
Essays on Transaction Costs and Food Diversity in Developing Countries

Dissertation

to obtain the Ph.D. degree

In the International Ph.D. Program for Agricultural Sciences in Goettingen (IPAG)

at the Faculty of Agricultural Sciences,

Georg-August-University Goettingen, Germany



presented by

Christoph David Steffen

born in: Kassel

Goettingen, July 2017

D7

1st Supervisor: Prof. Xiaohua Yu, PhD

2nd Supervisor: Prof. Dr. Bernhard Brümmer

3rd Supervisor: Prof. Dr. Stephan v. Cramon-Taubadel

Date of Dissertation: 13.07.2017

Essays on Transaction Costs and Food Diversity in Developing Countries

by

Christoph David Steffen

Summary:

This thesis addresses 3 different aspects of food and poverty related problems in developing countries. The first essay presents a new operational concept of transaction costs that firstly allows assessing the magnitude and secondly recognizing the non homogeneity of food products. This is realized by providing an estimate of the value of the good by means of a hedonic food price model. A model is proposed that decomposes unit values into spatial price factors and a value component that allows the comparison with a feasible value occurring in a situation without transaction costs. The model is estimated with a conditional mean stochastic frontier approach using data from Kenyan maize farmers. We find a magnitude of 12-18% for maize transactions in rural Kenya and identify drive time, market distance, education and counterparts in negotiations as main determinants.

The second essay is concerned with the latent demand structure for food diversity in India using data from 68th round of the CES Consumer Survey. We assume that consumers facing subsistence concerns favor calories over food diversity and once passing the subsistence threshold substitute away from staples towards a more varied diet. Latent classes and consumption patterns are identified by means of finite mixture models. Therefore we examine the link between food diversity indices and socioeconomic indicators and explain component memberships in order to characterize latent classes and evaluate nutritional implications. Two clearly distinct demand patterns for diversity can be identified, consistent with our assumptions. The identified classes differ substantially in terms of income, household composition and nutritional adequacy ratios

The third essay is concerned with the inference on nutrition from observed consumption. Measures of diversity have become popular tools to infer on nutritional adequacy from observed consumption. However the most common measures do not consider that equal distribution of food consumption does not reflect an optimal diet. The proposed index in this essay adjusts the existing concept of the healthy diversity index so that it is applicable for Indian dietary analysis and extends it for the analysis of household data. The results show that the modified HFD index is a superior predictor of nutritional adequacy compared to common measures like the Berry, Entropy or count index.

Contents

1	General Introduction and Summary	1
1.1	Transaction Costs and Smallholders in Developing Countries	1
1.2	Malnutrition and Food Diversity in Developing Countries	4
1.3	Summaries of Essays	7
2	Estimation of Transaction Costs for Maize Markets in Rural Kenya	10
2.1	Introduction	1
2.2	Literature Review	2
2.3	Theory	5
2.4	Econometric Model	7
2.4.1	Stochastic Frontier Model	7
2.4.2	Selection Bias	9
2.5	Data	10
2.6	Empirical Results	14
2.7	Conclusions	16
2.8	Appendix A	18
3	A Latent Class Analysis of the Demand for Food Diversity in India	26
3.1	Literature Review	28
3.2	Theory	30
3.3	Data	33
3.4	Empirical Model	34
3.5	Variables	37
3.6	Results	38

3.6.1	Model Selection	38
3.6.2	Characterization of Latent Classes	39
3.6.3	Demand Model Estimations	40
3.7	Conclusions	42
3.8	Appendix B	44
4	A Healthy Food Diversity Index for India	54
4.1	Developing a Healthy Food Diversity Index for India	56
4.2	Data	61
4.3	Empirical Strategy	62
4.3.1	Correlation with Nutrient Adequacy Ratios	63
4.3.2	Local Polynomial Regressions	64
4.3.3	Descriptive Statistics	64
4.4	Conclusion and Discussion	65
4.5	Appendix C	67
5	Conclusions	73

List of Tables

2.1	Characteristics of transactions and products	18
2.2	Household characteristics	19
2.3	Proxies for different types of transaction costs	20
2.4	Probit model on market participation	21
2.5	Stochastic frontier model for transaction costs (BC 1995)	22
2.6	Stochastic frontier model for transaction costs (Green, 2010)	23
2.7	Relative and absolute transaction costs	24
3.1	Expected class differences	44
3.2	Model selection criteria	45
3.3	Parameter estimates of the demand model	46
3.4	Average marginal effects of the demand model	47
3.5	Regression on posterior probabilities of component 1	48
3.6	Characterization of latent classes	49
3.7	Variable description	50
4.1	Balanced diet for gender and age groups (number of portions)	67
4.2	Recommended dietary allowances for indians by gender and age	68
4.3	Pearson correlation coefficients between nutrient adequacy ratios and diversity indices	69

List of Figures

2.1	Illustration of transaction costs as distance between observed and feasible price . . .	25
2.2	Kernel density estimates of transaction costs	25
3.1	Expected relation between income and food diversity	51
3.2	Local Polynomial Regression of the Staple Calorie Share on Log Income	52
3.3	Local Polynomial Regression of Posterior Probabilities on Mean Nutrient Adequacy	52
3.4	Histogram of the count index	53
3.5	Histogram of the count index by component	53
4.1	Kernel-weighted local polynomial regressions with 0.99 CI for group based indicators	70
4.2	Kernel-weighted local polynomial regressions with 0.99 CI for item based indicators	71
4.3	Income and food diversity	72
4.4	Kernel density estimates for rural and urban areas	72

Chapter 1

General Introduction and Summary

In 2015 the sustainable development goals also called Agenda 2030 were adopted by the Un General Assembly. The resolution represents the predecessor of the millennium development goals. The first two of new development goals are phrased “*End poverty in all its forms everywhere*”. And “*Zero hunger: end hunger, achieve food security and improved nutrition and promote sustainable agriculture*”. The new goals hence maintain the target to end poverty and hunger but direct more emphasis towards nutrition and agriculture. While extreme poverty has been greatly reduced there still remain 836 million people living in extreme poverty on less than 1.25 Dollars per day and 795 million are undernourished worldwide. This leaves plenty of challenges for researchers to develop a deeper understanding of the underlying problems.

