, on average. Further, age only reduces the differential effect of the drought shock on school progression among girls (column 4, row 3) relative to the normal rainfall. Turning to the flood shock, age increases the differential effects on school attendance

[80] Note that we are not able to examine how the effects of rainfall shocks on the share of expenditure on education vary with child age, since the data are aggregated at the household level by gender.

among boys (column 1, row 5) and girls (column 3, row 5) relative to the normal rainfall, on average. Further, age only increases the differential effect of the flood shock on school progression among girls (column 4, row 5) relative to the normal rainfall.

Table 4.5: Impact of rainfall shocks on schooling outcomes by age

Variables	Boys School attendance (1)	Changed grade level (2)	Girls School attendance (3)	Changed grade level (4)
Drought	0.298***	0.255	0.498***	0.418**
	(0.0618)	(0.166)	(0.0743)	(0.171)
Age	-0.0266***	-0.0211**	-0.0331***	-0.0229**
	(0.00398)	(0.00846)	(0.00439)	(0.00980)
Drought x Age	-0.0249***	-0.00843	-0.0407***	-0.0302**
	(0.00515)	(0.0118)	(0.00660)	(0.0127)
Floods	-0.102**	-0.0553	-0.153***	-0.289*
	(0.0483)	(0.168)	(0.0549)	(0.164)
Floods x Age	0.0113**	0.0129	0.0168***	0.0291**
	(0.00481)	(0.0135)	(0.00559)	(0.0132)
Other covariates	Yes	Yes	Yes	Yes
Child FE	Yes	Yes	Yes	Yes
N	2927	2448	2981	2470

Note: Linear regression absorbing multiple levels of fixed effects (reghdfe) estimator results in columns 1 through to 4. The dependent variable is an indicator for school attendance in columns 1 and 3, and an indicator for school progression in columns 2 and 4. The drought indicator takes on a value of one if the negative standardised deviation of rainfall is equal to or less than negative one, and zero otherwise, whereas the flood indicator takes on a value of one if the positive standardised deviation of rainfall is equal to or greater than positive one, and zero otherwise. Standard errors are clustered at the Traditional Authority level in parentheses * p<0.10, ** p<0.05, *** p<0.010.

Figure 4.1 presents the results of estimating equation [4.2] to show how the effects of the rainfall shocks on school attendance vary by child age. As we expected, the figure shows that the effect of the drought shock on school attendance is more negative for older boys (Panel A) and girls than younger boys and girls (Panel C). This means that the drought shock increases school attendance among younger boys and girls in lower primary school (*i.e.*, age ≤ 10, equivalent to lower grades $\leq$ 4), while those in upper primary school (*i.e.*, ages between 11 and 13, equivalent to grades between 5 and 8) are not affected relative to the normal rainfall. Conversely, the drought shock reduces school attendance among older boys and girls in secondary school relative to the normal rainfall. This effect is larger for upper secondary school (*i.e.*, age $= 18$, equivalent to grade $= 12$ or form $= 4$) than for lower secondary school (*i.e.*, age = 14, equivalent to grade = 9 or forms = 1) among both boys and girls. Moving on to the flood shock, the figure shows that the effect on school attendance is more positive for older boys (Panel B) and girls (Panel D) in contrast with our expectations. However, the effects are significant only for older boys and girls in upper primary and secondary schools (*i.e.*, ages between 10 and 18, equivalent to grades between 5 and 12). Thus, flood shocks increase school attendance among older boys and girls in upper primary and secondary school, while younger boys and girls in lower primary school are not affected relative to a normal rainfall. These findings are not surprising given that primary school is free, while secondary school is not free in Malawi. In contrast with Björkman-Nyqvist (2013) and Zimmermann (2020), our findings indicate that the effects of the drought and flood shocks on school attendance are similar between boys and girls in Malawi.

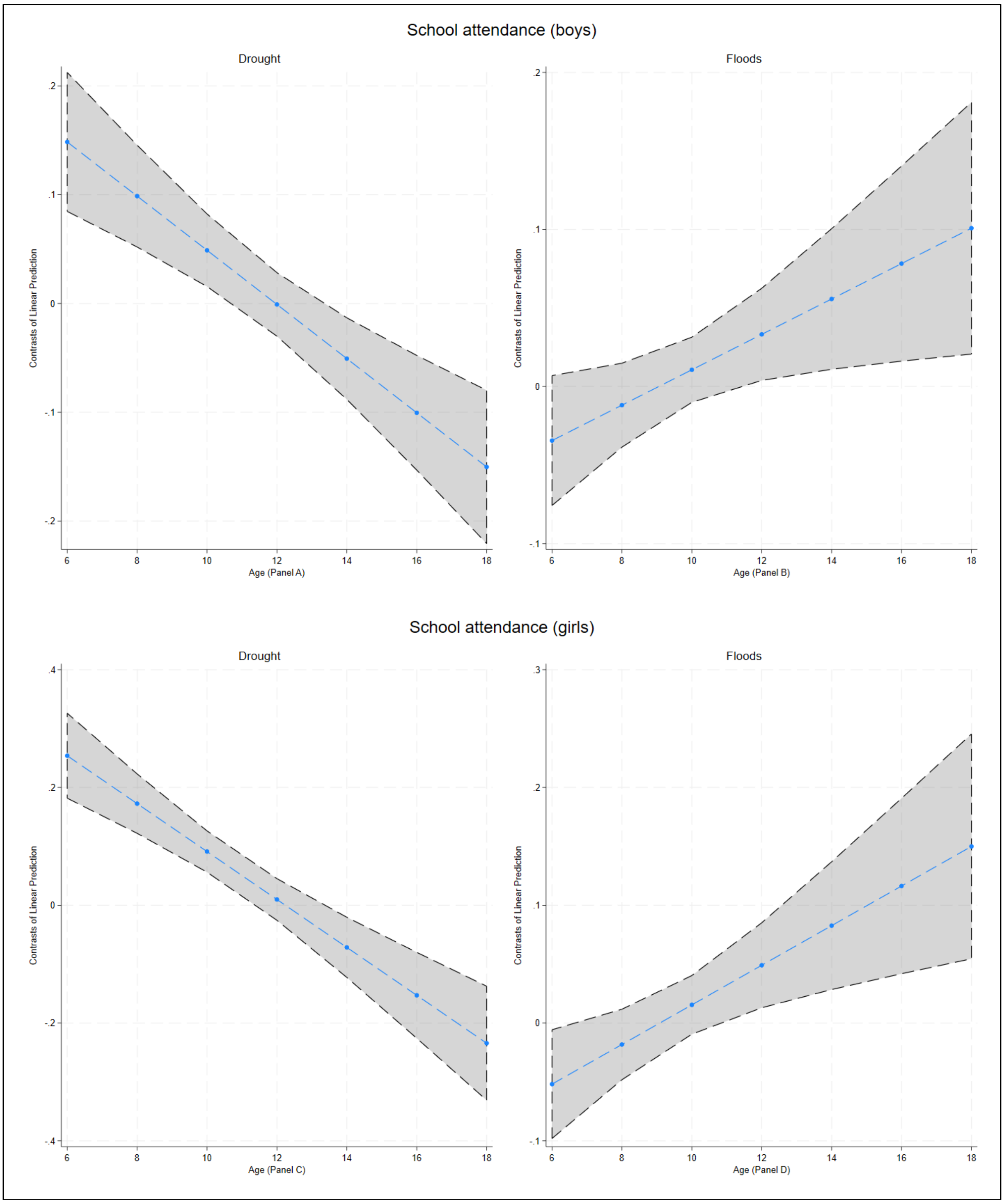

Figure 4.1: How impacts of rainfall shocks on school attendance vary with child age

Figure 4.2 presents the results of estimating equation [4.2] to show how the effects of the rainfall shocks on school progression vary by child age. The figure shows that the effect of the drought shock on school progression is more positive for younger boys than older boys (Panel E). However, the differential effects are significant only for ages between 6 and 16 (*i.e.*, grades between 1 and 10) at 95% confidence interval. This means that the drought shock increases school progression among boys, but the effect is smaller for older boys than for younger boys relative to a normal rainfall. Moving on to floods, the figure shows that the effect of the flood shock on school progression is more positive for older boys than younger boys (Panel F). however, the effects are significant only for ages between 12 and 16 (*i.e.*, grades between 7 and 10) at 95% confidence interval. Thus, the flood shock increases school progression among boys, but the effect is smaller for younger boys in upper primary school than for older boys in secondary school relative to a normal rainfall. Turning to girls, the figure shows that the effect of the drought shock on school progression is more negative among older girls (Panel G) than younger girls consistent with our expectations. However, the differential effects are significant only for ages between 6 and 10 (*i.e.*, grades between 1 and 5) at 95% confidence interval. Thus, the drought shock increases school progression among younger girls in lower primary school, but older girls in upper primary and secondary school are not affected relative to a normal rainfall. Moving on to floods, the figure shows that the effect of the flood shock on school progression is more positive for older girls than younger girls (Panel H). However, the effects are significant only for the age greater than 14 (*i.e.*, grades between 9 and 12) at 95% confidence interval. Thus, the flood shock increases school progression among girls in secondary school relative to a normal rainfall.

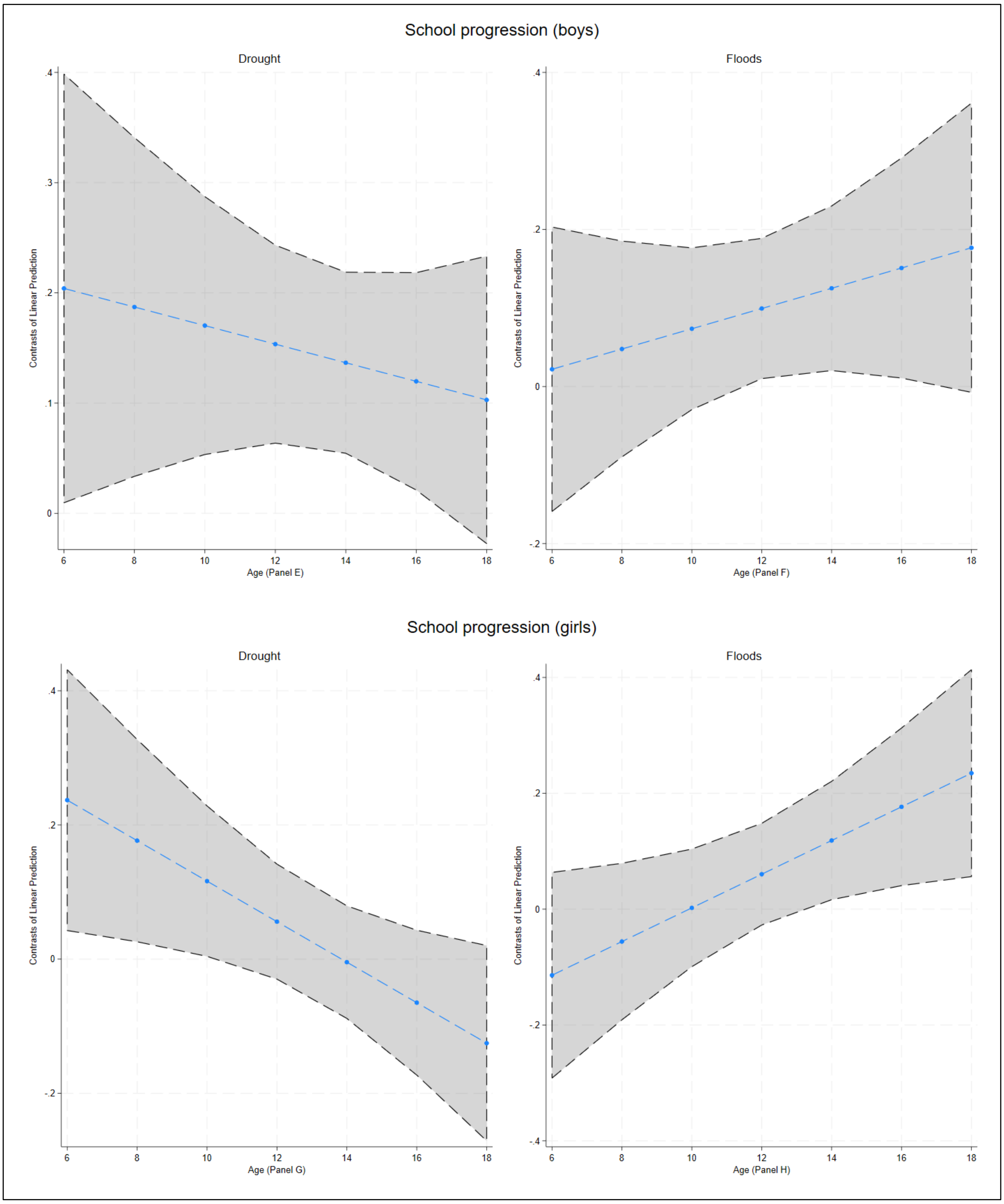

Figure 4.2: How impacts of rainfall shocks on school progression vary with child age

We now turn on to the alternative measures of rainfall shocks. We find similar results using the cut-off of equal to or less than minus 0.5, and zero otherwise for the drought indicator and the cut-off of equal to or greater than plus 0.5, and zero otherwise for the flood indicator (see figure C.3 in the appendix). Moving on to school progression, we find similar results for both the drought and flood indicators among boys relative to a normal rainfall (see figure C.4

in the appendix). However, there is no clear evidence of how the effects of rainfall shocks on school progression change with child age among girls at 95% confidence interval. These findings suggest that school attendance is less sensitive to the use of different cut-off points as alternative measures of rainfall shocks than school progression.

Re-estimating equation [4.2] with the continuous measure of the rainfall shock, we find that an increase in the standardised deviation of rainfall reduces school attendance among younger boys and girls in lower primary school (*i.e.*, ages between 6 and 10, equivalent to grades between 1 and 4), but it increases school attendance among older boys and girls in upper primary school (*i.e.*, ages between 12 and 14, equivalent to grades between 6 and 8) and in secondary school (see figure C.5 in the appendix). Turning to school progression, we find that an increase in the standardised deviation of rainfall reduces school progression among younger girls in lower primary school (*i.e.*, ages between 6 and 10, equivalent to grades between 1 and 4), but it increases school progression among older girls in upper secondary school (*i.e.*, ages between 16 and 18, equivalent to grades between 10 and 12 or forms 3 and 4) (see figure C.5 in the appendix). However, there is no clear evidence of how the effects of the changes in the standardised deviation of rainfall on school progression vary with child age among boys.

In summary, our findings confirms that effects of rainfall shocks on school attendance and progression vary with child age. Contrary to our expectations, the drought and flood shocks have opposite effects on schooling outcomes. The direction of the effect of the drought shock is negative, while the direction of the effect of the flood shock is positive on school attendance and progression. The effects of the drought shock are negative because households send younger children to school to gain from school meals and send older children to work in off-farm income activities to cope with the negative effects of the drought shock. Conversely, the effects of the flood shock are positive because the government and other development partners immediately provide support to affected households in the form of food, shelter, and nutritional supplies, which allows older children to attend school.[81] Table 4.6 summarises the results of how the effects of the rainfall shocks on school attendance and progression vary with child age. The table shows that the effects of rainfall shocks on school attendance are

[81] Unlike the flood shock, humanitarian assistance after the drought shock usually take place during the lean period (i.e., a few months before the next harvest).

similar between boys and girls, while the effects on school progression are different among boys and girls. In the next section we examine whether these findings from using the household level data are consistent at the school level.

Table 4.6: Summary of findings of how the effects of rainfall shocks on school attendance and progression vary with child age.

Rainfall shock	Boys		Girls	
	Ages (1 – 13)	**Ages (14 – 18)**	**Ages (1 – 13)**	**Ages (14 – 18)**
Drought shock	• Increases school attendance among younger boys in lower primary school. • Increases school progression.	• Reduces school attendance. • Increases school progression in lower secondary school.	• Increases school attendance among younger girls in lower primary school. • Increases school progression among younger girls in lower primary school	• Reduces school attendance.
Flood shock	• Increases school attendance among older boys in upper primary. • Increases school progression among older boys in upper primary school.	• Increases school attendance. • Increases school progression	• Increases school attendance among older girls in upper primary.	• Increases school attendance. • Increases school progression

Note: Ages 1 through to 13 represent primary school education, while ages 14 through to 18 represent secondary school education.

4.6.2 School level

This section uses school census data to examine whether the findings from using the household level data are consistent at the school level. Table 4.7 presents results of the effects of rainfall shocks on child schooling outcomes, equation [4.3] without control variables, where effects on boys are in columns 1 – 3 and effects on girls are in columns 4 – 6. The table indicates that there are no significant effects of rainfall shocks on school enrolment, the dropout rate, and the repetition rate among both boys and girls relative the normal rainfall at 95% confidence interval, on average. These findings suggest that the effects of rainfall shocks on schooling outcomes at the school level are not different between boys and girls. However, these findings mask important differential effects of the rainfall shocks on schooling outcomes that vary with child grade, which we explore in the next sub-section.

Table 4.7: Impact of rainfall shocks on schooling outcomes

	Boys			Girls		
Variables	**Enrolment (1)**	**Dropout rate (2)**	**Repetition rate (3)**	**Enrolment (4)**	**Dropout rate (5)**	**Repetition rate (6)**
Drought	-0.155	0.000000974	-0.00160	0.0657	-0.0000273	-0.000168
	(0.221)	(0.000534)	(0.00181)	(0.232)	(0.000614)	(0.00173)
Floods	0.290	-0.000324	0.000905	0.112	-0.000261	0.000632
	(0.303)	(0.000803)	(0.00225)	(0.320)	(0.000884)	(0.00217)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
School-grade FE	Yes	Yes	Yes	Yes	Yes	Yes
N	286328	239753	239800	285839	239721	239694

Note: Linear regression absorbing multiple levels of fixed effects (reghdfe) estimator results in columns 1 through to 6. The dependent variable is school enrolment in columns 1 and 4, school dropout rate in columns 2 and 5, and school repetition rate in columns 3 and 6. The drought indicator takes on a value of one if the negative standardised deviation of rainfall is equal to or less than negative one, and zero otherwise, whereas the flood indicator takes on a value of one if the positive standardised deviation of rainfall is equal to or greater than positive one, and zero otherwise. Standard errors are clustered at the Traditional Authority level in parentheses * p<0.10, ** p<0.05, *** p<0.010.

Heterogenous effects by school grade

Table 4.8 presents results of how the impacts of rainfall shocks on schooling outcomes vary with school grade, equation [4.4] without control variables, where effects on boys are in columns 1 – 3 and effects on girls are in columns 4 – 6. The table shows that school grade reduces the differential effect of the drought shock on school enrolment among boys (column 1, row 2) and the differential effect of the drought shock on the repetition rate among both boys and girls (columns 3 and 6, row 2) relative to the normal rainfall, on average. Moving on to the flood shock, school grade increases the differential effects on school enrolment among both boys and girls (columns 1 and 4, row 4) relative to the normal rainfall, on average.

Table 4.8: Impact of rainfall shocks on schooling outcomes by grade

	Boys			Girls		
Variables	**Enrolment (1)**	**Dropout rate (2)**	**Repetition rate (3)**	**Enrolment (4)**	**Dropout rate (5)**	**Repetition rate (6)**
Drought	1.020***	-0.00114	0.00623**	0.498	0.000913	0.00630**
	(0.354)	(0.000862)	(0.00265)	(0.390)	(0.00109)	(0.00262)
Drought x grade	-0.242***	0.000240*	-0.00164***	-0.0891*	-0.000200	-0.00136***
	(0.0470)	(0.000127)	(0.000304)	(0.0510)	(0.000188)	(0.000317)
Floods	-0.894**	0.000978	-0.000614	-0.640	0.000772	-0.00260
	(0.433)	(0.00158)	(0.00372)	(0.495)	(0.00168)	(0.00346)
Floods x grade	0.248***	-0.000278	0.000326	0.157***	-0.000222	0.000687
	(0.0539)	(0.000240)	(0.000449)	(0.0551)	(0.000317)	(0.000429)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
School-grade FE	Yes	Yes	Yes	Yes	Yes	Yes
N	286328	239753	239800	285839	239721	239694

Note: Linear regression absorbing multiple levels of fixed effects (reghdfe) estimator results in columns 1 through to 6. The dependent variable is school enrolment in columns 1 and 4, school dropout rate in columns 2 and 5, and school repetition rate in columns 3 and 6. The drought indicator takes on a value of one if the negative standardised deviation of rainfall is equal to or less than negative one, and zero otherwise, whereas the flood indicator takes on a value of one if the positive standardised deviation of rainfall is equal to or greater than positive one, and zero otherwise. Standard errors are clustered at the Traditional Authority level in parentheses * p<0.10, ** p<0.05, *** p<0.010.