1.1 Transaction Costs and Smallholders in Developing Countries

According to The World Bank (2016) 80% of the poor living at US\$1.90 poverty line live in rural areas of which 64% work in agriculture. Agricultural workers are over four times more likely than people employed in other sectors of the economy to be poor. The Development for Agriculture Report (The World Bank, 2008) estimates that out of 2.5 billion people involved in agriculture 1.5 Billion people are living in smallhold households. Today agriculture is still a key sector for sustainable development and poverty reduction.

Yet in developing countries agricultural goods remain imperfectly tradable due to high transaction costs. Market conditions determine by large the livelihood strategies of smallholders and the partic-

ipation in rural markets often fails as it promises no profitable outcomes. Producers are confronted with high transaction costs, insufficient and unequal access to information, imperfect competition, externalities, and insufficient provision of public goods. Especially in rural areas transaction costs and asymmetric information typically increase with lower population densities. Reducing transaction costs in food markets could substantially contribute to promote faster growth and benefit the poor.

The development of infrastructure has a major impact on rural development. It supports the relocation of labor, influences transportation costs and enhances access to markets and services. Road infrastructure is of particular importance. Access to roads, especially paved roads that can be used independent of the weather conditions, improve access to markets. This enables smallholders to charge better prices for their produce. Improved road networks can help to reduce transportation costs and risks but also contribute to the development of new local markets that benefit nearby smallholders. Vakis, Sadoulet and Janvry (2003) find that buyers may be more willing to come to the farmgate if the road conditions are better which reduces costs for the seller. Additionally a close by road can increase the value of farm land which eases access to credit for farmers.

Communication infrastructure is essential to obtain information and establish contact to traders and service providers. Smallholder households are dependent on good information to develop strategies and make business related decisions. New information technologies can help address some of these information disadvantages. Vakis, Sadoulet and Janvry (2003) evaluate the influence of price information and conclude that knowledge on the price significantly affects the choice of the market channel, the odds to sell the product on this market and increases the net price received.

Stable electricity networks support mechanization and the use of information technology and hence contribute improve output efficiency. The same is true for the access to water as irrigation land can be twice as productive as rainfed land (Rapsomanikis, 2015).

The provision of legal, financial and regulatory institutions decisively shapes the business environment. For smallholders this does for instance affect producer incentives, credit decisions, contract agreements and other aspects of their business. Weak institutions may however result in market failure. For instance undefined or weak property rights impede land sales and limit access to credit due to a lack of formal collateral (De Janvry and Sadoulet, 1998) Governments play an essential role in the provision of such public goods and failing to do so can increase costs to a degree that renders

business opportunities unprofitable. Ndulu (2006) states that costs of contract enforcement difficulties, inadequate infrastructure, crime, corruption and regulation can amount to over 25% of sales.

In most African countries high transaction costs impede food security among other factors such as low growth rates in agriculture and high population growth. The FAO (1996) pronounced that *"food security exists when all people, at all times, have physical and economic access to sufficient, safe, and nutritious food to meet their dietary needs and food preferences"*. This situation is however hard to establish in the presence of food market failures which can only partially be compensated by food imports and aid. High transaction costs thus affect well-being of poor households not only as producers but also consumers and contribute to the prevailing state of malnutrition.

Transaction costs economics recognizes that transactions do not occur in a frictionless environment but there are costs to be associated with an exchange. They are also referred to as institutional costs. As to the nature of institutions North (1991) defines that *"Institutions are the humanly devised constraints that structure political, economic and social interaction. They consist of both informal constraints (sanctions, taboos, customs, traditions, and codes of conduct), and formal rules (constitutions, laws, property rights)*. A common way to divide transaction costs into broad categories is to differ information, bargaining and monitoring costs (Furubotn and Richter, 2005). Information costs usually occur prior to the exchange and include the search for price information and potential buyers as well as the facilitation of the transactions. Bargaining, also labeled negotiation costs usually occur during an exchange and refer to all costs that are connected to the process of negotiation and reaching an agreement as well as costs related to payments. Monitoring or enforcement costs are costs that occur to ensure that the agreement is fulfilled.

Although the early stages of this literature reach back to 1937 when Coase (1937) described transaction costs as the costs of using the price mechanism, Masten (2000) emphasizes that transaction costs economics is just emerging from its infancy and agricultural transactions are a vastly unexplored area for applications and theory. Meanwhile there is a considerable body of literature that highlights the importance of this topic however very few works that address the theoretical and empirical challenges associated with transaction costs in food markets specifically.

1.2 Malnutrition and Food Diversity in Developing Countries

In 2015 the UN evaluated the successes of the Millennium Development Goals(MDG) that were presented in 2001 and adopted by 189 countries. The first goal on the agenda was to *"Eradicate extreme poverty and hunger"* with target 1.C *"Halve, between 1990 and 2015, the proportion of people who suffer from hunger"*. The estimates suggest a head count of 795 million globally undernourished people of which 780 million live in developing countries. The proportion dropped from 23.3% in 1990 to 12.9% in 2015 which is close but slightly below the targeted proportion. The Sustainable Development Goals, that were presented by the UN in 2015 target a broader scope of nutrition while the MDGs were focused on undernutrition. The FAO (2010) stated that more holistic approaches would be favorable since the only sustainable way of achieving these goals is to fight malnutrition in all its forms – undernutrition, micronutrient deficiencies, and overweight and obesity. Figures from the latest FAO report (2013) show that each of these forms is of global significance. In 2013 an estimated 26 percent of the world's children are stunted due to undernutrition. 2 billion people suffer from micronutrient deficiencies and 1.4 billion people are overweight, 500 million of them are obese.

The term undernourishment refers to the situation where food intake is insufficient to meet dietary energy requirements while the term undernutrition describes the outcome of insufficient food intake which can for instance be measured with a body mass index. Stunting is caused by undernutrition of and recurrent infections during early childhood or pre birth through an undernourished mother. It leads to permanent impaired cognitive functions and stunted growth.

Micronutrient deficiencies or malnutrition is the deficiency of one or more essential vitamin or mineral and is the outcome of an inadequate diet or disease. 163 million or 31% of all under five year old children were estimated to be vitamin A deficient in 2007. Vitamin A deficiency is mostly observed in developing countries and is the main cause of childhood blindness. It is further essential to the immune system which can be fatal in the absence of immunization. Iron deficiency can lead to anemia which affects the production of red blood cells or hemoglobin. This can negatively influence cognitive and growth development of children, pregnancy outcomes, maternal mortality and physical capacity. Iodine deficiency during early childhood or pregnancy impairs mental functions of 18 million children each year (FAO, 2013).

The WHO (2006) defines overweight and obesity as abnormal or excessive fat accumulation that

may impair health. An adult is "overweight" with a Body Mass Index(BMI) equal to or higher than 25, and "obese" with a BMI equal to or higher than 30. The BMI is an index that relates body weight and height by dividing weight in kg by squared body height. Overweight and obesity increase the risk of experiencing non-communicable diseases and health problems, including cardiovascular disease, diabetes, various cancers and osteoarthritis. The probability of incurrance increases with excess body fat.

According the Sustainable Development Report (2018) 32 million people died in 2016 due to cardiovascular disease, which is the the main type of noncommunicable disease, cancer, diabetes and chronic respiratory disease. In 2016 the probability of death from one of these causes was 18% for people between 30 and 70 years.

The manifestation of 2 different outcomes from malnutrition has been coined as "the double burden of malnutrition". This can for instance occur as obesity paired with micronutrient deficiency. Malnutrition represents a considerable social and economic burden. The FAO (2013) estimates the yearly global economic loss due to lost productivity and direct health care costs to be 5% of global gross domestic product equivalent to US\$3.5 trillion or US\$500 per head.

Thompson and Amoroso (2014) as well as the FAO (2013) are skeptical on the long term effect of supplement based actions to fight malnutrition and favor food-based solutions. Supplements are usually utilized in short term treatments for a specific deficiency. Beyond that nutritionists may not fully understand yet how the components of a diet interact so that good health is maintained. The root of prevailing malnutrition is an unbalanced diet of consumers. Food-based approaches tackle malnutrition by promoting dietary diversification as a long-term sustainable strategy. A variety of foods, given sufficient quantity, quality and favorable combinations can provide the required energy, micronutrients and fibre for a healthy nutrition.

Efforts to change food systems require complementary interventions in food system, public health, education and related policies (FAO, 2013). The development of the agricultural sector can reduce malnutrition due to wage increases. However beyond fostering growth agricultural policies need to encourage farmers to further diversify and shift production from staples towards more nutrient dense foods in order to facilitate a more nutritious and diverse food supply. However further mandatory interventions involve the entire food system, ranging from inputs and production, through processing, storage, transport and retailing, to consumption. The availability of a diversity of foods can be improved by modern and traditional supply chains. While considerable efforts are required

to alter a food system it is ultimately the consumer who chooses his diet. The government and other institutions can aid consumers to make better food choices. For instance by implementing food assistance programs, nutrition training or improved food environments. This can aid consumers to obtain the knowledge and information they need to make sound food choices or overcome poverty constraints with assistance programs. Setting incentives to influence consumption pattern towards a more diverse healthy diet can play an important role, but can also result in unintended consequences.

A consumer's demand for food diversity can be complicated to assess and an influential work from Jensen and Miller (2010), who analyze food consumption patterns in China indicates that there are more complex demand structures to be explored. They assume that consumers demand food attributes like diversity depending on their nutritional status. As caloric sufficiency depends on unobservable individual characteristics like absorption ratios this status is unknown and inferring solely from income can lead to flawed conclusions. Understanding the demand for food diversity is however essential information for governments to target program implementation or set incentives in order to successfully transform the food system and reduce malnutrition.

Besides the importance to understand the demand for food diversity evaluation methods of diets are useful tools to compare diets and measure successes. Food diversity indices are derived from the distribution of food items or groups within the consumption basket and have shown to be well correlated with nutrient adequacy (Hatloy et al., 1998) so that they allow inference on nutrition from observed consumption. So far hardly any diversity index was actually designed to measure diet quality. A noteworthy concept is the healthy food diversity index from Drescher, Thiele and Mensink (2007) who use recommended food group shares as index weights for the case of Germany. Further refinements and adjustments to developing countries could result in a valuable indicator to monitor changes in diet quality or infer on food security.

1.3 Summaries of Essays

This thesis addresses 3 very different aspects related to food, nutrition and poverty in developing countries and covers a new method to estimate transaction costs in Kenyan maize markets, a latent class analysis of the demand for food diversity in India and a food diversity index to analyze Indian consumption data.

The first essay presents a new operational concept of transaction costs that firstly allows assessing the magnitude and secondly recognizing the non homogeneity of food products.

Transaction costs have mostly been studied in theory while very few methods emerged to actually estimate them. Many of these methods are directed towards the application on financial markets where however the conditions of exchange differ significantly from those that are present with the exchange of agricultural goods (Kähkönen and Leathers, 1999). Transaction costs theory recognizes that transactions do not occur in a frictionless environment. For small scale farmers in developing countries transaction costs can present a barrier to market participation or drain their profits which can prevent small scale farmers from escaping poverty. Consequently models that are capable to analyze transaction costs can be a viable tool for development.

The proposed a model in this essay decomposes unit values into spatial price factors and a value component which allows the comparison with a feasible value occurring in a situation without transaction costs. We assume that agricultural goods are not homogeneous but differ in quality and hence value. Since the model requests estimation of potential outcomes a stochastic frontier approach is selected for the analysis. A conditional mean model is used to determine the unit value of a good in the first stage with determinants of the value of a good and model the mean difference to the value frontier with a set of transaction costs determinants. This procedure allows to simultaneously estimating the magnitude of transaction costs and identifying the main determinants. Due the presence of a potential selection bias our empirical framework includes Greene's (2010) sample selection model as well as the Battese and Coelli (1995) model.