Figure 4.3 presents the results of estimating equation [4.4] to show how the effects of the rainfall shocks on school enrolment vary by school grade. Grade 1 through to 8 represents primary school education, while grade 9 through to 12 (*i.e.*, equivalent to forms 1 – 4) represents secondary school education. Thus, grade 1 is equivalent to school starting age of 6 years and so on. As we expected, the figure shows that the effect of the drought shock on school enrolment is more negative for older boys (Panel A) and girls (Panel C) than younger boys and girls. However, the effects are significant only among boys for lower primary school (*i.e.*, grade = 1, equivalent to age = 6) and for both upper primary school (*i.e.*, grades between 7 and 8, equivalent to ages between 12 and 13) and secondary school (*i.e.*, grades between 9 and 12, equivalent to ages between 14 and 18) at 95% confidence interval. Thus, the drought shock increases school enrolment among younger boys in grade 1, while those in the middle primary school (*i.e.*, grades between 2 and 6) are not affected relative to the normal rainfall. Conversely, the drought shock reduces school enrolment among older boys in upper primary school (*i.e.*, grades between 7 and 8) and for those in secondary school (*i.e.*, grades between

9 and 12) relative to the normal rainfall. This effect is larger for boys in upper secondary school than for those in upper primary school. However, there is no clear evidence of how the effects of the drought shock change with school grade among girls at 95% confidence interval. Although the direction of the effects of the drought shock on school attendance and enrolment is the same among boys and girls, these findings suggest that the findings on the drought shock from using household level data are consistent at the school level among boys compared to girls. Moving on to the flood shock, the figure shows that the effect on school enrolment is more positive for older boys (Panel B) and girls (Panel D) than younger. However, the effects are significant only for older boys and girls in upper primary school (*i.e.*, grades between 7 and 8) and for those in secondary school (*i.e.*, grades between 9 and 12) at 95% confidence interval. This means that the flood shock increases school enrolment among older boys and girls in upper primary school and for those in secondary school, while younger boys and girls in lower primary school are not affected relative to a normal rainfall. These findings suggest that the findings on the flood shock from using household level data are consistent at the school level among boys and girls.

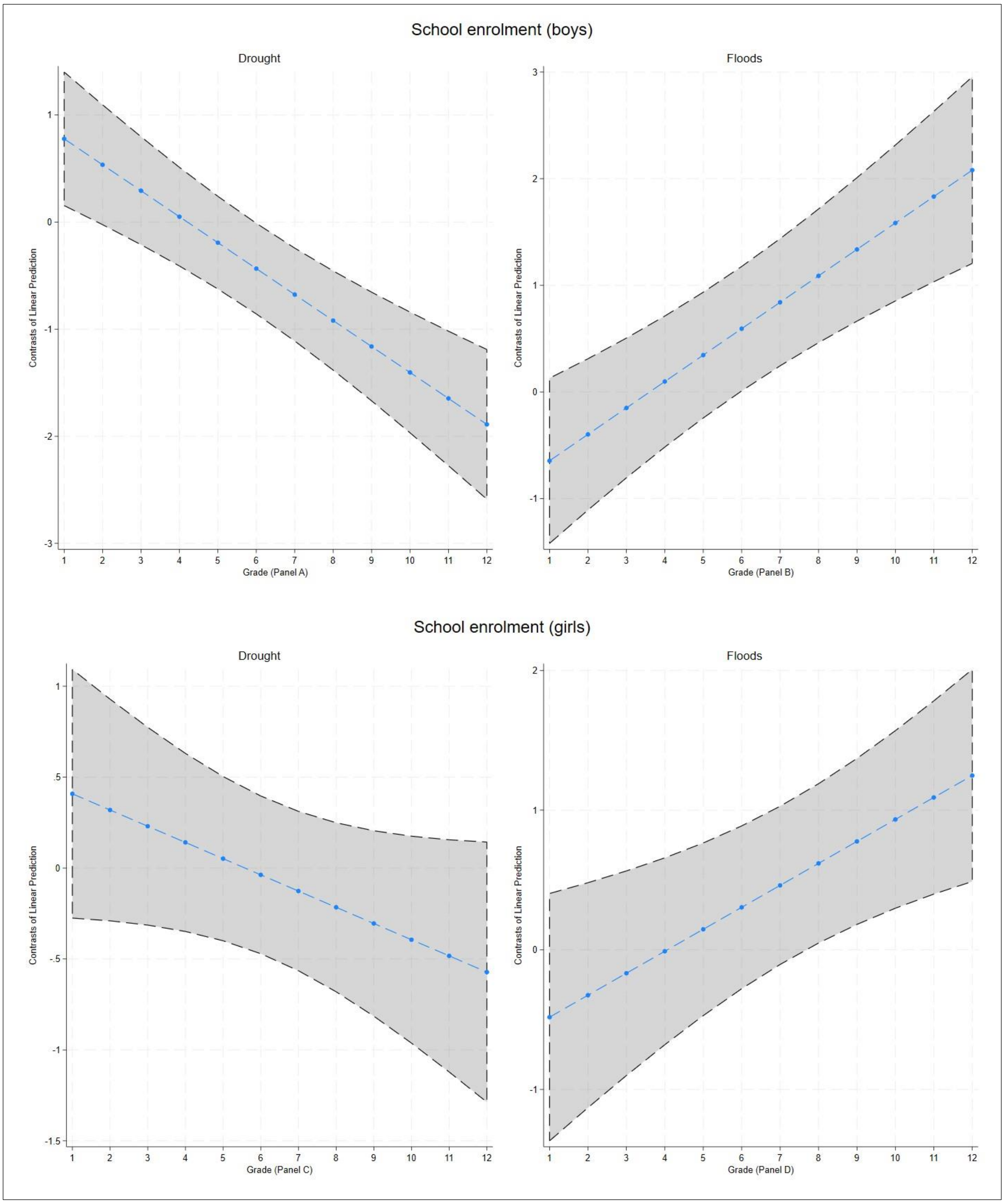

Figure 4.3: How impacts of rainfall shocks on school enrolment vary with school grade

Figure 4.4 presents the results of estimating equation [4.4] to show how the effects of the rainfall shocks on the school dropout rate vary by school grade. The figure reveals that there is no clear evidence of how the effects of rainfall shocks on school dropout change with school grade among both boys and girls at 95% confidence interval.

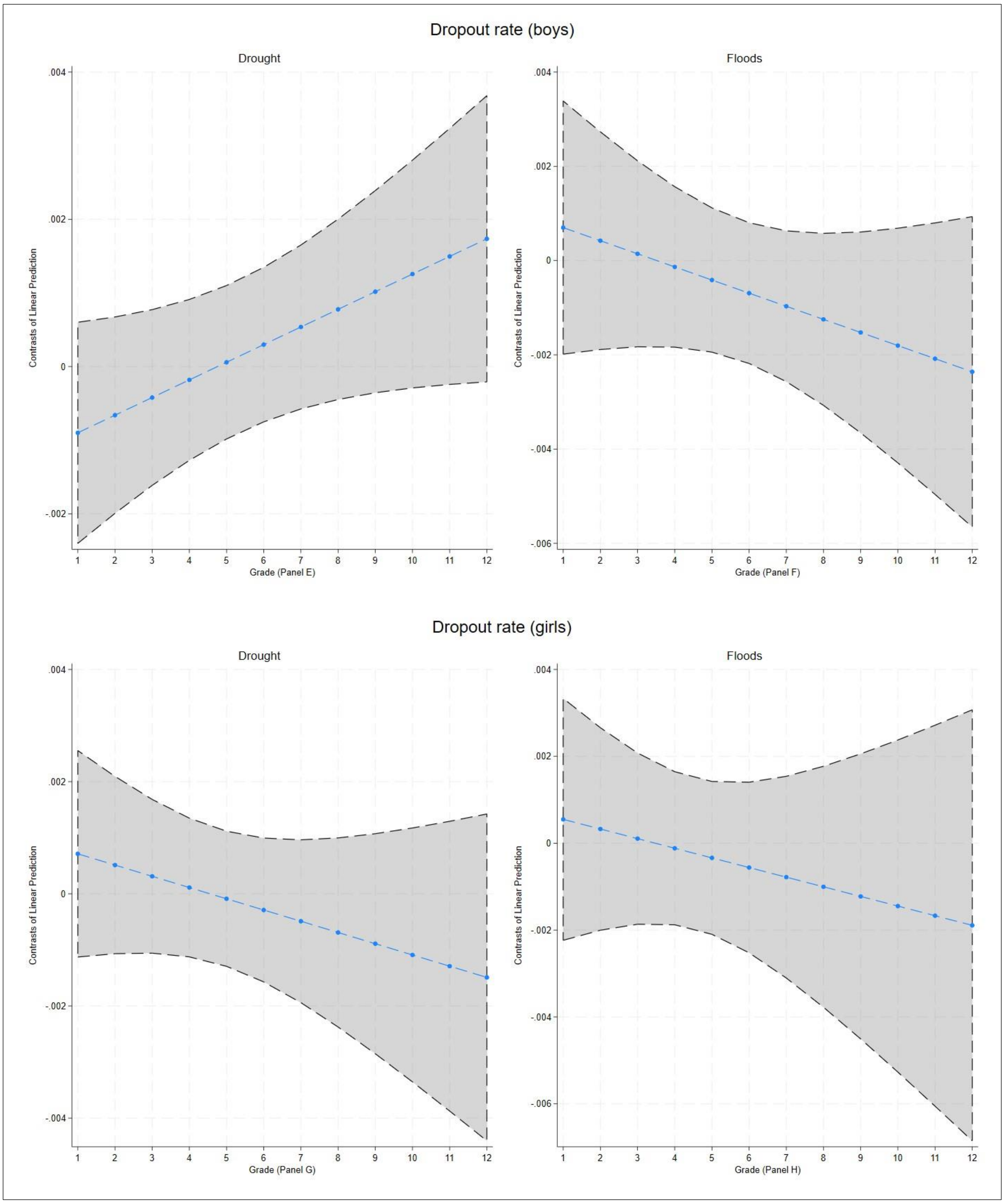

Figure 4.4: How impacts of rainfall shocks on school dropout vary with school grade

Figure 4.5 presents the results of estimating equation [4.4] to show how the effects of the rainfall shocks on the school repetition rate vary by school grade. The figure shows that the effect of the drought shock on the repetition rate is more negative for older boys (Panel A) and girls (Panel C) than for younger boys and girls. However, the effects are significant only for both boys and girls in upper primary school (*i.e.*, grades between 7 and 8) and for those in secondary school (*i.e.*, grades between 9 and 12) at 95% confidence interval. Thus, the drought shock reduces school repetition (*i.e.*, increases school progression) among older boys and girls in upper primary school (*i.e.*, grades between 7 and 8) and for those in secondary school (*i.e.*, grades between 9 and 12), while those in the lower primary school (*i.e.*, grades between 1 and 6) are not affected relative to the normal rainfall. This effect is larger for boys and girls in upper secondary school than for those in upper primary school. These findings suggest that the findings on the drought shock from using household level data are consistent at the school level among boys and girls. However, the only difference is that the effects of the drought shock on school progression from using household level data are significant among younger boys and girls in lower primary school, while with school level data the effects are significant among older boys and girls in secondary school. Turning on to the flood shock, the figure shows that there is no clear evidence of how the effects change with school grade among boys and girls at 95% confidence interval. These findings suggest that the findings on the flood shock from using household level data are consistent at the school level among boys and girls, except for boys in upper primary school.

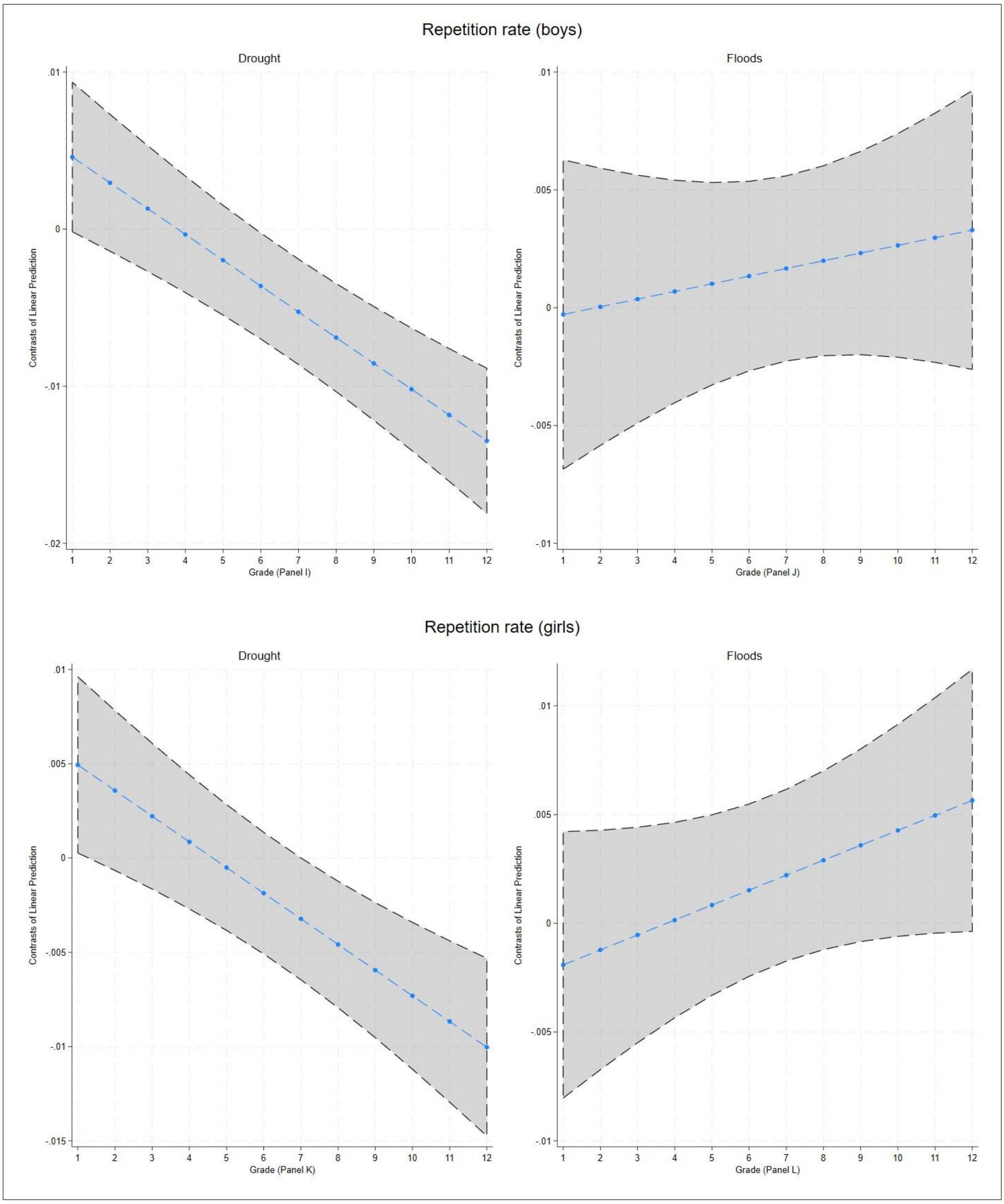

Figure 4.5: How impacts of rainfall shocks on school repetition vary with school grade

We now turn on to the alternative measures of rainfall shocks. We find similar results using the cut-off of equal to or less than minus 0.5, and zero otherwise for the drought indicator, and the cut-off of equal to or greater than plus 0.5, and zero otherwise for the flood indicator in equation [4.3], on average (see table C.4 in the appendix). We also find similar results using the continuous measure of the rainfall shock in equation [4.3] (see table C.5 in the appendix).

These findings suggest that the average effects of rainfall shocks on school enrolment, the dropout rate, and the repetition rate are not sensitive to alternative measures of rainfall shocks.

We also examine how the effects of the rainfall shocks on school enrolment vary by school grade, equation [4.4], using alternative measures of rainfall shocks. We find similar results using the cut-off of equal to or less than minus 0.5, and zero otherwise for the drought indicator, and the cut-off of equal to or greater than plus 0.5, and zero otherwise for the flood indicator (see figure C.6 in the appendix). Moving on to the dropout rate, we find that there is no clear evidence of how the effects of the rainfall shocks change with school grade among boys and girls relative to the normal rainfall at 95% confidence interval, except that the drought shock increases school the dropout rate among older boys in upper secondary school (*i.e.*, grades between 11 and 12) (see figure C.7 in the appendix). Turning to the repetition rate, we find similar results using the cut-off of equal to or less than minus 0.5, and zero otherwise for the drought indicator, and the cut-off of equal to or greater than plus 0.5, and zero otherwise for the flood indicator (see figure C.8 in the appendix). Thus, the findings are robust to the use of different cut-off point as the alternative measure of rainfall shocks.

Considering now the continuous measure of the rainfall shock in equation [4.4], we find that an increase in standardised deviation of rainfall reduces school enrolment among younger boys in grade 1 (*i.e.*, age $= 6$), but it increases school enrolment among older boys and girls in upper primary school (*i.e.*, grades between 6 and 8) and for those in secondary school (*i.e.*, grades between 9 and 12) at 95% confidence interval (see figure C.9 in the appendix). This finding is in contrast with Björkman-Nyqvist (2013) who found that the increase in deviation in rainfall from the mean increases enrolment only among older girls in Uganda. Moving on to the dropout rate, we find that the increase in standardised deviation of rainfall reduces school dropout among boys in grade 8 and for those in secondary school at 95% confidence interval (see figure C.10 in the appendix). Turning to the repetition rate, we find that the increase in standardised deviation of rainfall increases school repetition among boys and girls in upper primary school (*i.e.*, grades between 6 and 8), and for those in secondary school (*i.e.*, grades between 9 and 12) at 95% confidence interval (see figure C.11 in the appendix). Overall, these findings suggest that the results from using household level data on school attendance are consistent at the school level, while those on school progression are not consistent among boys and girls using the standardised deviation of rainfall as the alternative measure of rainfall shocks.

Overall, this sub-section has shown that the findings from the previous section from using household level data to examine how rainfall shocks affect schooling outcomes are generally consistent at the school level.

4.7 Conclusions and policy implications

This chapter investigates the extent to which rainfall shocks differentially affect schooling outcomes among boys and girls using an empirical application for Malawi. Rainfall shocks affect household income and consumption by impacting agricultural production in most developing countries. However, differential effects on schooling outcomes among boys and girls may arise when households treat boys and girls differently when coping from rainfall shocks. Consistent with the previous literature, we expect the drought and flood shocks to have negative effects on schooling outcomes. Further, we explore whether there are differential effects of rainfall shocks on schooling outcomes that varies with child age. We anticipate order girls to be affected by rainfall shocks, while boys and young girls to be insulated from the effects of rainfall shocks. Our measure of drought shock is a dummy indicator that takes on a value of one if the negative standardised deviation of rainfall from historical mean precipitation in the community is equal to or less than negative one, and zero otherwise, whereas the flood shock is a dummy indicator that takes on a value of one if the positive standardised deviation of rainfall from historical mean precipitation in the community is equal to or greater than positive one, and zero otherwise. We use both household level panel data (2010, 2013, and 2016) from IHPS and the administrative school-level census data (2010 – 2016) from the Ministry of Education in Malawi and apply the fixed effects estimator separately for boys and girls.