For the analysis we use the survey data on Kenyan semi subsistence farmers that was collected by the Tegemeo Institute in cooperation with the National Graduate Institute for Policy Studies (GRIPS). The surveys in Kenya were conducted in 2006-2007 following a randomized design and covers 725 households located in 93 sub-locations. The survey is well suited for the underlying

study as it contains details on the specific maize products that are required to determine value as well as details on each transaction that can serve as proxies for different types of transaction costs. The results suggest a magnitude of 12-18% for maize transactions in rural Kenya and identify drive time, market distance, education and counterparts in negotiations as main determinants.

The second essay is concerned with an estimation of the latent demand structure for food diversity in India.

The basic intuitions for a latent demand model for food diversity stem from Jensen and Miller (2011). They state that whether an individual has achieved calorie sufficiency or overcome hunger cannot be adequately determined by a calorie threshold due to individually varying needs. However individuals reveal their nutritional status in their consumption behavior. A consumer facing hunger strives for survival and favors calories over a diverse diet. Once having achieved caloric sufficiency consumers start to diversify as other food attributes such as diversity become more important than additional calories. The problem that caloric needs vary over individuals induces unobservable heterogeneity that lead to a latent demand pattern in the population.

We estimate the latent demand model by means of finite mixture models using a food item count index as dependent variable. The count data distribution and the number of components are selected by model selection criteria and identify a two component poisson model as the best fit. Further the determinants of posterior membership probabilities are analyzed in a beta regression model and characteristics of class members in terms of nutrient intake compared.

For our analysis we use the 68th round of the Consumer Expenditure Survey which was conducted in 2012. Over 4 sub rounds 101626 household were interviewed. The survey contains information on consumption expenditure over the last 30 days and provides a high level of detail on food expenditures. In total it provides quantity and expenditure information of 142 food items of which 127 remain for the analysis. Nutrient conversion tables for all food items were extracted from Nutrisurvey.de to evaluate nutritional implications of class memberships.

Estimates of the demand model reveal two distinct demand behaviors that are consistent with the initial intuitions. The more deprived class has a lower income elasticity of food diversity. Hence caloric sufficiency has not been attained yet and calories are favored as they are essential to survival. The classes differ widely in terms of nutrient intake. Under subsistence concerns more staples are demanded resulting in higher levels of malnutrition than the better off class that fares better with

all micro nutrients while maintaining a comparable calorie intake.

The third essay is concerned with the inference on nutrition from observed consumption and presents a new food diversity index for India. Measures of diversity have become popular tools to infer on nutritional adequacy from observed consumption. However the most common measures do not consider that equal distribution of food consumption does not reflect a healthy diet. Optimally a diversity index should react to favorable or unfavorable redistributions that are in line with dietary recommendations. The proposed index in this essay modifies the existing concept of the healthy diversity index from Drescher, Thiele and Mensink (2007) that considers recommended food group weights in the index calculation. The original index was matched to German dietary recommendations. The weights for the Indian version are derived from the official Indian dietary guidelines. Further the index is adjusted to be used on household level by taking into account the household composition. Lastly the new index is based on food groups rather than food items as these indices showed a better correlation with nutrient adequacy. For the index calculation and performance analysis we use the same data as in the second essay which is the NSS Consumer Expenditure Survey (2012) and conversion tables extracted from Nutrisurvey.de.

The performance of the index is tested by comparing correlation coefficients between various diversity indices and 12 different nutrient adequacy ratios. Indices that are considered for the analysis include the Berry index and the Entropy index of which versions based on food items and food groups were included, and food item counts. The Indian healthy food diversity index shows the highest correlation coefficients among all considered indices in 8 out of 12 cases including mean nutrient adequacy. Further we compare local polynomial regressions on mean nutrient adequacy. For the case of the Indian healthy food diversity index we find a very linear relation and the smallest confidence intervalls among all considered indices.

Chapter 2

Estimation of Transaction Costs for Maize Markets in Rural Kenya

Abstract

While there is rich literature covering theoretical concepts of transaction costs very few empirical strategies have been provided to estimate them. The theoretical framework proposed in this paper is based on a unit value decomposition and defines transaction costs as the difference between a unit value and a frontier value realized in a situation without transaction costs. Estimates of transaction costs are obtained by means of stochastic frontier models with the data from Kenyan maize farmers. We find a magnitude of 12-18% for maize transactions in rural Kenya and identify drive time, market distance, education and counterparts in negotiations as main determinants.

Key Words: transaction costs, stochastic frontier analysis, food prices, Kenya

JEL: D4 Q13 Q11

2.1 Introduction

Throughout history transaction costs have played a major role in economic theory and a variety of theories and definitions have emerged though empirical measuring encounters many high hurdles. Coase (1937) described them as the cost of using the price mechanism while Arrow (1969) viewed them as the costs of running the economic system. Despite the development of many path breaking theories, the development of practical operational concepts has been lagging behind and remains a challenge for empirical researchers. However transactions do not occur in a frictionless environment and understanding these frictions is among the key factors of economic development. For semi-subsistence farmers in developing countries overcoming transaction costs can be a decisive factor in escaping poverty. They can substantially drain farmer's profits (Vakis, Sadoulet and Janvry, 2003) or even represent a barrier to market participation (Key, Sadoulet and Janvry, 2000). Since many of these factors are unobservable it is a challenging task to develop a reliable measure.

In this paper transaction costs are defined as all costs that are related to the exchange of a good and range from information, bargaining and monitoring costs to costs related to physical transport. The analysis focuses on maize sales of Kenyan semi-subsistence farmers for whom transaction costs can play a major role to enter markets and overcome poverty. According to the Kenya Integrated Household Budget Survey 2006 which was conducted in 2005 Kenya had by then a population of 35.5 Million of which 28.36 million lived in rural areas. In total 6.45 million households were engaged in crop farming of which again 93% planted maize. About 75% of the maize production originates in small scale farms many of which are of subsistence or semi- subsistence. The current literature finds that high transaction costs, particularly the transportation costs, could impede farmers from participation of market transaction, which eventually entraps farmers in producing low-yielding food crops in Kenya and leads to subsistent life (Omamo, 1998; Alene et al. 2008; Barrett 2008; Olwande et al. 2015). Despite the undoubted importance of transaction costs in economics few models are so far concerned with the estimation of its magnitude although there is a considerable body of literature on determinants. The unobservable nature of transaction costs certainly poses one of the major challenges in achieving an appropriate measure. This issue is addressed in this paper by assuming there is a frontier value of products in the absence of transaction costs. A method is proposed to decompose unit values into spatial price factors and a quality value component following Deaton (1988) and Yu and Abler (2009). This quality value is heterogeneous across farms

and not fully observed by traders, so that it incurs the transaction costs. We can derive a measure of transaction costs as the difference between the unit value and a frontier value.

A statistical method that is appropriate for the prediction of potential outcomes is the stochastic frontier model. A conditional mean model is used to determine the unit value of a good in the first stage with determinants of the value of a good and model the mean difference to the value frontier with a set of transaction costs determinants. This procedure allows to simultaneously estimating the magnitude of transaction costs and identifying the main determinants. Due the presence of a potential selection bias our empirical framework includes Greene's (2010) sample selection model as well as the Battese and Coelli(1995) model. The estimations show that the proposed concept is well operational. The estimated magnitude of transaction costs is 12.1-18.2 % of the feasible value of maize which is in line with the previous literature. Transaction costs determinants that show significant are the counterpart in negotiations, the drive time to the next market, the distance to the capital Nairobi, the existence of a market in the sub-location and years of education of the household head.

The paper is structured as follows. First the literature on the here used categorization of transaction costs is discussed and recent findings on transaction costs of smallholders summarized. The theory section explains the decomposition of unit values into spatial price factors, value and transaction costs. The data section briefly presents the RePEAT data and discusses the relevance of proxies for both value and transaction costs categories. The econometric model section is concerned with the adequate estimation of the proposed transaction costs model under the consideration of self selection. The results section provides a detailed discussion of the empirical findings and the conclusions present a summary of the results and evaluate implications for policy and further research.

2.2 Literature Review

It is often argued that the concept of transaction costs started with Coase (1937) "The Nature of the Firm", where he referred to the costs to using of the price mechanism. The important contribution of transaction costs economics is the recognition that transactions do not happen in a frictionless economic environment. The development of Transaction Costs Economics was further accelerated with the rise of the new institutional economic theory with Oliver Williams as a main contributor.

Meanwhile many different strands have developed in the transaction costs related literature. While there is clearly no lag of powerful and useful theories the development of operational concepts could not keep pace and remains one of the main challenges in the transaction costs literature. One of the reasons is certainly the unobservable nature of institutions that is difficult to capture with empirical models.

In the course of agricultural exchange in developing countries farmers face a variety of institutional factors affecting transaction costs. Some of them are missing markets, information asymmetry, risk and uncertainty, non-separability of consumption and production, incomplete property rights and institutional failures (Cuevas, 2014).

An empirical analysis of transaction costs requires a basic definition to establish a framework and choose adequate proxy variables. It is straightforward to choose a categorization here that has already been established in the analysis of agricultural exchange. An operational concept that has appeared frequently in the literature proposes the differentiation of information, bargaining and monitoring costs. Information costs usually occur prior to the exchange and include the search for price information and potential buyers as well as the facilitation of the transactions. Bargaining, also labeled negotiation costs usually occur during an exchange and refer to all costs that are connected to the process of negotiation and reaching an agreement as well as costs related to payments. Bargaining costs can be significantly influenced by personal attributes such as education, gender and skills. Monitoring or enforcement costs are costs that occur to ensure that the agreement is fulfilled. Furubotn and Richter (2005) argue that transaction costs generally include the costs of resources utilized for the creation, maintenance, use, change, and so on of institutions and organizations. Applied to the transfer of existing property rights between individuals, transaction costs include the costs of information, negotiation, and enforcement. Hobbs (1995) is among the first empirical studies to show that this categorization is operational at the example of cattle marketing. While theorists generally describe transport costs as part of the production process empirical applications often highlight the importance of recognizing transport costs as a transaction costs category in the context of analyzing agricultural markets in developing countries (Omamo, 1998; Alene et al. 2008). Kähkönen and Leathers (1999) argue that a definition that assigns transport to the production process may be flawed if the actual costs are eventually paid by the buyer. This is a situation that can be observed in case of farmgate sales that make up for a substantial amount of transfers from semi-subsistent farmers in rural Kenya. Undoubtedly some transport costs occur in the course

of the production process such as costs related to the acquisition of seeds, fertilizer or other capital goods. Kähkönen and Leathers (1999) are concerned with transaction costs in agricultural markets in Zambia and Tanzania. They provide a detailed discussion of the challenges associated with the analysis of transaction costs in agricultural exchange. They list various examples on the substitutability of information related costs and physical costs. As a consequence they conclude that a definition of transaction costs that excludes physical costs may result in misleading conclusions on the efficiency of a transaction. Another argument for the inclusion of transfer costs is that costs related to physical handling of a transaction can also be linked to institutional impediments.

Li et al. (2012) estimated the transportation costs for Chinese food traders and find that these account for 40% of trade barriers in China. Key, Sadoulet and Janvry (2000) as well as Vakis, Sadoulet and Janvry (2003) analyze transaction costs in agricultural markets while differentiating between fixed and proportional transaction costs. While fixed transaction costs are categorized as above and are independent of the quantity transferred, proportional transaction costs vary with quantity. They may originate from different per unit transportation costs or quantity related price premiums that are a result of the negotiation process.