Results from the analysis show that households allocate more resources in boys' education during the periods of the flood shock, while resource allocation in girls' education is similar during the periods of the rainfall shock and the normal rainfall. Consistent with our expectations, we find that the effects of the rainfall shock on school attendance and progression vary with child age. We find that the drought shock increases school attendance among younger boys and girls in lower primary school, but it reduces school attendance among older boys and girls in secondary school relative to the normal rainfall. Although, children in upper primary school are not affected relative to the normal rainfall. Conversely, the flood shock increases school attendance among older boys and girls in upper primary and secondary school relative to a normal rainfall. However, younger boys and girls in lower primary school are not affected relative to the normal rainfall. Moving on to school progression, we find that the drought shock increases school progression among boys and younger girls in lower primary school, while the flood shock increases school progression among older boys in upper primary school and younger boys in lower secondary school, and among girls in secondary school relative to the normal rainfall. However, school progression among older girls in secondary school is not affected by the drought shock, while school progression among younger girls in primary school is not affected by the flood shock relative to a normal rainfall. Thus, the effects of rainfall shocks on school attendance are similar between boys and girls, while the effects on school progression are different among boys and girls. Overall, these findings are consistent at the school level.

Although government compensation, school meals, and social safety net programs during the periods of rainfall shocks might attenuate the effects of rainfall shocks on schooling outcomes, this study has demonstrated that rainfall shocks differentially affect schooling outcomes among boys and girls. Households invest more in boys' education than in girls' education during the periods of the flood shock. While the negative effects of the drought shock on school attendance are similar between boys and girls in primary school, school attendance of older boys and girls in secondary schools declines during the periods of the drought shock. Given that enrolment in secondary school is lower than in primary school, and that households are required to pay school fees in secondary school compared to primary school, there is a need to provide financial support to poor households during the periods of the flood shock to allow them to keep their children in secondary schools. While our analysis captures season to season variation in rainfall, there is a need to monitor schooling outcomes during the periods of extreme rainfall shocks to better understand how extreme rainfall shocks affect gender inequality in schooling in developing countries. Further, our outcomes of interest do not capture the quality of education (*i.e.*, the frequency of school attendance) that children receive (Amin et al., 2006b), therefore, there is a need to assess the quality of education during the periods of rainfall shocks to better understand how rainfall shocks affect the quality of education in developing countries.

4.8 Appendix C

Table C.1: Description of variables used in the analysis.

Variables	Description	Type
Child schooling outcomes		
Child school attendance	If the child attends school when the school is in session	Dummy
Child changed grade level	If child changed the grade level from the previous academic year	Dummy
Share of expenditure in child education	The proportion of inflation-adjusted value of expenditure on education to total household expenditure per year	Continuous
Independent variables		
# of males greater than 65 years old	The number of male members in the household greater than 65 years old.	Continuous
# of male adults 12-64 years old	The number of male adults in the household between the age of 12 and 65.	Continuous
# of females greater than 65 years old	The number of female members in the household greater than 65 years old.	Continuous
# of female adults 12-64 years old	The number of female adults in the household between the age of 12 and 65.	Continuous
# of children less than 12 years old	The number of children in the household below the age of 12.	Continuous
Marital status of head, =1 if has a spouse	Marital status of the household head.	Dummy
Gender of household head, =1 if male	Gender of the household head.	Dummy
Age of head in years	Age of the household head in years.	Continuous
Highest qualification of household head, none	If the household head does not have a formal qualification.	Dummy
Highest qualification of household head, primary	If the household head attended primary education.	Dummy
Highest qualification of household head, secondary	If the household head attended secondary education.	Dummy
Highest qualification of household head, tertiary	If the household head attended tertiary education.	Dummy
if HH was on any social safety net program	If the household was a beneficiary of a social safety net program.	Dummy
If HH head has salary/wage primary employment	If salary or wage job is primary employment for the household head	Dummy
If HH head participate in casual labour	If household has a member who participate in wage or casual labour.	Dummy
Value of assets (MWK'000)	Total value of assets for each household measured in Malawian Kwacha.	Continuous
Land area owned by HH in ha	Area of land owned by the household in acres	Continuous
HH Distance in (KMs) to Nearest Road	Household distance to the nearest trunk road in kilometres	Continuous

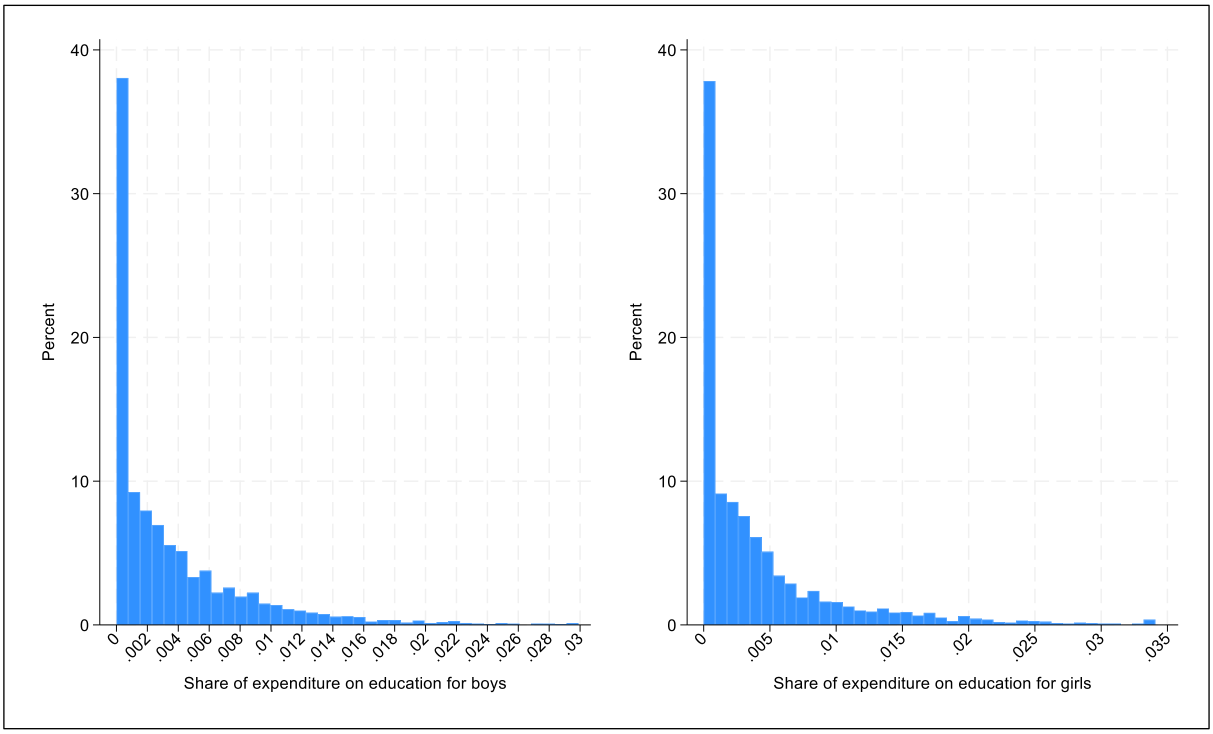

Figure C.1: Distribution of the share of expenditure on education from household data. The share of expenditure on education is trimmed at the top 5%.

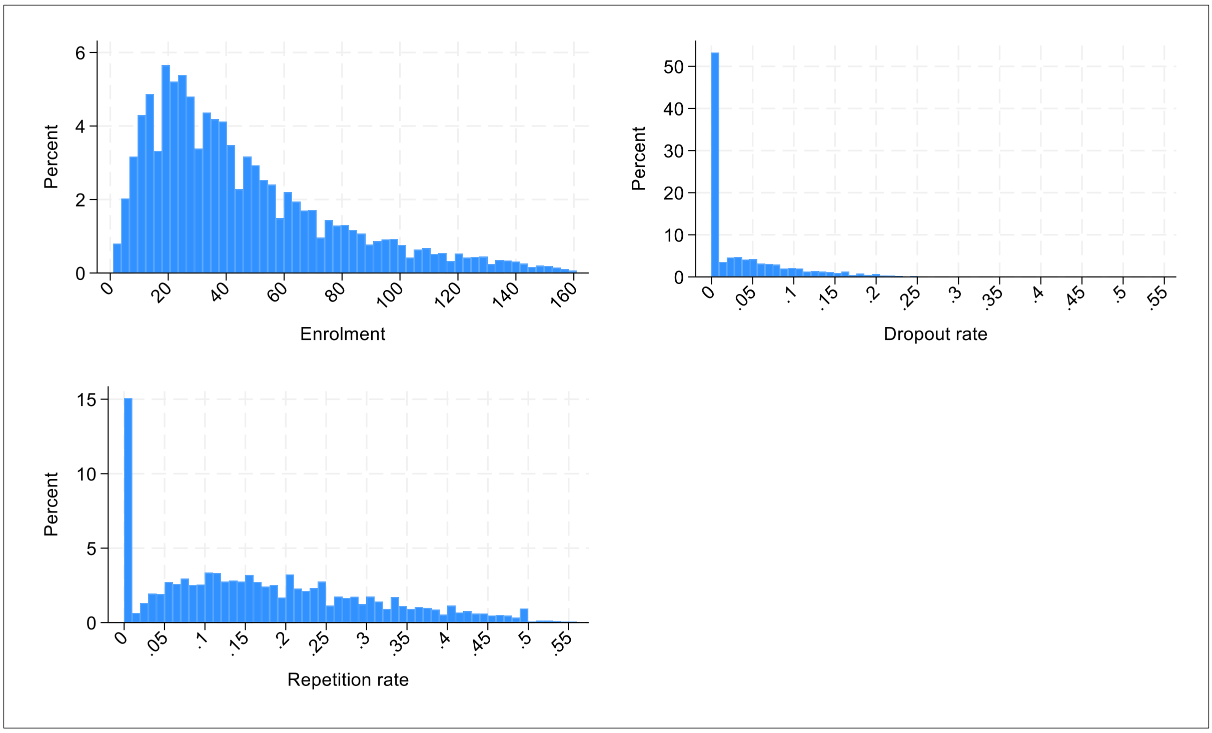

Figure C.2: Distribution of schooling outcomes from administrative data. All variables are trimmed at the top 5%.

Table C.2: Impact of rainfall shocks on child schooling outcomes

Variables	Boys			Girls		
	School attendance	Changed grade level	Share of school expenditure	School attendance	Changed grade level	Share of school expenditure
	(1)	(2)	(3)	(4)	(5)	(6)
Drought	0.0239*	0.0668	0.000118	-0.00203	0.00487	0.000188
	(0.0132)	(0.0414)	(0.000305)	(0.0187)	(0.0536)	(0.000355)
Floods	0.0210	0.0854*	0.000785***	0.0148	-0.0488	0.000261
	(0.0137)	(0.0436)	(0.000197)	(0.0158)	(0.0459)	(0.000366)
Other covariates	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Child FE	Yes	Yes	No	Yes	Yes	N0
Household FE	No	No	Yes	No	No	Yes
N	2927	2448	3085	2981	2470	3084

Note: Linear regression absorbing multiple levels of fixed effects (reghdfe) estimator results in columns 1 through to 6. The dependent variable is a dummy indicator for school attendance in columns 1 and 4, a dummy indicator for school progression in columns 2 and 4, and share of expenditure on education in columns 3 and 6. The drought indicator takes on a value of one if the negative standardised deviation of rainfall is equal to or less than negative 0.5, and zero otherwise, whereas the flood indicator takes on a value of one if the positive standardised deviation of rainfall is equal to or greater than positive 0.5, and zero otherwise. Standard errors are clustered at the Traditional Authority level in parentheses * p<0.10, ** p<0.05, *** p<0.010.

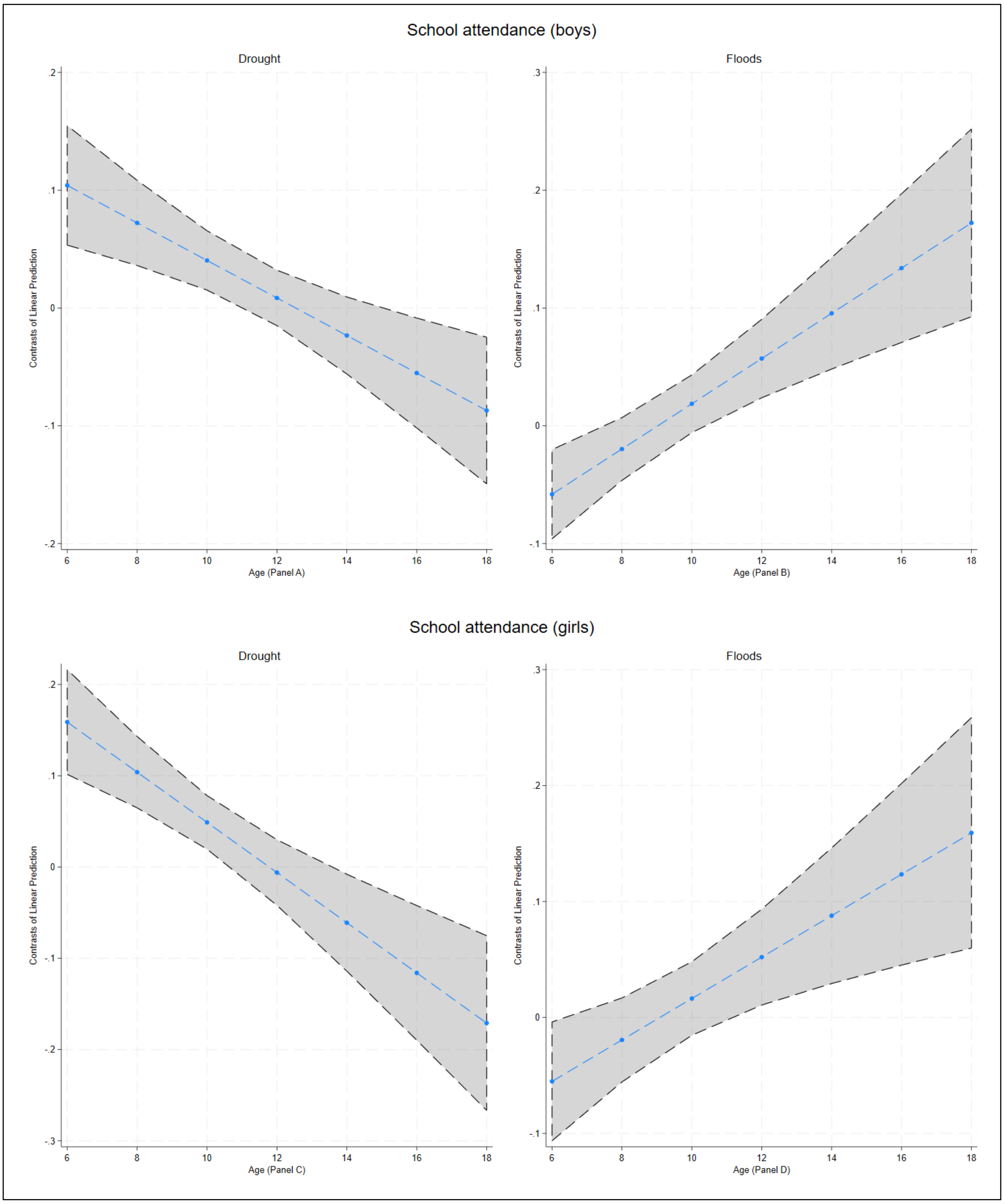

Figure C.3: How impacts of rainfall shocks on school attendance vary with child age

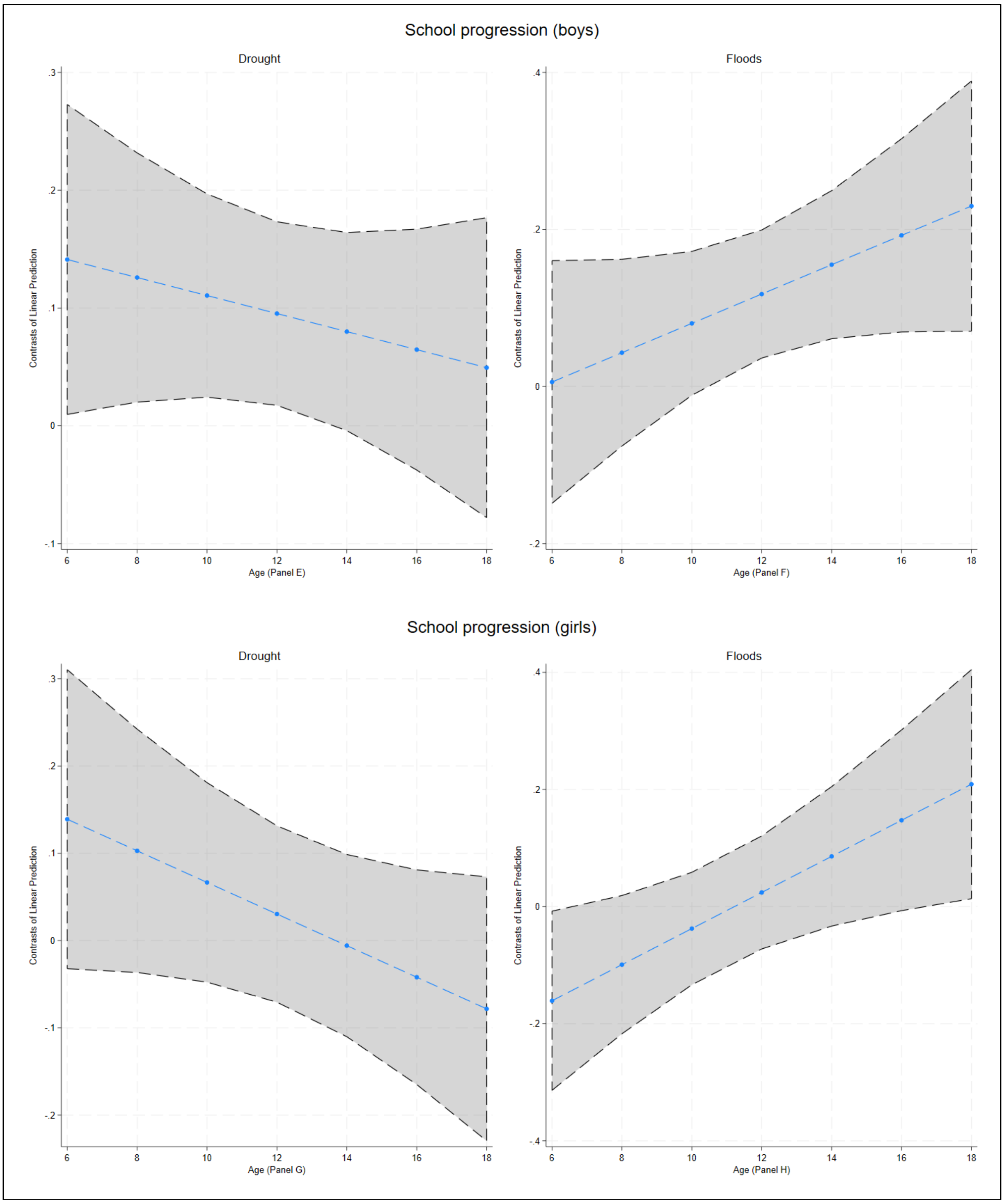

Figure C.4: How impacts of rainfall shocks on school progression vary with child age

Table C.3: Impact of rainfall shocks on child schooling outcomes

Variables	Boys			Girls		
	School attendance (1)	Changed grade level (2)	Share of school expenditure (3)	School attendance (4)	Changed grade level (5)	Share of school expenditure (6)
Standardised deviation of rainfall	0.00648	0.00811	0.000348*	0.00572	-0.0283	-0.000116
	(0.00681)	(0.0273)	(0.000195)	(0.00790)	(0.0312)	(0.000229)
Other covariates	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Child FE	Yes	Yes	No	Yes	Yes	N0
Household FE	No	No	Yes	No	No	Yes
N	2927	2448	3085	2981	2470	3084

Note: Linear regression absorbing multiple levels of fixed effects (reghdfe) estimator results in columns 1 through to 6. The dependent variable is a dummy indicator for school attendance in columns 1 and 4, a dummy indicator for school progression in columns 2 and 4 and share of expenditure on education in columns 3 and 6. Standard errors are clustered at the Traditional Authority level in parentheses * p<0.10, ** p<0.05, *** p<0.010.

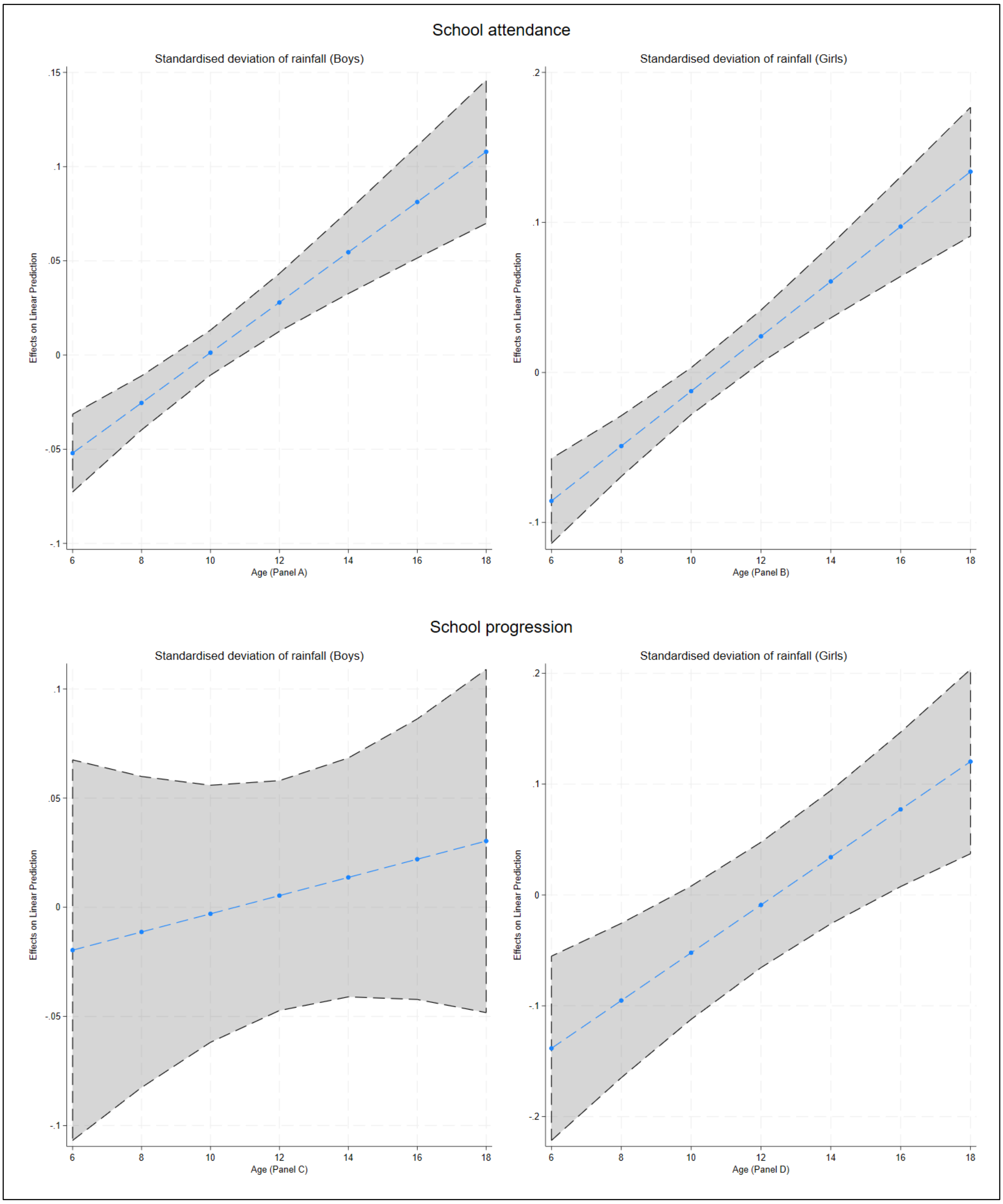

Figure C.5: How impacts of rainfall shocks on school attendance and progression vary with child age

Table C.4: Impact of rainfall shocks on schooling outcomes

Variables	Boys			Girls		
	Enrolment (1)	Dropout rate (2)	Repetition rate (3)	Enrolment (4)	Dropout rate (5)	Repetition rate (6)
Drought	-0.136	0.0000880	-0.000376	0.00224	0.000531	-0.00000563
	(0.206)	(0.000455)	(0.00141)	(0.222)	(0.000553)	(0.00138)
Floods	0.0563	-0.000503	0.000683	0.0225	-0.000261	-0.00116
	(0.225)	(0.000526)	(0.00170)	(0.233)	(0.000584)	(0.00153)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
School-grade FE	Yes	Yes	Yes	Yes	Yes	Yes
N	286328	239753	239800	285839	239721	239694

Note: Linear regression absorbing multiple levels of fixed effects (reghdfe) estimator results in columns 1 through to 6. The dependent variable is school enrolment in columns 1 and 4, school dropout rate in columns 2 and 5, and school repetition rate in columns 3 and 6. The drought indicator takes on a value of one if the negative standardised deviation of rainfall is equal to or less than negative 0.5, and zero otherwise, whereas the flood indicator takes on a value of one if the positive standardised deviation of rainfall is equal to or greater than positive 0.5, and zero otherwise. Standard errors are clustered at the Traditional Authority level in parentheses * p<0.10, ** p<0.05, *** p<0.010.

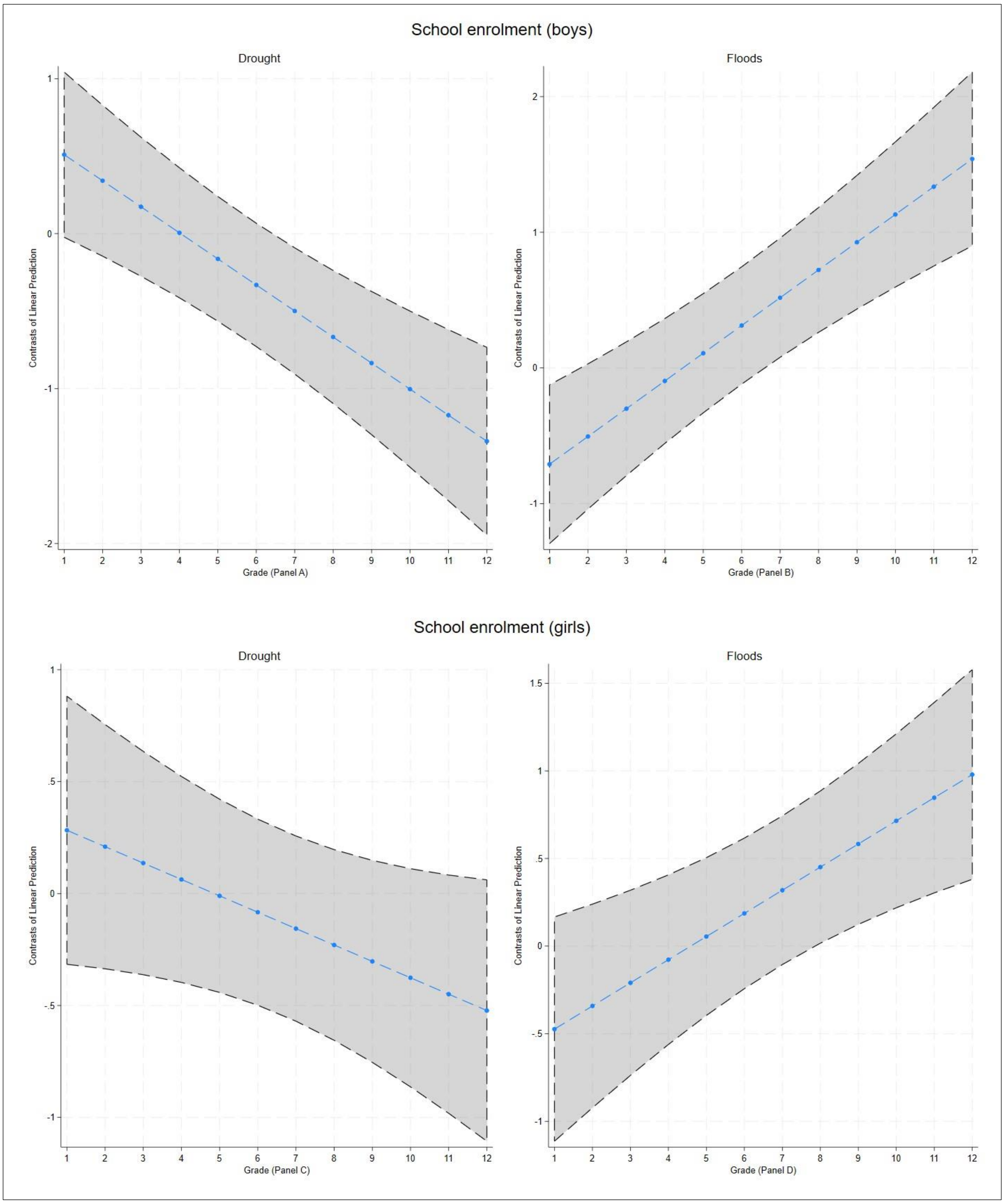

Figure C.6: How impacts of rainfall shocks on school enrolment vary with school grade

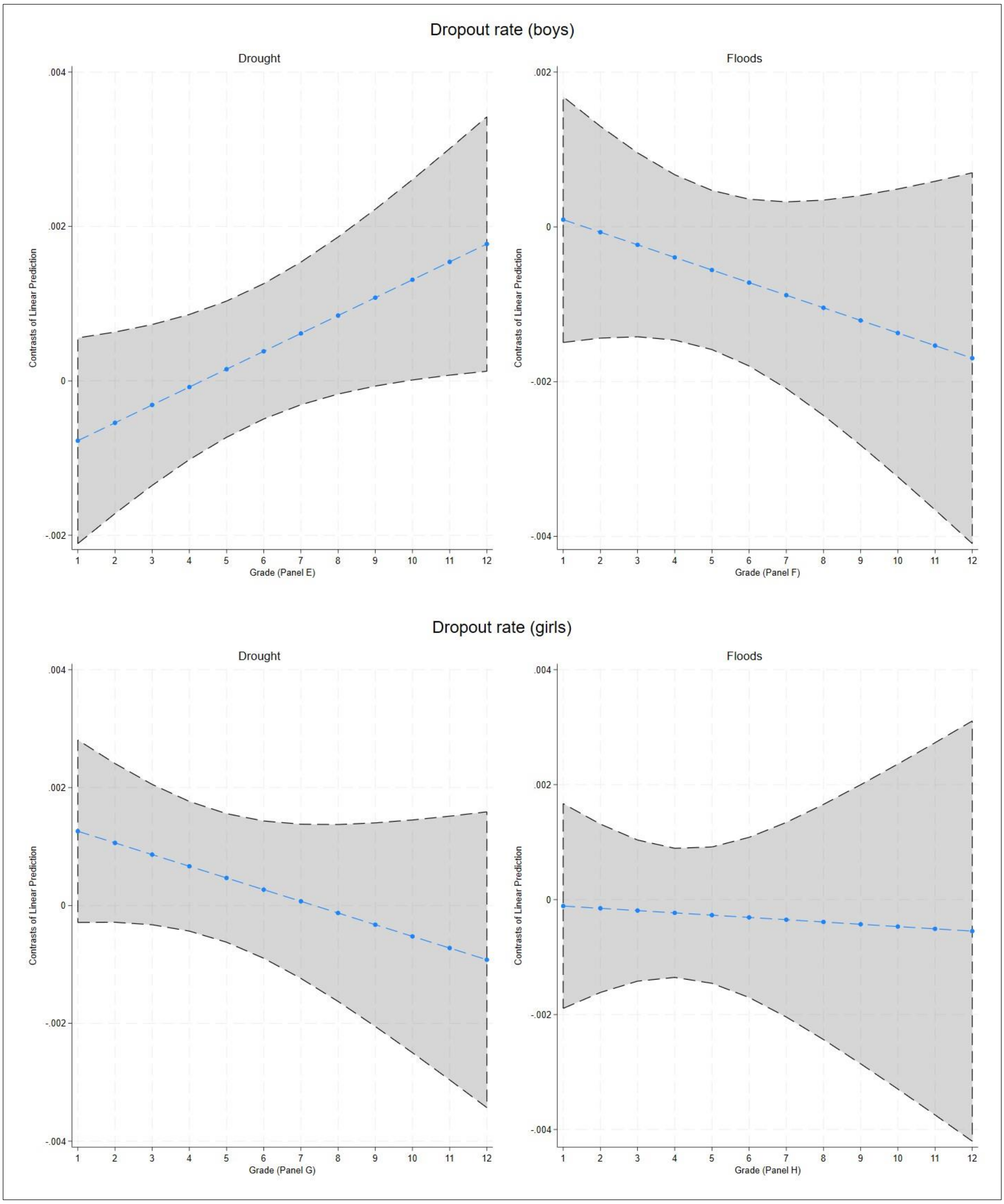

Figure C.7: How impacts of rainfall shocks on school dropout vary with school grade

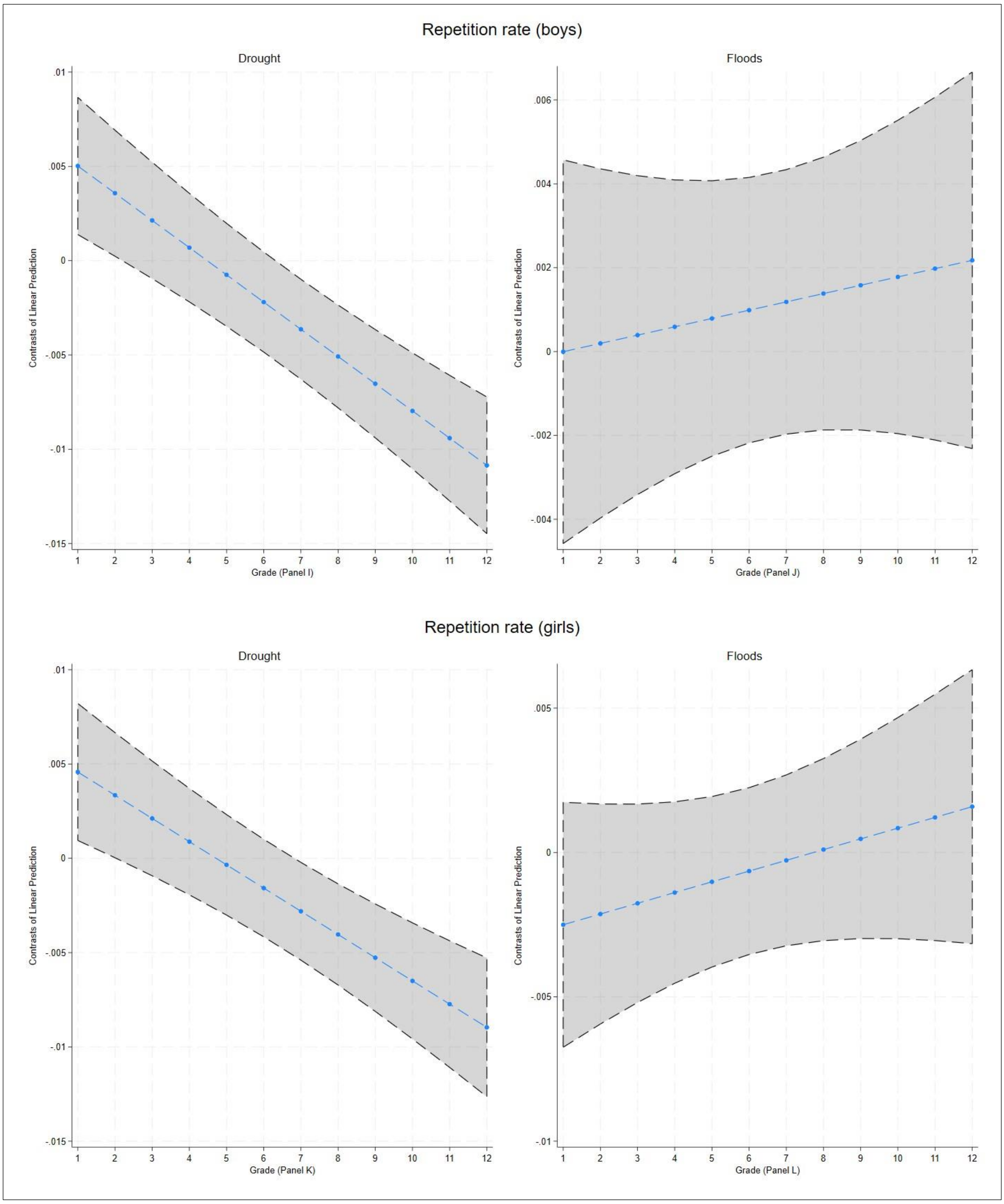

Figure C.8: How impacts of rainfall shocks on school repetition vary with school grade

Table C.5: Impact of rainfall shocks on schooling outcomes

Variables	Boys			Girls		
	Enrolment	Dropout rate	Repetition rate	Enrolment	Dropout rate	Repetition rate
	(1)	(2)	(3)	(4)	(5)	(6)
Standardised deviation of rainfall	0.221*	-0.000288	0.00106	0.146	-0.000436	0.000343
	(0.131)	(0.000300)	(0.000885)	(0.138)	(0.000353)	(0.000854)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
School-grade FE	Yes	Yes	Yes	Yes	Yes	Yes
N	286328	239753	239800	285839	239721	239694

Note: Linear regression absorbing multiple levels of fixed effects (reghdfe) estimator results in columns 1 through to 6. The dependent variable is school enrolment in columns 1 and 4, school dropout rate in columns 2 and 5, and school repetition rate in columns 3 and 6. Standard errors are clustered at the Traditional Authority level in parentheses * p<0.10, ** p<0.05, *** p<0.010.

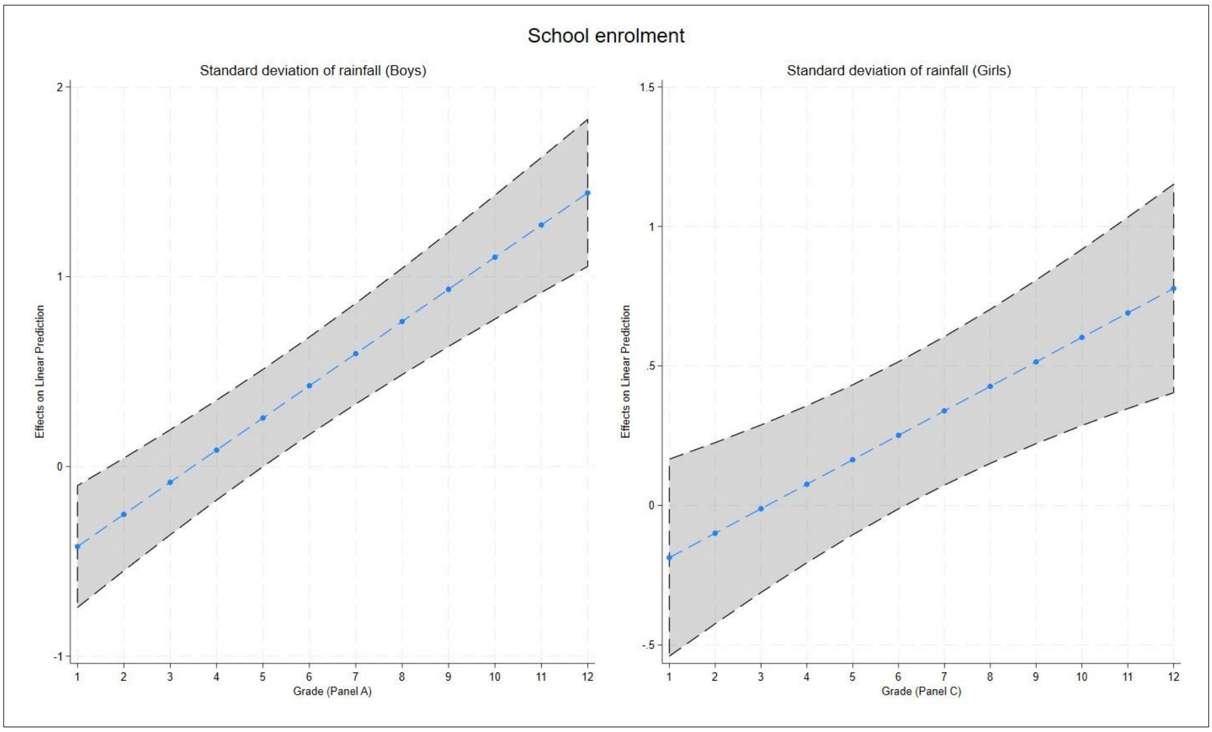

Figure C.9: How impacts of rainfall shocks on school enrolment vary with school grade