Vakis, Sadoulet and Janvry (2003) are one of the very few studies that propose a method to estimate the magnitude of transaction costs at the example of Peruvian potato farmers. They are concerned with transaction costs as a barrier to market participation and estimate transaction costs by means of a conditional logit market choice model. The results indicate transportation costs of 10-15 percent of the price received by the farmers in integrated and 30 % in isolated areas. Special attention is here paid to market price information that has a very significant impact on transaction costs and the choice of the market channel. It can reduce fixed transaction costs by the equivalent of doubling the price received by the farmer. One of the first estimates of the magnitude of transaction costs for agricultural households comes from Renkow, Hallstrom and Karanja (2003) who estimate the magnitude by means of MLE using a sample of Kenyan maize farmers. Their results indicate fixed transaction costs of 15 % on average in their sample. The literature also shows that transaction costs vary significantly across farmers or traders.

Besides draining farmers' profits, transactions costs are the embodiment of barriers to market participation by resource-poor smallholders and as factors responsible for significant market failures in developing countries. For instance Escoba and Cavero (2004) show at the example of poor potato farmers in Peru that high transaction costs can represent an exclusion mechanism to agro industrial

markets. Osebeyo and Aye (2014) confirm this pattern and find that transaction costs have a significant influence on market participation. Key, Sadoulet, and Janvry (2000) provide a household supply response model for market participation with transaction cost as decisive component. They show that the decision to participate in markets of Mexican maize farmers depends significantly on fixed as well as proportional transaction costs.

2.3 Theory

In order to estimate transaction costs it is necessary to first consider some determinants of the value of a good. A food group item such as maize or bananas is practically not a homogenous good and consists of many different items. The properties and hence the price of a food group item may vary depending on the variety and quality of the items. In Sub Saharan Africa many farmers use local varieties. Due to the individual selection of farmers over generations, these local seeds have a very broad gene pool in contrast to improved hybrid seeds. As a consequence quality and taste can differ significantly across regions (Fafchamps, 2004). Anticipating this heterogeneity, traders tend to inspect goods at each transaction to overcome the associated asymmetric information. Hybrid seeds tend to lead to a lower variety in quality since they are less prone to wrong treatment and climate. This might be preferred for industrial processes if a constant input quality is required for a certain output quality (Fafchamps, 2004). Such circumstances may be anticipated by traders as well in order to meet industrial demands. Other differences in quality can be attributed to soil properties, fertilizer usage, differences in climate or damages. Hence a variety of factors must be taken into account to calculate a reference price and the mere consideration of a market price would ignore some decisive factors. Since prices also differ across locations and traders take individual properties into consideration before setting a price, the magnitude of transaction costs cannot be estimated without the consideration of individual good properties and spatial variations of prices. An appropriate model for the considerations of quality and spatial variations has been introduced by Deaton (1988) and Yu and Abler (2009). For a particular food group i (e.g. maize), we assume there are M different items with different quality, and define the exogenous price vector p_{ij} of food group i in region j as

$$p_{ij} = \lambda_{ij} p_i \tag{2.1}$$

Where λ_{ij} is a scaler and reflects the regional price factor for food group i in region j while p_i is a

price vector capturing the price effect that arises due to differences in quality, independent of the region, such as taste, variety, moisture level, organic or not, etc. For instance, if a kg maize has 4000 items, which have different quality. The true value of this kg maize should be determined by each item's value.

In practice, a farmer could also produce many different items for food group i with different qualities, and his output quantity vector is q_{ij} . Thus the total revenue from the food group i is E_{ij} and $E_{ij} = p'_{ij}q_{ij}$, and the total quantity sold in the market is $Q_{ij} = \Theta'_i q_{ij}$, where Θ_i denotes a vector of ones. The unit value for the food group i can now be derived as follows:

$$V_{ij} = \frac{E_{ij}}{Q_{ij}} = \frac{p'_{ij}q_{ij}}{\Theta'_i q_{ij}} = \lambda_{ij} \left(\frac{p'_{ij}q_{ij}}{\Theta'_i q_{ij}} \right) = \lambda_{ij} v_{ij}(p_i, q_{ij}) \quad (2.2)$$

The unit value is obtained by dividing the total revenue E_{ij} by the total quantity Q_{ij} . Price differences across regions can be removed by factoring them out in λ_{ij} . Then v_{ij} denotes the average price of different food items for food group i in region j while λ_{ij} represents the price differences across regions. v_{ij} can be regarded as the quality value of food group i , which is determined by the composition of different items q_{ij} with different qualities. Due to transaction costs, the quality information cannot be fully observed, measured, monitored or realized by the traders or even the farmers. When taking logs the unit values can be expressed in an additive relation of spatial price factors and the value of quality:

$$\ln V_{ij} = \ln \lambda_{ij} + \ln v_{ij}(p_i, q_{ij}) \quad (2.3)$$

Equation (2.3) shows that the unit value or the receive price of a farm is determined by a regional effect λ_{ij} and a quality factor $\ln v_{ij}(p_i, q_{ij})$ which is asymmetry between farmers and buyers due to transaction costs. In order to find a measure for transaction costs by means of a hedonic price model a situation needs to be modeled in which zero transaction costs occur so that the deviation from the observed unit value can be derived. Hence define V_{ij}^* as the maximum unit value paid for good i that may be obtained by farmers (or traded in the market between farmers and traders) in the absence of transaction costs. This implies that observed unit value can never exceed the feasible one so that $V_{ij}^* \geq V_{ij}$. As the regional factor λ_{ij} is often fixed and known to both trade partners, we can assume

$$\ln V_{ij}^* = \ln \lambda_{ij} + \ln v_{ij}^*(p_i, q_{ij}) - t_{ij} \quad (2.4)$$

Where $\ln v_{ij}^*(p_i, q_{ij})$ can be defined as the frontier quality. Then the transaction cost can be defined as the distance between the frontier and the realized values. Consequently its share as a fraction of the frontier can be calculated as follows:

$$t_{ij} = \ln V_{ij}^* - \ln V_{ij} = \ln \left(\frac{v_{ij}^*(p_i, q_{ij})}{v_{ij}(p_i, q_{ij})} \right) \quad (2.5)$$

Combining Equation (2.3), (2.4) and (2.5), yields,

$$\ln V_{ij} = \ln \lambda_{ij} + \ln v_{ij}^*(p_i, q_{ij}) - t_{ij} \quad (2.6)$$

In this representation of unit values we can explain the value of a good in terms of a value frontier and its difference to the realized unit value. This model now consists of three components: firstly a spatial price factor for each product that varies over regions and secondly the frontier quality and thirdly a component t_{ij} which represents the transaction costs. As the frontier quality is unknown, the given structure in Equation (2.6) can be estimated by a stochastic frontier approach.