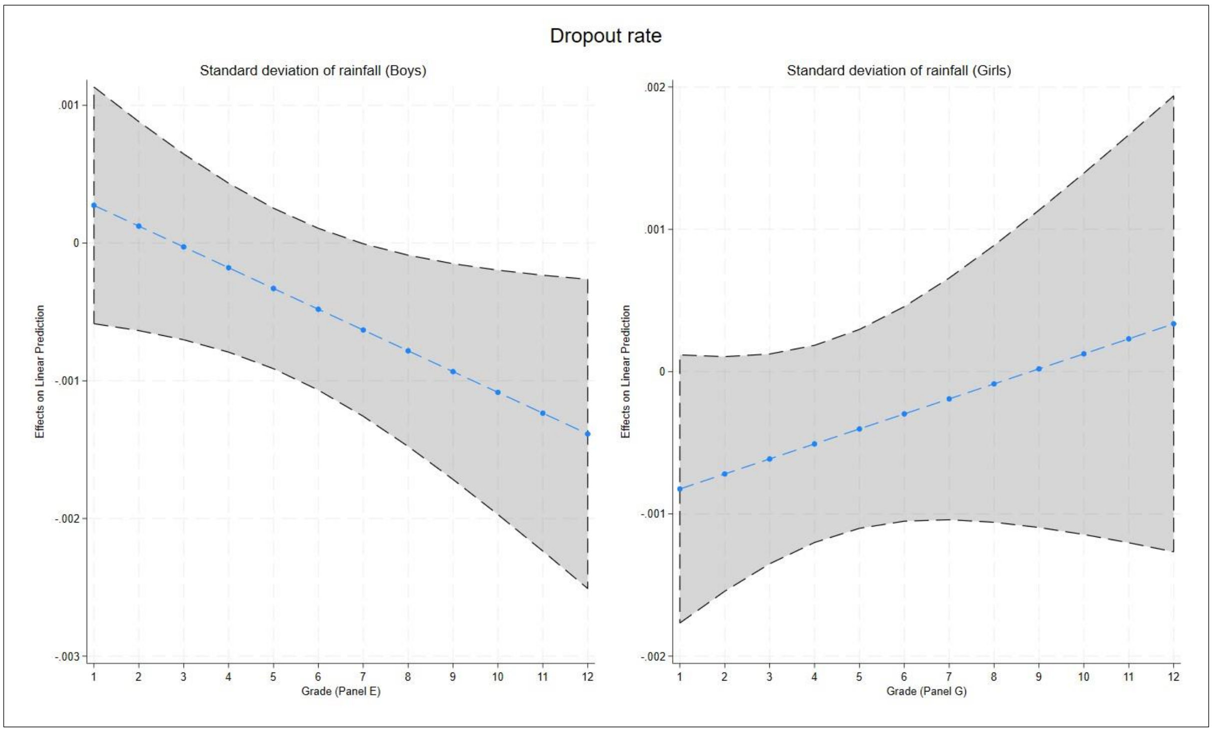

Figure C.10: How impacts of rainfall shocks on school dropout vary with school grade

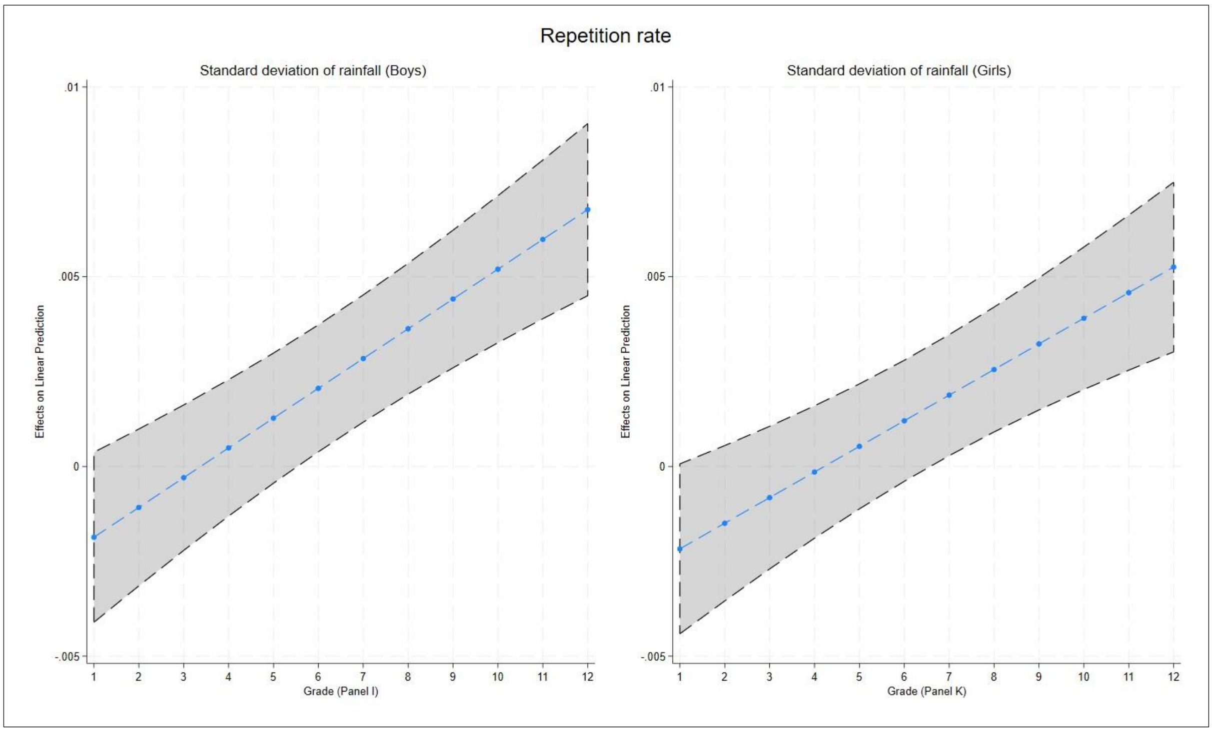

Figure C.11: How impacts of rainfall shocks on school repetition vary with school grade

Chapter 5

Conclusions

This thesis is motivated by the effects of economic shocks arising from the change in fuel prices on functionality of food markets and household well-being, and rainfall shocks on schooling outcomes in SSA using an empirical application for Malawi. The first empirical chapter examines how transport costs are associated with spatial or regional inequalities in affordability of various foods across markets. This chapter contributes to the literature that systematically look at the implications of fuel price changes on food affordability and nutrition. The second empirical chapter provides insights on how the removal of the fuel subsidy differentially affected households. This chapter contributes to the literature that investigates the effects of transport costs on production and consumption of farm produce by households in developing countries. The final empirical chapter investigates how rainfall shocks differentially affect schooling outcomes in both primary and secondary education among boys and girls. This chapter contributes to the literature that examines the effects of rainfall shocks or changing climate on schooling outcomes in developing countries.

5.1 Summary of findings

The second chapter systematically examines how transport costs influence spatial or regional inequalities in affordability of various foods across markets in Malawi. This chapter uses monthly consumer price monitoring data, monthly average diesel prices and the route distance over paved roads between the market pairs. Estimating the panel data non-linear dyadic regression model using the Poisson pseudo-likelihood regression estimator with multiple levels of fixed effects to examine short run relationships and the instrumental variable Poisson pseudo maximum likelihood estimator to examine long run relationships between transport costs and price differences of various foods across markets, the analysis reveals that the endogenous increase in transport costs is associated with a reduction in overall spatial inequalities in affordability of food across markets in the short run, on average. This counterintuitive influence is driven by processed foods but not by perishable foods and nutrient-dense foods for which the changes in transport costs is associated with the increase in overall price differences across markets in the short run. Examining the relationship between transport costs and price differentials for each food item, we find that spatial inequality in food affordability widens for maize flour dehulled maize grain, maize grain (private), maize grain (ADMARC), brown beans, and eggs in the short run. Thus, an increase in transport costs lowers the incentive of traders to move these food items from a lower price market to a higher price one, sustaining spatial inequalities in prices across markets. Overall, the magnitudes of the relationships between transport costs and price differentials are smaller for market pairs that are closer to each other for all foods under consideration. In addition, we find that spatial inequality in food affordability widens for most foods under investigation, except for brown beans, usipa, utaka, and tomatoes in the long run. Therefore, there are both food security and nutritional implications of increases in transport costs, hence, the need to promote food affordability and nutrition.

The third chapter builds on the previous chapter to investigate the extent to which the reform to Malawi's fuel policy adopted in 2012 increase or decrease agricultural production and consumption of staple maize grain differentially among households. This chapter uses three waves of nationally representative panel data from IHPS, which were implemented before and after the fuel reform. Using the data between 2010 and 2013 to estimate immediate differential effects of the policy reform, and the data between 2010 and 2016 to estimate persistent differential effects on maize production and consumption among households via a fixed effects estimator, the results confirm that there are heterogeneous differential impacts of the fuel price reform on households that vary with household status and market access. Contrary to our expectations, we find that there are both short- and long-term consequences of the fuel policy reform on staple maize production and consumption. Overall, the results indicate that households that are in autarky in remote areas increased maize production more than those closer to the market but lost in consumption as transport costs of accessing markets increased. Households that are net buyers that reside closer to the market increased maize production, consumption, and became less prone to maize insecurity, while those that reside in remote locations lost in non-food consumption and became more prone to maize insecurity relative to households that are in autarky. Conversely, households that are net sellers that reside in remote locations lost in non-food consumption and maize consumption, while those that reside closer to the market lost in consumption, non-food consumption and non-maize food consumption relative to households that are in autarky.

The final empirical chapter investigates the extent to which rainfall shocks differentially affect schooling outcomes in both primary and secondary education among boys and girls in Malawi. Rainfall shock variables are the drought indicator that takes on a value of one if the negative standardised deviation of rainfall from historical mean precipitation in the community is equal to or less than negative one, and zero otherwise, and the flood indicator that takes on a value of one if the positive standardised deviation of rainfall from historical mean precipitation in the community is equal to or greater than positive one, and zero otherwise. Using both household level panel data (2010, 2013, and 2016) from IHPS and the school census administrative data (2010 – 2016) from the Ministry of Education in Malawi and apply the fixed effects estimator separately for boys and girls, the analysis reveals that there is differential treatment in children's education whereby households allocate more resources in boys' education during the periods of the flood shock, while resource allocation in girls' education is similar during the periods of the rainfall shock and the normal rainfall. The effects of rainfall shocks on school attendance are similar between boys and girls, while the effects on school progression are different among boys and girls. For example, we find that the drought shock increases school attendance among younger boys and girls in lower primary school, but it reduces school attendance among older boys and girls in secondary school relative to the normal rainfall. Conversely, the flood shock increases school attendance among older boys and girls in upper primary and secondary school relative to a normal rainfall. Moving on to school progression, we find that the drought shock increases school progression among boys and younger girls in lower primary school, while the flood shock increases school progression among older boys in upper primary school and younger boys in lower secondary school, and among girls in secondary school relative to the normal rainfall. Overall, these findings are consistent at the school level.

5.2 Policy implications, limitations, and future research

The first empirical chapter has demonstrated how the increase in transport costs, as impacted by the increase in fuel price, is associated with the price dispersion of various foods across markets in Malawi. Overall, transport costs shock is associated with the decrease in spatial inequality in overall food affordability across markets in the short run. However, spatial inequality in food affordability widens for maize flour dehulled maize grain, maize grain (private), maize grain (ADMARC), brown beans, and eggs in the short run. Given that these food items are important in a Malawian diet, these findings indicate that there are both food security and nutritional implications of changes in transport costs. Since the increase in transport costs will limit trade, increase consumer prices, and reduce food affordability across markets there is need to devise strategies that will lower search costs to allow market traders to easily organise larger loads that will minimise the effect of fuel costs on distance, which is associated with poor market integration across the country. Examining whether the increase in trucking competition improves market integration of various foods is an area for further research. According to Fafchamps & Gabre-Madhin (2006), either removing taxes on diesel fuel that large trucks use or removing toll road fees for vehicles carrying food items across the country would lower transport costs for market traders. Whether removing taxes on diesel fuel prices or toll road fees will reduce transport costs is an area for further research. Another potential area for further research is to examine general equilibrium effects of increases in fuel costs on the economy. What we do know is that increasing market integration of food over time will allow market traders to organise and transport larger loads that will lower transport costs per unit volume that will in turn reduce the effect of fuel costs on distance and promote trade from surplus locations to deficit locations. In the longer term, there is need to consider investment in least-costs transport alternatives to road transport such as rail transportation (Donaldson, 2018; Zant, 2018) to increase food affordability and improve nutrition across the country.

The second empirical chapter has demonstrated that fuel subsidy removal, which increases transport costs has both short- and long-term consequences on households' agricultural production and consumption. One of the limitations of our identification strategy is that households might switch across the groups after the policy reform whereby those that did not sell any maize grain before the policy might start selling maize grain after the policy or those that sold maize grain before the policy might stop selling maize grain after the policy likely leading to biased results, our estimation does not capture any differential effect of the policy reform on households' entry and exit across the groups. Further, the country experienced a few crises over the period under investigation such as intensification of the border dispute with neighbouring Tanzania in 2011 (Mahony et al., 2014), a constitutional crisis in 2011/2012 (Cammack, 2012), flood shock in 2014/2015, and drought shock in 2015/2016 agricultural seasons (Floodlist, 2015; Government of Malawi, 2016c), which may lead to an overestimation of the effects of the fuel price reform on our outcomes of interest. Despite these limitations, our study has demonstrated that fuel subsidy removal, which increases transport costs has both short- and long-term consequences on households' agricultural production and consumption.

Unlike in Ethiopia where the government implemented a food subsidy scheme to reduce welfare effects of higher food prices in consumption centres at the time of fuel subsidy removal, the government of Malawi removed the fuel subsidy without a safety net programme to protect households from the effects of the policy reform. Fuje (2019) finds that ultra-poor households that are part of the food subsidy scheme gained from the fuel policy reform compared to those that did not participate in the scheme. Overall, our results indicate that households that are in autarky in remote areas increased maize production more than those closer to the market but lost in consumption. Further, households that are net buyers that reside closer to the market increased maize production, consumption, and became less prone to maize insecurity, while those that reside in remote locations lost in non-food consumption and became more prone to maize insecurity relative to households that are in autarky. Conversely, households that are net sellers that reside in remote locations lost in non-food consumption and maize consumption, while those that reside closer to the market lost in consumption, non-food consumption and non-maize food consumption relative to households that are in autarky. During the first 6 months of the subsidy removal, the government of Malawi saved about MWK36 billion (Kasalika, 2013). Therefore, the lesson for other countries that are considering rescaling or removing fuel subsidies is to contain welfare differential effects of the reform on households that reside both closer to markets or consumption centres and in remote areas. One way this can be achieved is to use the money that is saved to implement social safety net programmes at the time of subsidy removal like in Ethiopia. Such programs can target autarkic households, net buyers, and net sellers that reside in remote locations more than those that reside closer to the market or consumption centre to mitigate negative differential effects of the reform. For countries that are already implementing social safety net programmes such as input subsidies, expanding the coverage of these programmes can lead to an increase in cultivation of staple crops such as maize to increase food availability in the short term (Chibwana et al., 2010). Lowering structural barriers to trade in rural areas through investment in modern transport infrastructure such as railways and information communication technologies will be vital to reduce transport costs between the farm and consumption centre in the long-term.

The final empirical chapter has demonstrated that rainfall shocks differentially affect schooling outcomes in primary and secondary education among boys and girls. Although government compensation, school meals, and social safety net programs during the periods of rainfall shocks might attenuate the effects of rainfall shocks on schooling outcomes, the results indicate that households invest more in boys' education than in girls' education during the periods of the flood shock. While the negative effects of the drought shock on school attendance are similar between boys and girls in primary school, school attendance of older boys and girls in secondary schools declines during the periods of the drought shock. Given that enrolment in secondary school is lower than in primary school, and that households are required to pay school fees in secondary school compared to primary school, there is a need to provide financial support to poor households during the periods of the flood shock to allow them to keep their children in secondary schools. While our analysis captures season to season variation in rainfall, there is a need to monitor schooling outcomes during the periods of extreme rainfall shocks to better understand how extreme rainfall shocks affect gender inequality in schooling in developing countries. Further, our outcomes of interest do not capture the quality of education (*i.e.*, the frequency of school attendance) that children receive (Amin et al., 2006b), therefore, there is a need to assess the quality of education during the periods of rainfall shocks to better understand how rainfall shocks affect the quality of education in developing countries.

Bibliography

Abbott, P. C., Hurt, C., & Tyner, W. E. (2008). What's driving food prices? In *Archives of Ophthalmology* (Vol. 126, Issue 7). https://doi.org/10.1001/archopht.126.7.1012

Abiona, O. (2017). Adverse Effects of Early Life Extreme Precipitation Shocks on Short-term Health and Adulthood Welfare Outcomes. *Review of Development Economics*, *21*(4), 1229–1254. https://doi.org/10.1111/rode.12310

Adams, R. H., & Cuecuecha, A. (2010). Remittances, household expenditure and investment in guatemala. *World Development*, *38*(11), 1626–1641. https://doi.org/10.1016/j.worlddev.2010.03.003

African Institute of Corporate Citizenship. (2016). *Malawi Rice Outlook* (Vol. 2000). http://www.aiccafrica.org/images/documents/Malawi_Rice_Outlook__Sept-_Nov_2016.pdf

Aggarwal, S. (2018). Do rural roads create pathways out of poverty? Evidence from India. *Journal of Development Economics*, *133*(March), 375–395. https://doi.org/10.1016/j.jdeveco.2018.01.004

Aguilar, A., & Vicarelli, M. (2022). El Niño and children: Medium-term effects of early-life weather shocks on cognitive and health outcomes. *World Development*, *150*, 105690. https://doi.org/10.1016/j.worlddev.2021.105690

Aker, J. C. (2010a). Information from Markets Near and Far : Mobile Phones and Agricultural Markets in Niger. *American Economic Journal : Applied Economics*, *2*(3), 46–59.

Aker, J. C. (2010b). Rainfall shocks, markets and food crises: The effect of drought on grain markets in Niger. *Revue d'Economie Du Developpement*, *24*(1), 71–108.

Aker, J. C., & Fafchamps, M. (2015). Mobile phone coverage and producer markets: Evidence from west Africa. *World Bank Economic Review*, *29*(2), 262–292. https://doi.org/10.1093/wber/lhu006

Aker, J. C., Klein, M. W., O'Connell, S. A., & Yang, M. (2014). Borders, ethnicity and trade. *Journal of Development Economics*, *107*, 1–16. https://doi.org/10.1016/j.jdeveco.2013.10.004

Aksoy, M. A., & Isik-Dikmelik, A. (2008). Are low food prices pro-poor? Net food buyers and sellers in low-income countries. In *The World Bank - Policy Research Working Paper* (No. 4642). https://doi.org/10.1037/h0095951

Alene, A. D., Manyong, V. M., Omanya, G., Mignouna, H. D., Bokanga, M., & Odhiambo, G. (2008). Smallholder market participation under transactions costs: Maize supply and fertilizer demand in Kenya. *Food Policy*, *33*(4), 318–328. https://doi.org/10.1016/j.foodpol.2007.12.001

Allison, E. H., & Mvula, P. M. (2002). *Fisheries management policies and livelihoods: Malawi* (LADDER Working Paper No.22; Issue 22). https://assets.publishing.service.gov.uk/media/57a08d45e5274a31e000176c/Ladder-wp22.pdf

Alvi, E., & Dendir, S. (2011). Weathering the storms: Credit receipt and child labor in the aftermath of the great floods (1998) in Bangladesh. *World Development*, *39*(8), 1398–1409. https://doi.org/10.1016/j.worlddev.2011.01.003

Amare, M., Jensen, N. D., Shiferaw, B., & Denno, J. (2018). Rainfall shocks and agricultural productivity : Implication for rural household consumption. *Agricultural Systems*, *166*(March), 79–89. https://doi.org/10.1016/j.agsy.2018.07.014

Amin, S., Quayes, S., & Rives, J. M. (2006a). Are Children and Parents Substitutes or Complements in the Family Labor Supply Decision in Bangladesh? *The Journal of Developing Areas*, *40*(1), 15–37. https://doi.org/10.1353/jda.2007.0002

Amin, S., Quayes, S., & Rives, J. M. (2006b). Market work and household work as deterrents to schooling in Bangladesh. *World Development*, *34*(7), 1271–1286. https://doi.org/10.1016/j.worlddev.2005.10.006

Amjadi, a., & Yeats, A. J. (1995). Have Transport Costs Contributed to the Relative Decline of Sub-Saharan African Exports? *Policy Research Working Paper*, *December*.

Andersson, C., Bezabih, M., & Mannberg, A. (2017). The Ethiopian Commodity Exchange and spatial price dispersion. *Food Policy*, *66*, 1–11. https://doi.org/10.1016/j.foodpol.2016.11.003

Andree, B. P. J. (2021). Estimating food price inflation from partial surveys. In *Policy Research Working Paper - World Bank* (Issue 9886).

Ansah, I. G. K., Gardebroek, C., & Ihle, R. (2020). Shock interactions, coping strategy choices and household food security. *Climate and Development*, 1–13. https://doi.org/10.1080/17565529.2020.1785832

Arndt, C., Benfica, R., Maximiano, N., Nucifora, A. M. D., & Thurlow, J. T. (2008). Higher fuel and food prices: Impacts and responses for Mozambique. *Agricultural Economics*, *39*(supplement), 497–511. https://doi.org/10.1111/j.1574-0862.2008.00355.x

Aronow, P. M., Samii, C., & Assenova, V. A. (2015). Cluster-robust variance estimation for dyadic data. *Political Analysis*, *23*(4), 564–577. https://doi.org/10.1093/pan/mpv018

Arze del Granado, F. J., Coady, D., & Gillingham, R. (2012). The Unequal Benefits of Fuel Subsidies: A Review of Evidence for Developing Countries. *World Development*, *40*(11), 2234–2248. https://doi.org/10.1016/j.worlddev.2012.05.005

Asfaw, S., & Maggio, G. (2018). Gender, Weather Shocks and Welfare: Evidence from Malawi. *Journal of Development Studies*, *54*(2), 271–291. https://doi.org/10.1080/00220388.2017.1283016

Atkin, D., & Donaldson, D. (2015). Who's getting globalized? The size and implications of intra-national trade costs. In *National Bureau of Economic Research Working Paper Series* (No. 21439). http://www.nber.org/papers/w21439

Ba, M., & Mughal, M. (2022). Weather Shocks, Coping Strategies and Household Well-being: Evidence from Rural Mauritania. *Journal of Development Studies*, *58*(3), 482–502. https://doi.org/10.1080/00220388.2021.1983168

Babu, S. C., & Chapasuka, E. (1997). Mitigating the effects of drought through food security and nutrition monitoring : Lessons from Malawi. *Food Policy*, *18*(1), 1–13.

Bacon, R., & Kojima, M. (2006). *Coping with Higher Oil Prices*. *August*, 223. https://www.esmap.org/sites/default/files/esmap-files/FR323-06_GBL_Coping_with_Higher_Oil_Prices.pdf

Baez, J. E., Lucchetti, L., Genoni, M. E., Salazar, M., Baez, J. E., Lucchetti, L., Genoni, M. E., & Salazar, M. (2017). Gone with the Storm : Rainfall Shocks and Household Wellbeing in Guatemala Gone with the Storm : Rainfall Shocks and Household Wellbeing in Guatemala. *The Journal of Development Studies*, *53*(8), 1253–1271. https://doi.org/10.1080/00220388.2016.1224853

Baltagi, B. H. (2021). Econometric Analysis of Panel Data. In *Springer Texts in Business and Economics* (6th ed.). Springer. https://doi.org/10.1007/978-3-030-53953-5

Banda, L. J., & Dzanja, J. L. (2006). Goat marketing systems and channels in Nkhoma and Nsundwe areas. *3rd Bunda College Dissemination Conference Proceedings*, 30–36. https://www.cabi.org/gara/FullTextPDF/2008/20083326853.pdf

Baquie, S., & Fuje, H. (2020). Vulnerability to Poverty following Extreme Weather Events in Malawi. In *Policy Research Working Paper* (No. 9435; Issue October). https://doi.org/10.1596/1813-9450-9435

Barrett, C. B., & Dorosh, P. A. (1996). Farmers' Welfare and Changing Food Prices: Nonparametric Evidence from Rice in Madagascar. *American Journal of Agricultural Economics*, *78*(3), 656–669. https://doi.org/10.2307/1243283

Barrett, C. B., & Li, J. R. (2002). Distinguishing between equilibrium and integration in spatial price analysis. *American Journal of Agricultural Economics*, *84*(2), 292–307. https://doi.org/10.1111/1467-8276.00298

Barro, R. J., & Lee, J.-W. (2015). International Comparisons of School Enrollment and Educational Attainment, 1870–2010. In *Education Matters*. https://doi.org/10.1093/acprof:oso/9780199379231.003.0002

Baulch, B. (1997). Transfer Costs, Spatial Arbitrage, and Testing for Food Market Integration. *American Journal of Agricultural Economics*, *79*(2), 477–487. https://doi.org/10.2307/1244145

Beguería, S., & Vicente-Serrano, S. M. (2017). Package "SPEI." *American Journal of Agricultural Economics*, *March*, 16. https://doi.org/10.1175/2009JCLI2909.1.http

Beguería, S., Vicente-Serrano, S. M., Reig, F., & Latorre, B. (2014). Standardized precipitation evapotranspiration index (SPEI) revisited: Parameter fitting, evapotranspiration models, tools, datasets and drought monitoring. *American Journal of Agricultural Economics*, *34*(10), 3001–3023. https://doi.org/10.1002/joc.3887

Behrman, J. R. (1988). Intrahousehold Allocation of Nutrients in Rural India : Are Boys Favored ? Do Parents Exhibit Inequality Aversion? *Oxford Economic Papers*, *40*(1), 32–54.

Benfica, R. (2014). Welfare and distributional impacts of price shocks in Malawi: a non-parametric approach. *Food Security*, *6*(1), 131–145. https://doi.org/10.1007/s12571-

013-0324-2

Benson, T. (1999). *Area-Specific Fertilizer Recommendations for Hybrid Maize Grown by Malawian Smallholders: A Manual for Field Assistants. Report prepared for the Maize Productivity Task Force*. https://www.cabi.org/GARA/FullTextPDF/2008/20083326757.pdf

Benson, T., & Weerdt, J. De. (2023). *Employment options and challenges for rural households in Malawi An agriculture and rural employment analysis of the fifth Malawi Integrated Household Survey , 2019 / 20*. *2023*(April). https://massp.ifpri.info/files/2023/04/Working-Paper-40-Malawi-CEM-ag-background-paper-Apr2023-revised.pdf

Björkman-Nyqvist, M. (2013). Income shocks and gender gaps in education: Evidence from Uganda. *Journal of Development Economics*, *105*, 237–253. https://doi.org/10.1016/j.jdeveco.2013.07.013

Blundell, R., Gri, R., & Windmeijer, F. (2002). Individual effects and dynamics in count data models. *Journal of Econometrics*, *108*, 113–131.

Borgomeo, E., Vadheim, B., Woldeyes, F. B., Alamirew, T., Tamru, S., Charles, K. J., Kebede, S., & Walker, O. (2018). The Distributional and Multi-Sectoral Impacts of Rainfall Shocks: Evidence From Computable General Equilibrium Modelling for the Awash Basin, Ethiopia. *Ecological Economics*, *146*(December 2017), 621–632. https://doi.org/10.1016/j.ecolecon.2017.11.038

Botha, B. N., Nkoka, F. S., & Mwumvaneza, V. (2018). Hard Hit by El Nino: Experiences, Responses, and Options for Malawi. In *World Bank Group*. https://doi.org/10.1596/30037

Boutin, D. (2014). Climate vulnerability, communities' resilience and child labour. *Revue d'Economie Politique*, *124*(4), 625–638. https://doi.org/10.3917/redp.244.0625

Branco, D., & Féres, J. (2021). Weather Shocks and Labor Allocation: Evidence from Rural Brazil. *American Journal of Agricultural Economics*, *103*(4), 1359–1377. https://doi.org/10.1111/ajae.12171

Buddy, J. W. (1993). Changing Food Prices and Rural Welfare : A Nonparametric Examination of the Côte d ' Ivoire. *Economic Development and Cultural Change*, *41*(3), 587–603.

Burke, W. J., Jayne, T. S., & Sitko, N. J. (2020). Do Medium-scale Farms Improve Market Access Conditions for Zambian Smallholders? *Journal of Agricultural Economics*, *71*(2), 517–533. https://doi.org/10.1111/1477-9552.12360

Cameron, A. C., & Miller, D. (2014). Robust Inference for Dyadic Data. In *Unpublished manuscript*. https://cameron.econ.ucdavis.edu/research/dyadic_cameron_miller_december2014_with_tables.pdf

Cameron, A. C., & Trivedi, K. P. (2013). Regression analysis of count data. In *Econometric Society Monograph No.53* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9781139013567

Cammack, D. (2012). Malawi in crisis, 2011-12. *Review of African Political Economy*, *39*(132), 375–388. https://doi.org/10.1080/03056244.2012.688651

Carpena, F. (2019). How do droughts impact household food consumption and nutritional intake? A study of rural India. *World Development*, *122*, 349–369. https://doi.org/10.1016/j.worlddev.2019.06.005

Carrillo, B. (2020). Early Rainfall Shocks and Later-Life Outcomes: Evidence from Colombia. *World Bank Economic Review*, *34*(1), 179–209. https://doi.org/10.1093/wber/lhy014

Chamberlain, G. (1982). Multivariate regression models for panel data. *Journal of Econometrics*, *18*(1), 5–46. https://doi.org/10.1016/0304-4076(82)90094-X

Cherry, T. L., & List, J. A. (2002). Aggregation bias in the economic model of crime. *Economics Letters*, *75*(1), 81–86. https://doi.org/10.1016/S0165-1765(01)00597-3

Chibwana, C., Fisher, M., Jumbe, C., Masters, W., & Shively, G. (2010). *Measuring Impact of Malawi's Farm Input Subsidy Program*. 37. https://doi.org/https://dx.doi.org/10.2139/ssrn.1860867

Chibwana, C., Fisher, M., & Shively, G. (2012). Cropland Allocation Effects of Agricultural Input Subsidies in Malawi. *World Development*, *40*(1), 124–133. https://doi.org/10.1016/j.worlddev.2011.04.022

Chigwa, C. F. (2012). Assesment of goat production systems and social-economic factors affecting demand of goats and goat meat in Malawi. In *Ruforum* (Issue September). https://ruforum.org/sites/default/files/Chimwemwe Chigwa, F..pdf

Chimombo, J. (2009). Changing patterns of access to basic education in Malawi: A story of a mixed bag? *Comparative Education*, *45*(2), 297–312. https://doi.org/10.1080/03050060902921003

Chirwa, W. E. (2006). *Commercialisation of food crops in rural Malawi : insights from the household survey* (No. 04; Issue August). https://www.researchgate.net/publication/228998083_Commercialisation_of_food_crops_in_rural_Malawi_insights_from_the_household_survey

Chirwa, W. E. (2009). *Determinants of Marketing Channels among Smallholder Maize Farmers in Malawi* (No. 03; Vol. 3). https://www.researchgate.net/publication/228546518_Determinants_of_Marketing_Channels_among_Smallholder_Maize_Farmers_in_Malawi

Christiaensen, L., & Demery, L. (2018). *Agriculture in Africa; Agriculture in Africa: Telling Myths from Facts : Telling Myths from Facts. Directions in Development—Agriculture and Rural Development*. World Bank. https://doi.org/10.1201/9781003079057-13

Coady, D., El-Said, M., Gillingham, R., Kpodar, K., Medas, P., & Newhouse, D. (2006). The Magnitude and Distribution of Fuel Subsidies: Evidence From Bolivia, Ghana, Jordan, Mali, and Sri Lanka. *IMF Working Papers*, *06*(247), 1. https://doi.org/10.5089/9781451865073.001

Coates, J., Rogers, B. L., Webb, P., Maxwell, D., Houser, R., & Mcdonald, C. (2007). Diet Diversity Study. In *World Food Programme, Emergency Needs Assessment Service*

(ODAN). https://nutrition.tufts.edu/sites/default/files/fpan/wp40_dietdiversity.pdf

Commercial Agriculture for Smallholders and Agribusiness. (2020). Poultry Sector Strategy – Malawi. In *Poultry Sector Strategy – Malawi* (Issue April). https://www.casaprogramme.com/wp-content/uploads/CASA-Malawi-PoultrySector-analysis-report.pdf

Competition and Fair Trading Commission. (2016). Competition Assessment in Malawi Transport Sector. In *A paper prepared for ACER 2016 Conference, Livingstone, Zambia*. https://static1.squarespace.com/static/52246331e4b0a46e5f1b8ce5/t/56f1352c22482e727503e388/1458648367884/Fexter+Katungwe_+Competition+in+transport+in+Malawi.pdf

Correia, S., Guimar, P., & Zylkin, T. (2019). *PPMLHDFE : Fast Poisson Estimation with High-Dimensional Fixed Effects*. *1*, 1–19. https://arxiv.org/abs/1903.01690

Cragg, J. G. (1971). Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods. *Econometrica*, *39*(5), 829. https://doi.org/10.2307/1909582

Damania, R., Berg, C., Russ, J., Barra, A. F., Nash, J., & Ali, R. (2016). Agricultural technology choice and transport. *American Journal of Agricultural Economics*, *99*(1), 265–284. https://doi.org/10.1093/ajae/aav073

Damania, R., Desbureaux, S., Hyland, M., Islam, A., Moore, S., Rodella, A.-S., Russ, J., & Zaveri, E. (2020). Uncharted Waters. The New Economics of Water Scarcity and Variability. In *Oxford Review of Economic Policy* (Vol. 36, Issue 1). World Bank Group. https://doi.org/10.1093/oxrep/grz027

Dana, J., Rohrbach, D., & Syroka, J. (2007). *Risk Management Tools for Malawi Food Security*. https://documents1.worldbank.org/curated/en/376991468276878847/pdf/703950WP0P10550038B00PUBLIC00malawi.pdf

Darmon, N., & Drewnowski, A. (2015). Contribution of food prices and diet cost to socioeconomic disparities in diet quality and health: A systematic review and analysis. *Nutrition Reviews*, *73*(10), 643–660. https://doi.org/10.1093/nutrit/nuv027

de Janvry, A., Fafchamps, M., & Sadoulet, E. (1991). Peasant Household Behaviour with Missing Markets: Some Paradoxes Explained. *The Economic Journal*, *101*(409), 1400. https://doi.org/10.2307/2234892

Deaton, A. (1989). Rice prices and income distribution in Thailand: a non-parametric analysis. *Economic Journal*, *99*(395 Supp), 1–37. https://doi.org/10.2307/2234068

Deaton, Angus, & Zaidi, S. (2002). Guidelines for constructing consumption aggregates for welfare analysis. In *World Bank Living Standards Measurement Study Working Paper* (Vol. 135). https://www.princeton.edu/~deaton/downloads/Guidelines_for_Constructing_Consumption_Aggregates.pdf

Dell, M., Jones, B. F., & Olken, B. A. (2014). What do we learn from the weather? The new climate-economy literature. *Journal of Economic Literature*, *52*(3), 740–798.

https://doi.org/10.1257/jel.52.3.740

Derlagen, C. (2012). Analysis of incentives and disincentives for maize in Malawi. In *Monitoring African Food and Agricultural Policies* (Issue October). https://www.fao.org/3/at563e/at563e.pdf

Devereux, S. (2002). State of Disaster: Causes, Consequences and Policy Lessons from Malawi. In *ActionAid Report* (Issue June). papers2://publication/uuid/C89C3FCA-9B02-4925-BF6E-C9DF713DFA11

Diagne, A. (1998). *Impact of access to credit on income and food security in Malawi* (No. 46; Issue FCND DISCUSSION PAPER). https://ebrary.ifpri.org/utils/getfile/collection/p15738coll2/id/125691/filename/125722.pdf

Dietrich, S., Giuffrida, V., Martorano, B., & Schmerzeck, G. (2022). COVID-19 policy responses mobility and food prices. *American Journal of Agricultural Economics*, *104*, 569–588. https://doi.org/10.1111/ajae.12278

Dillon, B. M., & Barrett, C. B. (2016). Global Oil Prices and Local Food Prices: Evidence from East Africa. *American Journal of Agricultural Economics*, *98*(1), 154–171. https://doi.org/10.1093/ajae/aav040

Dimova, R., & Gbakou, M. (2013). The Global Food Crisis: Disaster, Opportunity or Non-event? Household Level Evidence from Côte d'Ivoire. *World Development*, *46*, 185–196. https://doi.org/10.1016/j.worlddev.2013.02.007

Dittus, K. L., Hillers, V. N., & Beerman, K. A. (1995). Benefits and barriers to fruit and vegetable intake: relationship between attitudes and consumption. *American Journal of Agricultural Economics*, *27*(3), 120–126. https://doi.org/10.1016/S0022-3182(12)80802-8

Donaldson, D. (2018). Railroads of the Raj: Estimating the impact of transportation infrastructure. *American Economic Review*, *108*(4–5), 899–934. https://doi.org/10.1257/aer.20101199

Dorward, A., Chirwa, E., Kelly, V. A., Jayne, T. S., Slater, R., & Boughton, D. (2008). *Evaluation of the 2006/7 Agricultural Input Subsidy Programme, Malawi: Final Report*. *March*, 1–131. https://eprints.soas.ac.uk/16730/1/MarchReportRevMayNoTrack.pdf

Drewnowski, A., Darmon, N., & Briend, A. (2004). Replacing fats and sweets with vegetables and fruits - A question of cost. *American Journal of Public Health*, *94*(9), 1555–1559. https://doi.org/10.2105/AJPH.94.9.1555

Drewnowski, A., & Specter, S. E. (2004). Poverty and obesity: The role of energy density and energy costs. *American Journal of Clinical Nutrition*, *79*(1), 6–16. https://doi.org/10.1093/ajcn/79.1.6

Dube, O., & Vargas, J. F. (2013). Commodity price shocks and civil conflict: Evidence from Colombia. *American Journal of Agricultural Economics*, *80*(4), 1384–1421. https://doi.org/10.1093/restud/rdt009

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. https://doi.org/10.1086/701828

Eckard, E. W. (2004). The "law of one price" in 1901. *Economic Inquiry*, *42*(1), 101–110. https://doi.org/10.1093/ei/cbh047

Edmonds, E. V. (2007). Chapter 57 Child Labor. In *Handbook of Development Economics* (Vol. 4, Issue 07). https://doi.org/10.1016/S1573-4471(07)04057-0

Ellis, F., & Manda, E. (2012). Seasonal Food Crises and Policy Responses: A Narrative Account of Three Food Security Crises in Malawi. *World Development*, *40*(7), 1407–1417. https://doi.org/10.1016/j.worlddev.2012.03.005

Enke, S. (1951). Equilibrium among Spatially Separated Markets : Solution by Electric Analogue. *Econometrica*, *19*(1), 40–47. https://www.jstor.org/stable/pdf/1907907.pdf?refreqid=excelsior%3A1ac71dc767dc679ca32d1c0cec1832de&ab_segments=&origin=

Fackler, P. L., & Goodwin, B. K. (2001). Chapter 17 Spatial price analysis. In *Marketing, Distribution and Consumers* (Vol. 1, pp. 971–1024). Elsevier. https://doi.org/https://doi.org/10.1016/S1574-0072(01)10025-3

Fafchamps, M., & Gabre-Madhin, E. (2006). Agricultural markets in Benin and Malawi. *AfJARE*, *1*(0), 67–94. https://web.stanford.edu/~fafchamp/afjae.pdf

Fafchamps, M., & Gubert, F. (2007). The formation of risk sharing networks. *Journal of Development Economics*, *83*(2), 326–350. https://doi.org/10.1016/j.jdeveco.2006.05.005

Famine Early Warning Systems Network. (2018). *Malawi Enhanced Market Analysis Report. Washington, DC*. *September*, 144. https://fews.net/southern-africa/malawi/enhanced-market-analysis/september-2018

FAO and FHI 360. (2016). *Minimum Dietary Diversity for Women: A Guide for Measurement* (Issue ISBN 978-92-5-109153-1 (FAO)). http://www.fao.org/3/a-i5486e.pdf