Note that the difference between the frontier price and the actual price is defined as the transaction costs in our study, but we do not explicitly indicate who pays the transaction costs. Let's take an example for transportation cost here. If the realized prices are the farmgate prices as most cases occur in the maize market of Kenya in this study, the transportation costs are actually paid by the buyers. If farmers deliver maize and pay the transportation costs, the transportation costs will be added to the farmgate price, but the "actual" farmgate price does not change. Empirically, the trade places will be controlled for the transportation costs.

2.4 Econometric Model

2.4.1 Stochastic Frontier Model

With the aim to estimate transaction costs, we are interested in the deviation of an observed price (or unit value) from a frontier price that is achievable under optimal conditions rather than a deviation from the average which would be given by a common regression model. We hence require

a model that allows us to estimate the distance between an observed and a maximum feasible price as illustrated in figure 2.1. Although it has never been used in this context, a stochastic frontier model fulfills these criteria. The stochastic frontier model is used to estimate equation (2.6) which has been derived in the theory section while keeping in mind that the value frontier consists of a regional base price and the feasible price effect that is in nature hedonic. This equation can be estimated by a standard stochastic frontier model as suggested by Aigner, Lovell and Schmidt (1977).

$$\ln V_i = x_i \beta + (v_i - u_i) \quad (2.7)$$

V_i refers to the to the unit value of a good that was observed in a transaction for farmer i . x_i is a $k \times 1$ vector containing the determinants of quality and spatial variations and β is the vector of parameters to be estimated. $v_i - u_i$ is the decomposed random error term. The symmetric disturbance $v_i \sim N(0, \sigma_v^2)$ captures unobserved heterogeneities and measurement errors and is assumed to be independently and identically distributed. The term u_i presents a measure of transaction costs which is in the theoretical model referred to as $\ln\left(\frac{v_{ij}^*(p_i, q_{ij})}{v_{ij}(p_i, q_{ij})}\right) > 0$. It is non-negative, follows a one-sided distribution and is assumed to be distributed independently of v_i . In the context of the underlying analysis it is desirable to model transaction cost with a set of exogenous variables to identify determinants. This feature is available from conditional mean models. The first conditional mean models to parameterize the mean of the transaction costs in order to study exogenous effects on transaction costs which originates from of Kumbhakar, Ghosh, and McGuckin (1991) and Huang and Liu (1994), The model was extended to the case of panel data by Battese and Coelli (1993, 1995).

In the empirical model the feasible value V_i^* can be estimated by means of $x_i \beta$ The estimation of the frontier model yields estimates of u_i that in turn allow for a calculation of transaction costs. Note that u_i denotes the difference between the logs of expected feasible and observed values. Here u_i is modeled as a function of z_i , and assumes u_i following a truncated normal distribution:

$$u_i \sim N^+(\mu_i, \sigma_u^2), \quad (2.8)$$

$$\mu_i = z_i \delta + \omega_i$$

N^+ refers to a truncated normal distribution. z_i represents a vector of transaction cost determinants and δ the respective parameter to be estimated. In this way u_i can be modeled as a function of transaction costs determinants z_i with parameters δ, ω_i is a random error following a truncated

normal distribution. Relative transaction costs (t) as the share of the feasible values can then be calculated with

$$t = 1 - \exp(-u_i) \quad (2.9)$$

The total transaction costs (T) can be calculated as a share of the predicted frontier values:

$$T = \exp(x_i\beta) * (1 - \exp(-u_i)) = \exp(x_i\beta) * t \quad (2.10)$$

2.4.2 Selection Bias

In our sample only a fraction of the maize producers decide to participate in the market. Despite transaction costs being constituted through a market transaction, each producer has to face them prior to the decision whether to participate in a market. Transaction costs may be a decisive determinant of this decision and many households fail to participate in a market due to transaction costs. Thus market participants cannot be considered representative for all producers and are a biased represent. Since unit values can only be observed when farmers participate, the data used in the transaction cost model is subject to truncation. According to Heckman (1979) a selection bias can occur if the dependent variable can only be observed when passing a certain threshold. The threshold that needs to be passed in order to observe a price and hence conclude on transaction costs is the participation in the market. If there are unobserved costs related to market participation ignoring this problem will yield biased estimates.

Accounting for selection bias adequately in a conditional mean model is however challenging as all existing solutions have different weaknesses. The empirical approach in this paper largely leans on Wollni and Brümmer (2012) who analyzed the productive efficiency of specialty and conventional coffee farmers in Costa Rica under self selection and provide a detailed discussion on the dilemma situation. Their self selection framework includes the Greene (2010) selection model as well as the Battese and Coelli (1995) model. Following Heckman we apply a two step procedure. In the first step the probability of market participation is estimated by means of a probit model. The model predictions are then used to calculate the inverse Mills ratio and include it in the stochastic frontier model as an additional regressor.