Filmer, D., Friedman, J., Kandpal, E., & Onishi, J. (2021). Cash Transfers, Food Prices, and Nutrition Impacts on Ineligible Children. *The Review of Economics and Statistics*, 1–45. https://doi.org/https://doi.org/10.1162/rest_a_01061

Finkelstein, A. (2007). The aggregate effects of health insurance: evidence from the introduction of Medicare. *Quarterly Journal of Economics*, *122*(1), 1–37. https://doi.org/https://doi.org/10.1162/qjec.122.1.1

Floodlist. (2015). *Malawi Floods – 48 Dead as President Declares State of Disaster World ' s Media Start to Pay Attention*. *February*, 1–2. https://floodlist.com/africa/malawi-floods-48-dead-president-declares-state-disaster

Food and Agriculture Organization of the United Nations. (1968). *Food Composition Table for use in Africa*. https://www.fao.org/3/X6877E/X6877E02.htm

Food and Agriculture Organization of the United Nations. (2016). Influencing food environments for healthy diets. In *Food and agriculture organization of the united* *nations*. http://www.fao.org/3/a-i6484e.pdf

Fuje, H. (2019). Fossil fuel subsidy reforms, spatial market integration, and welfare: Evidence from a natural experiment in Ethiopia. *American Journal of Agricultural Economics*, *101*(1), 270–290. https://doi.org/10.1093/ajae/aay026

Fuje, H. (2020). Transportation Cost, Fuel Subsidies, and Commodity Prices. *World Trade Review*, *19*(S1), S88–S97. https://doi.org/10.1017/S1474745620000336

Funk, C. C., Peterson, P. J., Landsfeld, M. F., Pedreros, D. H., Verdin, J. P., Rowland, J. D., Romero, B. E., Husak, G. J., Michaelsen, J. C., & Verdin, A. P. (2014). A Quasi-Global Precipitation Time Series for Drought Monitoring. *U.S. Geological Survey Data Series*, *832*, 4. https://doi.org/https://doi.org/10.3133/ds832

Garrett, T. (2002). Aggregated Versus Disaggregated Data in Regression Analysis: Implications for Inference. *Economics Letters*, *81*, 61–65. https://doi.org/10.1016/S0165-1765(03)00149-6

Gilbert, C. L., Christiaensen, L., & Kaminski, J. (2017). Food price seasonality in Africa: Measurement and extent. *Food Policy*, *67*, 119–132. https://doi.org/10.1016/j.foodpol.2016.09.016

Goetz, S. J. (1992). A Selectivity Model of Household Food Marketing Behavior in Sub-Saharan Africa. *American Journal of Agricultural Economics*, *74*(2), 444–452. https://doi.org/10.2307/1242498

Gourieroux, A. C., Monfort, A., & Trognon, A. (1984). Pseudo Maximum Likelihood Methods : Theory. *Econometrica*, *52*(3), 681–700. https://doi.org/https://doi.org/10.2307/1913471

Government of Malawi. (2016a). *National Agriculture Policy*. https://cepa.rmportal.net/Library/government-publications/national-agriculture-policy-2016/view

Government of Malawi. (2016b). *National Education Policy*. https://planipolis.iiep.unesco.org/sites/default/files/ressources/national_education_policy.pdf

Government of Malawi. (2016c). *The Malawi Vulnerability Assessment Committee (MVAC): National Food and Nutrition Security Forecast, April 2016 to March 2017.* (Issue July). https://reliefweb.int/report/malawi/malawi-vulnerability-assessment-committee-mvac-national-food-security-forecast-april

Government of Malawi. (2018). *National liquid fuels and gas emergency plan for liquid fuels and gas supply: Vol. LV* (Issue 38). https://gazettes.africa/archive/mw/2018/mw-government-gazette-dated-2018-09-14-no-38.pdf

Government of Malawi. (2020a). *Malawi Poverty Report 2020* (Issue August).

Government of Malawi. (2020b). *The Fifth Integrated Household Survey (IHS5) 2020 Report*. https://doi.org/https://doi.org/10.48529/mpyk-ds48

Government of Malawi. (2021a). *Background to the National Oil Company of Malawi*

(NOCMA) Limited. https://www.nocma.mw/company-background/

Government of Malawi. (2021b). *National Oil Company Considerations*. http://www.reforms.gov.mw/psrmu/national-oil-company-malawi

Government of Malawi. (2022). *Malawi Education Statistics Report: Education Management Information System (EMIS)*. https://www.education.gov.mw/index.php/edu-resources/2022-education-statistics

Graham, B. S. (2017). An Econometric Model of Network Formation With Degree Heterogeneity. *Econometrica*, *85*(4), 1033–1063. https://doi.org/10.3982/ecta12679

Groot, L., & Oostveen, T. (2019). Welfare effects of energy subsidy reform in developing countries. *Review of Development Economics*, *23*(4), 1926–1944. https://doi.org/10.1111/rode.12619

Hasan, S. A. (2016). The impact of the 2005–2010 rice price increase on consumption in rural Bangladesh. *Agricultural Economics*, *47*(4), 423–433. https://doi.org/10.1111/agec.12241

Hausman, J., Hall, B. H., & Griliches, Z. (1984). Econometric Models for Count Data with an Application to the Patents-R & D Relationship. *Econometrica*, *52*(4), 909–938. https://doi.org/https://doi.org/10.2307/1911191

Headey, D., Hirvonen, K., Hoddinott, J., & Stifel, D. (2019). Rural Food Markets and Child Nutrition. *American Journal of Agricultural Economics*, *101*(5), 1311–1327. https://doi.org/10.1093/ajae/aaz032

Heim, R. R. J. (2002). A Review of Twentieth Century Drought Indices Used in the United States. *Bulletin of the American Meteorological Society*, *83*(8), 1149–1166. https://doi.org/https://doi.org/10.1175/1520-0477-83.8.1149

Hill, S., Rothchild, D., & Cameron, C. (1998). Tactical Information and the Diffusion of Protests. In D. A. Lake & D. Rothchild (Eds.), *The International Spread of Ethnic Conflict. Fear, diffusion, and escalation* (pp. 61–88). Princeton University Press. https://doi.org/https://doi.org/10.1515/9780691219752-007

Hirvonen, K., Hoddinott, J., Minten, B., & Stifel, D. (2017). Children's Diets, Nutrition Knowledge, and Access to Markets. *World Development*, *95*, 303–315. https://doi.org/10.1016/j.worlddev.2017.02.031

Hoddinott, J., Headey, D., & Dereje, M. (2015). Cows, Missing Milk Markets, and Nutrition in Rural Ethiopia. *American Journal of Agricultural Economics*, *51*(8), 958–975. https://doi.org/10.1080/00220388.2015.1018903

Hoddinott, J., & Yohannes, Y. (2002). Dietary Diversity as a Household Food Security Indicator. In *Food and Nutrition Technical Assistance Project* (Vol. 136, Issue May). http://www.fantaproject.org/research/dietary-diversity-household-food-security

Innovex Development Consulting Ltd. (2020). *Establishment of Wholesalers' and Retailers' Margins in the Tanzania Petroleum Downstream Industry* (Issue October). https://www.ewura.go.tz/wp-content/uploads/2020/12/Report-on-the-Establishment-of-Wholesalers-and-Retailers-Margins-in-the-Tanzania-Petroleum-Downstream-

Industry.pdf

International Dietary Data Expansion Project. (2018). Household food expenditure share. In *Data4Diets: Building Blocks for Diet-related Food Security Analysis*. https://inddex.nutrition.tufts.edu/printpdf/138

International Labour Office and United Nations Children's Fund. (2021). Child Labour: Global estimates 2020, trends and the road forward. In *ILO and UNICEF* (Vol. 6, Issue 2). https://www.ilo.org/wcmsp5/groups/public/---ed_norm/---ipec/documents/publication/wcms_797515.pdf

International Monetary Fund. (2013). *Energy subsidy reform in Sub-Saharan Africa: experiences and lessons*. https://doi.org/10.5089/9781484366547.087

Ito, T., & Kurosaki, T. (2009). Weather Risk Wages in Kind and the Off-Farm Labor Supply of Agricultural Households in a Developing Country. *American Journal of Agricultural Economics*, *91*(3), 697–710. https://doi.org/10.1111/j.1467-8276.2009.01270.x

Jann, B. (2014). Plotting regression coefficients and other estimates. *Stata Journal*, *14*(4), 708–737. https://doi.org/10.1177/1536867x1401400402

Jayne, T. S. (1994). Do High Food Marketing Costs Constrain Cash Crop Production ? Evidence from Zimbabwe. *Economic Development and Cultural Change*, *42*(2), 387–402. https://doi.org/https://www.journals.uchicago.edu/doi/10.1086/452086

Jayne, T. S., Zulu, B., & Nijhoff, J. J. (2006). Stabilizing food markets in eastern and southern Africa. *Food Policy*, *31*, 328–341. https://doi.org/10.1080/00220388.2014.903244

Jensen, R. (2007). The digital provide: information (technology), market performance, and welfare in the South Indian fisheries sector. *The Quarterly Journal of Economics*, *122*(3). https://doi.org/https://doi.org/10.1162/qjec.122.3.879

Jessoe, K., Manning, D. T., & Taylor, J. E. (2016). Climate Change and Labour Allocation in Rural Mexico: Evidence from Annual Fluctuations in Weather. *Economic Journal*, *128*(608), 230–261. https://doi.org/10.1111/ecoj.12448

Jodha, N. S. (1978). Effectiveness of farmers' adjustments to risk. *Economic and Political Weekly*, *13*(25), 38–48. https://www.jstor.org/stable/4366744

Jones, A. D. (2017). On-farm crop species richness is associated with household diet diversity and quality in subsistence- and market-oriented farming households in Malawi. *Journal of Nutrition*, *147*(1), 86–96. https://doi.org/10.3945/jn.116.235879

Jones, A. D., Shrinivas, A., & Bezner-Kerr, R. (2014). Farm production diversity is associated with greater household dietary diversity in Malawi: Findings from nationally representative data. *Food Policy*, *46*, 1–12. https://doi.org/10.1016/j.foodpol.2014.02.001

Kachule, R., & Franzel, S. (2009). The status of fruit production, processing and marketing in Malawi. *World Agroforestry Centre*, *87*, 4–44. https://doi.org/https://doi.org/10.5716/WP16350.PDF

Kaiyatsa, S., Bai, Y., Schneider, K. R., Herforth, A. W., & Masters, W. A. (2019). Do remote rural people pay higher prices for more nutritious foods? Evidence from 130,975 price observations at rural markets in Malawi, 2007-2017. *2019 Agricultural & Applied Economics Association Annual Meeting*, *1*, 1–2. https://doi.org/.1037//0033-2909.I26.1.78

Kasalika, J. (2013, February 18). Malawi saves K36bn on auto-fuel pricing. *The Nation*. https://mwnation.com/malawi-saves-k36bn-on-auto-fuel-pricing/

Kaunda, C. (2016). *Promotion of Efficient Public Transport in Malawi*. https://wedocs.unep.org/bitstream/item/21643/Promotion of Efficient Public Transport in Malawi%2C Chimwemwe kaunda.pdf?sequence=9

Kaya, C. M., & Derin, L. (2023). Parameters and methods used in flood susceptibility mapping: a review. *Journal of Water and Climate Change*, *14*(6), 1935–1960. https://doi.org/10.2166/wcc.2023.035

Kazakis, N., Kougias, I., & Patsialis, T. (2015). Assessment of flood hazard areas at a regional scale using an index-based approach and Analytical Hierarchy Process: Application in Rhodope-Evros region, Greece. *Science of the Total Environment*, *538*, 555–563. https://doi.org/10.1016/j.scitotenv.2015.08.055

Key, N., Sadoulet, E., & De de Janvry, A. (2000). Transactions costs and agricultural household supply response. *American Journal of Agricultural Economics*, *82*(2), 245–259. https://doi.org/10.1111/0002-9092.00022

Kochar, A. (1999). Smoothing consumption by smoothing income: Hours-of-work responses to idiosyncratic agricultural shocks in Rural India. *Review of Economics and Statistics*, *81*(1), 50–61. https://doi.org/10.1162/003465399767923818

Kojima, M. (2009). Government Response to Oil Price Volatility. *Development*, *July*, 2009. http://siteresources.worldbank.org/INTOGMC/Resources/10-govt_response-hyperlinked.pdf

Kojima, M. (2012). Oil Price Risks and Pump Price Adjustments. *World Bank*, *October*, 1–34. http://rru.worldbank.org/

Kojima, M. (2013). *Petroleum Product Pricing and Complementary Policies Experience of 65 Developing Countries Since 2009* (WPS 6396; Policy Research Working Paper, Issue April). https://documents.worldbank.org/en/publication/documents-reports/documentdetail/287081468139496312/petroleum-product-pricing-and-complementary-policies-experience-of-65-developing-countries-since-2009

Kojima, M., Matthews, W., & Sexsmith, F. (2010). *Petroleum markets in Sub-Saharan Africa : analysis and assessment of 12 countries* (No. 15; Extractive Industries and Development Series, Issue March). http://documents.worldbank.org/curated/en/907171467990361424/Petroleum-markets-in-Sub-Saharan-Africa-analysis-and-assessment-of-12-countries

Koppmair, S., Kassie, M., & Qaim, M. (2017). Farm production, market access and dietary diversity in Malawi. *Public Health Nutrition*, *20*(2), 325–335. https://doi.org/10.1017/S1368980016002135

Kpodar, K., & Djiofack, C. (2009). The distributional effects of oil price changes on household income: Evidence from Mali. *Journal of African Economies*, *19*(2), 205–236. https://doi.org/10.1093/jae/ejp023

Kunaka, C., Raballand, G., & Fitzmaurice, M. (2018). How trucking services have improved and may contribute to economic development: The case of east Africa. In *Industries without Smokestacks: Industrialization in Africa Reconsidered* (Issue December, pp. 133–150). https://doi.org/10.1093/oso/9780198821885.003.0007

Lall, S. V., Wang, H., & Munthali, T. (2009). *Explaining High Transport Costs within Malawi - Bad Roads or Lack of Trucking Competition ? Vol. November* (No. WPS5133; Policy Research Working Paper). https://openknowledge.worldbank.org/server/api/core/bitstreams/aa26b1b9-e942-5fc9-adf0-457d94bd51af/content

Letta, M., Montalbano, P., & Pierre, G. (2021). Weather shocks, traders' expectations, and food prices. *American Journal of Agricultural Economics*, 1–20. https://doi.org/10.1111/ajae.12258

Limão, N., & Venables, A. J. (2001). Infrastructure, Geographical Disadvantage, Transport Costs, and Trade. *The World Bank Economic Review*, *15*(3), 451–479. https://doi.org/https://doi.org/10.1093/wber/15.3.451

Maccini, B. S., & Yang, D. (2009). Under the Weather : Health , Schooling , and Economic Consequences of Early-Life Rainfall. *The American Economic Review*, *99*(3), 1006–1026. https://doi.org/10.1257/aer.99.3.1006

Mahony, C., Clark, H., Bolwell, M., Simcock, T., Potter, R., & Meng, J. (2014). Where Politics Borders Law : The Malawi-Tanzania Boundary Dispute. In *New Zealand Centre for Human Rights Law, Policy and Practice* (Vol. 21, Issue February). https://cdn.auckland.ac.nz/assets/humanrights/Research/MalawiTanzania-NZCHRLPP-final.pdf

Manea, R. E. (2020). School Feeding Programmes, Education and Food Security in Rural Malawi. In *Geneva, Graduate Institute of International and Development Studies, Centre for International Environmental Studies*. https://repository.graduateinstitute.ch/record/298512

Mason, N. M., & Ricker-Gilbert, J. (2013). Disrupting Demand for Commercial Seed: Input Subsidies in Malawi and Zambia. *World Development*, *45*, 75–91. https://doi.org/10.1016/j.worlddev.2012.11.006

Matita, M., Chirwa, E. W., Johnston, D., Mazalale, J., Smith, R., & Walls, H. (2021). Does household participation in food markets increase dietary diversity? Evidence from rural Malawi. *Global Food Security*, *28*(January 2020), 100486. https://doi.org/10.1016/j.gfs.2020.100486

McCarthy, N., Kilic, T., de la Fuente, A., & Brubaker, J. (2017). Shelter from the Storm? Household-Level Impacts of, and Responses to, the 2015 Floods in Malawi. In *Policy Research Working Paper* (No. 8189; Issue September). https://doi.org/10.1596/1813-9450-8189

McLaughlin, S. M., Bozzola, M., & Nugent, A. (2023). Changing Climate, Changing Food

Consumption? Impact of Weather Shocks on Nutrition in Malawi. *Journal of Development Studies*, *0*(0), 1–22. https://doi.org/10.1080/00220388.2023.2244634

Minot, N. (2010). Staple food prices in Malawi. *Variation in Staple Food Prices: Causes, Consequence, and Policy Options*, *January 2010*. https://www.researchgate.net/publication/46470944_Staple_Food_Prices_in_Malawi

Minot, N., & Dewina, R. (2015). Are we overestimating the negative impact of higher food prices? Evidence from Ghana. *Agricultural Economics*, *46*(4), 579–593. https://doi.org/10.1111/agec.12183

Minot, N., & Goletti, F. (1998). Export Liberalization and Household Welfare: The Case of Rice in Vietnam. *American Journal of Agricultural Economics*, *80*(4), 738–749. https://doi.org/10.2307/1244060

Minten, B., Koru, B., & Stifel, D. (2013). The last mile(s) in modern input distribution: Pricing, profitability, and adoption. *Agricultural Economics*, *44*(6), 629–646. https://doi.org/10.1111/agec.12078

Minten, B., & Kyle, S. (1999). The effect of distance and road quality on food collection, marketing margins, and traders' wages: Evidence from the former Zaire. *Journal of Development Economics*, *60*(2), 467–495. https://doi.org/10.1016/S0304-3878(99)00049-8

Mishra, A. K., & Goodwin, B. K. (1997). Farm Income Variability and the Supply of Off-Farm Labor. *American Journal of Agricultural Economics*, *79*(3), 880–887. https://doi.org/10.2307/1244429

Monsivais, P., Mclain, J., & Drewnowski, A. (2010). The rising disparity in the price of healthful foods: 2004-2008. *Food Policy*, *35*(6), 514–520. https://doi.org/10.1016/j.foodpol.2010.06.004

Moore, D. E. (1998). *MVTEST: Stata module to perform multivariate F tests*. https://econpapers.repec.org/RePEc:boc:bocode:s345804

Mu, R., & van de Walle, D. (2011). Rural roads and local market development in Vietnam. *Journal of Development Studies*, *47*(5), 709–734. https://doi.org/10.1080/00220381003599436

Mundlak, Y. (1978). On the Pooling of Time Series and Cross Section Data. In *Econometrica* (Vol. 46, Issue 1, p. 69). https://doi.org/10.2307/1913646

Nandi, R., Nedumaran, S., & Ravula, P. (2021). The interplay between food market access and farm household dietary diversity in low and middle income countries: A systematic review of literature. *Global Food Security*, *28*, 100484. https://doi.org/10.1016/j.gfs.2020.100484

Nankhuni, F. J., & Findeis, J. L. (2004). Natural resource-collection work and children's schooling in Malawi. *Agricultural Economics*, *31*(2-3 SPEC. ISS.), 123–134. https://doi.org/10.1016/j.agecon.2004.09.022

National Statistical Office. (2013a). *Malawi Integrated Household Panel Survey (IHPS). Geovariables:Description*.

https://www.unhcr.org/publications/manuals/4d9352319/unhcr-protection-training-manual-european-border-entry-officials-2-legal.html?query=excom 1989

National Statistical Office. (2013b). *Malawi Labour Force Survey 2013*. https://webapps.ilo.org/surveyLib/index.php/catalog/6870

National Statistical Office. (2014a). *Integrated Household Panel Survey 2010-2013 Household socio-economic characteristics report*. http://siteresources.worldbank.org/INTLSMS/Resources/3358986-1233781970982/5800988-1271185595871/6964312-1404828635943/IHPS_BID_FINAL.pdf

National Statistical Office. (2014b). *Malawi Integrated Household Panel Survey (IHPS) 2013 Basic Information Document* (Issue September). https://microdata.worldbank.org/index.php/catalog/2248/download/41259

National Statistical Office. (2017). *Integrated Household Survey 2016-2017: Household socio-economic characteristics report* (Issue November). http://www.nsomalawi.mw/images/stories/data_on_line/economics/ihs/IHS4/IHS4 REPORT.pdf

National Statistical Office of Malawi. (2017). Malawi Integrated Household Panel Survey (IHPS) 2016: Basic Information Document. *Minister of Economic Planning and Development, Malawi Government*, *September*, 1–133. https://microdata.worldbank.org/index.php/catalog/2939/download/48116

Nguyen, G., & Nguyen, T. T. (2020). Exposure to weather shocks: A comparison between self-reported record and extreme weather data. *Economic Analysis and Policy*, *65*, 117–138. https://doi.org/10.1016/j.eap.2019.11.009

Nguyen, T., Nguyen, T. T., & Grote, U. (2020). Multiple shocks and households' choice of coping strategies in rural Cambodia. *Ecological Economics*, *167*(January 2019), 106442. https://doi.org/10.1016/j.ecolecon.2019.106442

Nikoloski, Z., Christiaensen, L., & Hill, R. (2018). Household shocks and coping mechanism: evidence from Sub-Saharan Africa. In *Agriculture in Africa: Telling Myths from Facts. Directions in Development—Agriculture and Rural Development* (Issue November 2017, pp. 123–134). World Bank. https://core.ac.uk/download/pdf/132202311.pdf

Nkegbe, P. K., & Abdul Mumin, Y. (2022). Impact of community development initiatives and access to community markets on household food security and nutrition in Ghana. *Food Policy*, *May*, 102282. https://doi.org/10.1016/j.foodpol.2022.102282

Nordman, C. J., Sharma, S., & Sunder, N. (2022). Here Comes the Rain Again : Productivity Shocks , Educational Investments , and Child Work. *Economic Development and Cultural Change*, *70*(3). https://doi.org/https://doi.org/10.1086/713937

Nübler, L., Austrian, K., Maluccio, J. A., & Pinchoff, J. (2021). Rainfall shocks, cognitive development and educational attainment among adolescents in a drought-prone region in Kenya. *Environment and Development Economics*, *26*(5–6), 466–487. https://doi.org/10.1017/S1355770X20000406

Nyama World. (2017). *Nyama World Case Study*.

https://www.norfund.no/app/uploads/2020/04/Nyama-World-development-impact-case-study.pdf

Ochieng, D. O., Botha, R., & Baulch, B. (2019). Structure, conduct and performance of maize markets in Malawi. In *MaSSP Working Paper - Malawi Strategy Support Program* (Issue No.29). https://cgspace.cgiar.org/server/api/core/bitstreams/c9b811e9-da38-4d5b-a172-d7f52f0f0acb/content

Olabisi, M., Obekpa, H. O., & Liverpool-Tasie, S. (2021). Is growing your own food necessary for dietary diversity? Evidence from Nigeria. *Food Policy*, *104*(July), 102144. https://doi.org/10.1016/j.foodpol.2021.102144

Omamo, S. W. (1998). Transport Costs and Smallholder Cropping Choices: An Application to Siaya District, Kenya. *American Journal of Agricultural Economics*, *80*(1), 116–123. https://doi.org/10.2307/3180274

Orazem, P. F., & Gunnarsson, V. (2003). *Child labour, school attendance and academic performance: A review* (International Labour Office/International Programme on the Elimination of Child Labour). http://ilo-mirror.library.cornell.edu/public/english/standards/ipec/publ/policy/edu_schoolattend_2003.pdf

Pauw, K., Beck, U., & Mussa, R. (2016). Did rapid smallholder-led agricultural growth fail to reduce rural poverty? Making sense of Malawi's poverty puzzle. In C. Arndt, A. McKay, & F. Tarp (Eds.), *Growth and Poverty in Sub-Saharan Africa* (pp. 1–10). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198744795.003.0001

Pauw, K., Thurlow, J., Bachu, M., & Van Seventer, D. E. (2011). The economic costs of extreme weather events: A hydrometeorological CGE analysis for Malawi. *Environment and Development Economics*, *16*(2), 177–198. https://doi.org/10.1017/S1355770X10000471

Pauw, K., Verduzco-gallo, I., & Ecker, O. (2015). Agriculture , Food Security & Nutrition. In N. Aberman & T. Roopnaraine (Eds.), *Mapping the Linkages Between Agriculture, Food Security & Nutrition in Malawi* (pp. 1–11). https://massp.ifpri.info/files/2015/05/Full-report_4-20-15_2sm.pdf

Pierce, J. R., & Schott, P. K. (2016). American Economic Association The Surprisingly Swift Decline of US Manufacturing Employment. *American Economic Review*, *106*(7), 1632–1662. https://doi.org/10.1257/aer.20131578

Pondani, M. (2022, March 30). Malawi importing 20 , 000mt bananas. *The Daily Times*, 4–6. https://times.mw/malawi-importing-20000mt-bananas/

Pratik, D., & Sar, J. (2020). Identification and validation of potential flood hazard areausing GIS-based multi-criteria analysis and satellitedata-derived water index. *Journal of Flood Risk Management*.

Psaki, S. R., McCarthy, K. J., & Mensch, B. S. (2018). Measuring Gender Equality in Education: Lessons from Trends in 43 Countries. *Population and Development Review*, *44*(1), 117–142. https://doi.org/10.1111/padr.12121

Randell, H., & Gray, C. (2016). Climate variability and educational attainment: Evidence from rural Ethiopia. *Global Environmental Change*, *41*, 111–123. https://doi.org/10.1016/j.gloenvcha.2016.09.006

Ray, R. (2000a). Analysis of child labour in Peru and Pakistan: A comparative study. *Journal of Population Economics*, *13*(1), 3–19. https://doi.org/10.1007/s001480050119

Ray, R. (2000b). Child labor, child schooling, and their interaction with adult labor: Empirical evidence for Peru and Pakistan. *World Bank Economic Review*, *14*(2), 347–367. https://doi.org/10.1093/wber/14.2.347

Ricker-Gilbert, J., Jumbe, C., & Chamberlin, J. (2014). How does population density influence agricultural intensification and productivity? Evidence from Malawi. *Food Policy*, *48*, 114–128. https://doi.org/10.1016/j.foodpol.2014.02.006

Rideout, K., Mah, C. L., & Minaker, L. M. (2015). *Food Environments: An Introduction for Public Health Practice*. *December*, 1–7. http://www.ncceh.ca/documents/evidence-review/food-environments-introduction-public-health-practice

Rizet, C., & Gwet, H. (1998). Transport de Marchandises: Une Comparaison Internationale des Prix du Camionnage—Afrique, Asie du Sud Est, Amérique Centrale. *Recherche–Transports–Sécurités*, *60*(1), 69–85. https://doi.org/https://doi.org/10.1016/S0761-8980(98)90040-1

Rizet, C., & Hine, J. L. (1993). A comparison of the costs and productivity of road freight transport in africa and pakistan: Foreign summaries. *Transport Reviews*, *13*(2), 151–165. https://doi.org/10.1080/01441649308716841

Robinson, B., & Wakeford, J. (2013). *Oil shock vulnerabilities and impacts: Case Study of Malawi*. *June*, 25 pp. http://r4d.dfid.gov.uk/Output/193547/Default.aspx

Roehner, B. M. (1996). The Role of Transportation Costs in the Economics of Commodity Markets The Role of Transportation Costs in the Economics of Commodity Markets. *American Journal of Agricultural Economics*, *78*(2), 339–353. https://www.jstor.org/stable/1243707

Rosales-Rueda, M. (2018). The impact of early life shocks on human capital formation: evidence from El Niño floods in Ecuador. *Journal of Health Economics*, *62*, 13–44. https://doi.org/10.1016/j.jhealeco.2018.07.003

Rose, E. (2001). Ex ante and ex post labor supply response to risk in a low-income area. *Journal of Development Economics*, *64*(2), 371–388. https://doi.org/10.1016/S0304-3878(00)00142-5

Rosenzweig, B. M. R., & Schultz, T. P. (1982). Market Opportunities , Genetic Endowments , and Intrafamily Resource Distribution : Child Survival in Rural India. *The American Economic Review*, *72*(4), 803–815. https://www.jstor.org/stable/1810018

Ruel, M. T., Garrett, J. L., Hawkes, C., & Cohen, M. J. (2010). The food, fuel, and financial crises affect the urban and rural poor disproportionately: A review of the evidence. *Journal of Nutrition*, *140*(1). https://doi.org/10.3945/jn.109.110791

Saha, A. (1994). A two-season agricultural household model of output and price uncertainty. *Journal of Development Economics*, *45*(2), 245–269. https://doi.org/10.1016/0304-

3878(94)90032-9

Salazar-Espinoza, C., Jones, S., & Tarp, F. (2015). Weather shocks and cropland decisions in rural Mozambique. *Food Policy*, *53*, 9–21. https://doi.org/10.1016/j.foodpol.2015.03.003

Salazar, C., Ayalew, H., & Fisker, P. (2019). Weather Shocks and Spatial Market Efficiency: Evidence from Mozambique. *Journal of Development Studies*, *55*(9), 1967–1982. https://doi.org/10.1080/00220388.2018.1528352

Samuelson, P. A. (1952). Spatial Price Equilibrium and Linear Programming. *The American Economic Review Economic*, *42*(3), 283–303. https://www.jstor.org/stable/1810381

Santos Silva, J. M. C., & Tenreyro, S. (2006). The log of gravity. *The Review of Economics and Statistics*, *88*(4), 641–658. https://doi.org/10.1080/00036846.2011.599786

Schmidt, R. (1969). Cattle in Malawi's southern region. *The Society of Malawi Journal*, *22*(2), 57–72. https://www.jstor.org/stable/pdf/29778215

Schneider, K. R., Christiaensen, L., Webb, P., & Masters, W. A. (2023). Assessing the affordability of nutrient-adequate diets. *American Journal of Agricultural Economics*, *105*(2), 503–524. https://doi.org/https://doi.org/10.1111/ajae.12334

Sesmero, J., Ricker-Gilbert, J., & Cook, A. (2018). How do african farm households respond to changes in current and past weather patterns? A structural panel data analysis from Malawi. *American Journal of Agricultural Economics*, *100*(1), 115–144. https://doi.org/10.1093/ajae/aax068

Shah, M., & Steinberg, B. M. (2017). Drought of Opportunities : Contemporaneous and Long-Term Impacts of Rainfall Shocks on Human Capital. *Journal of Political Economy*, *125*(2). https://www.journals.uchicago.edu/doi/abs/10.1086/690828?journalCode=jpe

Sheahan, M., & Barrett, C. B. (2017). Ten striking facts about agricultural input use in Sub-Saharan Africa. *Food Policy*, *67*, 12–25. https://doi.org/10.1016/j.foodpol.2016.09.010

Sibande, L., Bailey, A., & Davidova, S. (2017). The impact of farm input subsidies on maize marketing in Malawi. *Food Policy*, *69*, 190–206. https://doi.org/10.1016/j.foodpol.2017.04.001

Sibhatu, K. T., Krishna, V. V., & Qaim, M. (2015). Production diversity and dietary diversity in smallholder farm households. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(34), 10657–10662. https://doi.org/10.1073/pnas.1510982112

Sibhatu, K. T., & Qaim, M. (2017). Rural food security, subsistence agriculture, and seasonality. *American Journal of Agricultural Economics*, *12*(10), 1–15. https://doi.org/10.1371/journal.pone.0186406

Smith, L. C., & Subandoro, A. (2007). Measuring Food Security Using Household Expenditure Surveys. In *Food Security in Practice technical guide series*. https://doi.org/10.2499/0896297675

Snapp, S. S., & Fisher, M. (2015). "Filling the maize basket" supports crop diversity and quality of household diet in Malawi. *Food Security*, *7*(1), 83–96. https://doi.org/10.1007/s12571-014-0410-0

Stifel, D., & Minten, B. (2017). Market Access, Well-being, and Nutrition: Evidence from Ethiopia. *World Development*, *90*, 229–241. https://doi.org/10.1016/j.worlddev.2016.09.009

Storeygard, A. (2016). *Farther on down the Road : Transport Costs , Trade and Urban Growth in Sub-Saharan Africa*. *April*, 1263–1295. https://doi.org/10.1093/restud/rdw020

Sun, L., & Shapiro, J. (2022). A Linear Panel Model with Heterogeneous Coefficients and Variation in Exposure. *Journal of Economic Perspectives*, *36*(4), 193–204. https://doi.org/doi.org/10.1257/jep.36.4.193

Swindale, A. (2005). Measuring Household Food Consumption: A Technical Guide. *Measurement*, *53*(12), 1967–1979. http://www.sciencedirect.com/science/article/B6VRG-4VS40F6-1/2/561070ef0adc5144d6bb75c60a666551

Swindale, A., & Bilinsky, P. (2006). Household Dietary Diversity Score (HDDS) for measurement of household food access: Indicator guide. *Food and Nutrition Technical Assistance*, 11. https://doi.org/10.1017/CBO9781107415324.004

Takayama, T., & Judge, G. G. (1971). *Spatial and temporal price allocation models*. Amsterdam, North-Holland.

Teravaninthorn, S., & Raballand, G. (2009). *Infrastructure Transport Prices and Costs in Africa. A Review of the International Corridors*. https://openknowledge.worldbank.org/bitstream/handle/10986/6610/461810PUB0Box3101OFFICIAL0USE0ONLY1.pdf?sequence=1

Thomson Reuters. (2022). *Spot Prices for Crude Oil and Petroleum Products*. http://www.eia.gov/dnav/pet/pet_pri_spt_s1_d.htm%0A

Tobacco Reporter. (2023, June 1). From Imports to Orchards Rebuilding the banana value chain. *Tobacco Reporter*, 10–12. https://tobaccoreporter.com/2023/06/01/from-imports-to-orchards/

Tobin, J. (1958). Liquidity preference as behavior towards risk. *Review of Economic Studies*, *25*(2), 65–86. https://doi.org/10.2307/2296205

Tran, N., Peart, J., Rodriguez, U., Chan, C., Hoong, Y., Siamudaala, V., Njaya, F., & Kafakoma, R. (2022). *The future of fish supply-demand in Malawi and its implications for nutrition security and poverty reduction The future of fish supply-demand in Malawi and its implications for nutrition security and poverty reduction* (No. 2022–21). https://digitalarchive.worldfishcenter.org/bitstream/handle/20.500.12348/5358/0ab45416ac19a2926e2c4cbae24d6ff3.pdf?sequence=2

Trivedi, P. K. (2010). Recent Developments in Panel Models for Count Data. *Mexican Stata Userts Group Meeting*, 1–79. https://core.ac.uk/download/pdf/6306771.pdf

UNICEF. (2018). *Education is a key investment for child survival, growth, development and well-being*. https://www.unicef.org/esa/sites/unicef.org.esa/files/2019-05/UNICEF-

Malawi-2018-Costing-Quality-Secondary-Education-for-all-Children.pdf

United Nations Children's Fund-Malawi. (2022). Malawi Education Fact Sheets 2022: Analyses for learning and equity using MICS data. In *MICS-Eagle*. https://data.unicef.org/wp-content/uploads/2022/12/2022Malawi_Factsheet_InDesign-FINAL-2.pdf

United States Department of Labor. (2022). Findings on the Worst Forms of Child Labor. In *Child Labor and Forced Labor Reports*. https://www.dol.gov/sites/dolgov/files/ILAB/child_labor_reports/tda2022/2022-Findings-on-the-Worst-Forms-of-Child-Labor.pdf

Usman, M. A., & Haile, M. G. (2022). Market access, household dietary diversity and food security: Evidence from Eastern Africa. *Food Policy*, *113*(September 2021), 102374. https://doi.org/10.1016/j.foodpol.2022.102374

Vandercasteelen, J., Beyene, S. T., Minten, B., & Swinnen, J. (2018). Big cities, small towns, and poor farmers: Evidence from Ethiopia. *World Development*, *106*, 393–406. https://doi.org/10.1016/j.worlddev.2018.03.006

Vicente-Serrano, S. M., Beguería, S., & López-Moreno, J. I. (2010). A multiscalar drought index sensitive to global warming: The standardized precipitation evapotranspiration index. *Journal of Climate*, *23*(7), 1696–1718. https://doi.org/10.1175/2009JCLI2909.1

Volpe, R., Roeger, E., & Leibtag, E. (2013). How Transportation Costs Affect Fresh Fruit and Vegetable Prices. *ERR-160, U.S. Department of Agriculture, Economic Research Service*, *November*, 1–38. https://www.ers.usda.gov/webdocs/publications/45165/41077_err160.pdf

Weidinger, M. (2021). *Agricultural child labour in response to climate change* (Issue September) [Maastricht University]. https://mathiasweidinger.com/mppthesis_final.pdf

Windmeijer, A. G. F., & Santos Silva, J. M. C. (1997). Endogeneity in Count Data Models : An Application to Demand for Health Care. *Journal of Applied Econometrics*, *12*(3), 281–294. https://www.jstor.org/stable/pdf/2285250.pdf

Wooldridge, J. M. (1999). Distribution-free estimation of some nonlinear panel data models. *Journal of Econometrics*, *90*, 77–97. https://doi.org/https://doi.org/10.1016/S0304-4076(98)00033-5

Wooldridge, J. M. (2005). Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity. *Journal of Applied Econometrics*, *20*(1), 39–54. https://doi.org/10.1002/jae.770

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data. 2nd ed. Cambridge, Mass.: MIT Press.* 2010.

Wooldridge, J. M. (2019). Correlated random effects models with unbalanced panels. *Journal of Econometrics*, *211*(1), 137–150. https://doi.org/10.1016/j.jeconom.2018.12.010

Wooldridge, J. M. (2020). *Introductory Econometrics: A Modern Approach* (7th ed.). Cengage

Learning, Inc.

Wooldridge, J. M. (2021). *Two-Way Fixed Effects , the Two-Way Mundlak Regression , and Difference-in-Differences Estimators*. 1–77. https://dx.doi.org/10.2139/ssrn.3906345

Workman, J. P., Phiri, P. H. W., & Coppock, D. L. (1998). Cattle marketing in Malawi. *Rangelands*, *20*(1), 14–16. http://hdl.handle.net/10150/639109

World Bank. (2010). *Note on Conversion Factors for Food Item-Non-standard Measurement Unit Combinations In the Malawi Third Integrated Household Survey (IHS3) 2010/11 Data*. https://microdata.worldbank.org/index.php/catalog/1003/download/40802

World Bank Group, United Nations, & European Union. (2016). *Malawi Drought 2015-2016: Post-Disaster Needs Assessment.* http://hdl.handle.net/10986/25781

World Food Programme. (2009). Market Analysis Tool - How to Estimate Household Net-Seller / Buyer Status and the Welfare Impact of Shocks? In *World Food Programme Division of Communications, Public Policy and Private Partnerships* (Issue December). https://documents.wfp.org/stellent/groups/public/documents/manual_guide_proced/wfp214630.pdf?_ga=2.109784517.1743320223.1724108794-2137537041.1710933039

World Food Programme. (2021). *School Feeding* (Issue May). https://docs.wfp.org/api/documents/WFP-0000131157/download/?_ga=2.179308582.360811445.1710933039-2137537041.1710933039

Xia, F., & Deininger, K. (2019). Spillover Effects of Tobacco Farms on the Labor Supply, Education, and Health of Children: Evidence from Malawi. *American Journal of Agricultural Economics*, *101*(4), 1181–1202. https://doi.org/10.1093/ajae/aaz011

Zanello, G., Shankar, B., & Poole, N. (2019). Buy or make? Agricultural production diversity, markets and dietary diversity in Afghanistan. *Food Policy*, *87*(July 2018), 101731. https://doi.org/10.1016/j.foodpol.2019.101731

Zant, W. (2013). How is the liberalization of food markets progressing? Market integration and transaction costs in subsistence economies. *World Bank Economic Review*, *27*(1), 28–54. https://doi.org/10.1093/wber/lhs017

Zant, W. (2018). Trains, trade, and transaction costs: How does domestic trade by rail affect market prices of Malawi agricultural commodities? *World Bank Economic Review*, *32*(2), 334–356. https://doi.org/10.1093/wber/lhx011

Zimmermann, L. (2020). Remember when it rained – Schooling responses to shocks in India. *World Development*, *126*, 104705. https://doi.org/10.1016/j.worlddev.2019.104705