The two-step Heckman approach in combination with a stochastic frontier model has however been criticized as biased by Greene (2009) since the Heckman model is not an adequate solution for

non-linear models. It has grown increasingly popular to solve selection biases in stochastic frontier analysis by means of propensity score matching in order to achieve unbiased estimates of differences in technology. However since we can only observe transaction costs after a farmer’s decision to participate in the market, matching procedures are not an option. Greene (2009) proposed a selection corrected stochastic frontier based on a maximum simulated likelihood estimator that can consistently account for a selection bias, assuming that the unobserved factors of the selection equation are correlated with the error term of the stochastic frontier model. Greene’s solution represents a special case of the Heckman 2-step estimator. The model does, however, not allow for conditional mean modeling. This is a potential shortcoming for our analysis since we are especially interested in the determinants of transaction costs. Hence we regress the transaction costs on related variables and predict transaction costs of the Greene’s selection model. Such a procedure will yield biased results since the first step is misspecified. Schmidt and Wang (2002) as well as Schmidt (2011) discuss the shortcomings of a 2-step procedure versus possible 1-step procedures. Schmidt and Wang (2002) perform a Monte Carlo simulation and find a substantial bias in two-step procedures while examining several model parameters. The effects of z on u are biased downwards even if x and z are uncorrelated. An 1-step solution in which the transaction costs distribution can be directly affected by z is hence superior but cannot correctly account for self selection.

In contrast to the Battese and Coelli (1995) model that follows a truncated normal distribution Greenes self-selection model assumes a half normal distribution of transaction costs so that $u_i \sim N^+(0, \sigma_u^2)$. With μ the Battese and Coelli model has one additional parameter to be estimated which makes it more flexible. As transaction costs in the Greene’s method follow a half-normal distribution this implies that increasing transaction costs become increasingly less likely (Kumbhakar and Lovell, 2000). Note that, compared with other distribution assumption (e.g. truncated normal distribution), the half normal model might generate lower transaction costs since it forces the model value of transaction costs to be 0 and is therefore more restrictive.

2.5 Data

We use the survey data collected by the Tegemeo Institute in cooperation with the National Graduate Institute for Policy Studies (GRIPS). The data was collected as part of the Research on Poverty and Environment and Agricultural Technology Project (RePEAT). The objective of the survey was

the identification of agricultural technologies and farming systems that contribute to increased agricultural productivity, the sustainable use of resources and reduced poverty in Kenya (Yamano et. al 2004). The surveys in Kenya were conducted in 2006-2007 following a randomized design. It covered 725 households located in 93 sub-locations. The survey contains detailed information on the farm activities and household characteristics as well as information on the villages. The choice in favor of this rather dated survey was made since it provides the information required to test our model adequately. This includes sufficient details on each transaction to find enough proxies that allow capturing several categories of transaction costs as well as sufficient specifics on the cultivation of maize for the hedonic value model. Since the data contains details on the plots where the maize that was sold was planted it allows the value of a transacted good to vary in value even if it has been produced by the same household. This level of detail makes this data set particularly interesting for this study despite the uncertainties that arise from the use of different weight measures.

After removing all households from the data that did not produce maize and dropping observations with missing values, the data set used in the analysis contains 510 households of which 328 did not participate in the market. For the remaining 182 households there are 258 observed transaction for the stochastic frontier model. Descriptive statistics for transactions and sold maize are listed in Table 2.1 of Appendix A and Table 2.2 provides an overview on household characteristics.

The stochastic frontier model contains Kenyan Shillings per kilogram of maize as dependent variable. The weight measures for maize trade price, represented by Shillings per kilogram had to be calculated from a range of different weight measures. Along with the survey data on different weight measures was collected to construct conversion tables. These were used here as well to convert all measures to kilogram. Since only very few attributes of the product are directly observable, like for instance the variety, factors are included that can influence the hedonic value of the maize due to its influence on the production process. In order to find a proxy for quality we used the amount of chemical fertilizer in kg per acre to account for the influence of chemical fertilizer on the price. We also tried further differentiations of chemical fertilizers such as the separate consideration of NPK fertilizer, which is often mentioned to be of special importance for maize as for instance stated by Matsumoto and Yamano (2009). Niaz et al.(2015) find that the application of nitrogen can enhance the yield as well as the quality of maize grains. Although they also show that the application pattern of fertilizer matters it is here only possible to account of the quantity that was applied per hectare. In the estimations the type of the chemical fertilizer did not seem to matter.

Additionally we added kg of organic fertilizer per acre, which can be manure, ash or compost. The total amount of land in acres available for a household has been added to test the hypothesis that quality is more difficult to control when managing a larger land mass. We further added a dummy that indicated if the household owns the parcel from where the sold goods originated as a household might be willing to invest more into a parcel that it owns than into a rented parcel. Yamano et al. (2005) find that farmers for instance apply more fertilizer to plots for which they possess a land title. They do generally suspect that farmers are less willing to invest into plots that could be taken away from them in the near future. The willingness to invest may positively influence the quality of the products. On the other hands more than half of the parcels were inherited. It is likely that they have been cultivated over a long time and may suffer from the depletion. Hence the sign of this dummy could go either way. A damage dummy signals if the harvests experienced any damage in this sub-location. The information on damage includes damage caused by rain, floods, insects, animals and disease in the area. While this information is only available for the sub-location there is some probability that the harvest involved in the transaction was affected as well. In our sample such damages occurred in the surrounding of 17.5 % of the households. We assume that there is a lower price for damaged goods as the damage could lead to low quality. The educational level of the household head measured in years of education is supposed to capture the ability to process agricultural information and hence introduce innovations. For the same reason the mean education in years of all household members was added as they are likely to be involved in the production process as well. The size of the household is related to the number of people from the household that help out on the field. Since the sample covers semi subsistence and subsistence farmers household members basically grow their own food and hence are expected to do this with a different motivation than hired labor. Concerning the age of the household head we do not have specific expectation on the sign of its impact on prices. It can reflect experience but there can also be decreasing productivity with increasing age. Further a dummy was included indicating if the household head is female as the gender of the household head might influence production technologies. Information on the variety used was included with a dummy indicating whether improved hybrid seeds have been used (=1) or local varieties. In 74 % of the transactions maize from hybrid seeds were transferred while the remaining 26 % were traditional varieties. A delivery of the goods might increase its value for a buyer a dummy indicates whether the maize was sold at the farmgate. 77 % of all transactions were farmgate sales. The remaining 23 % of the transactions took place

at local markets, trading centers, and schools. The spatial variation is covered by 4 province dummies. These were the Western Province (12%), Nyanza (28%) and Rift Valley (20%). The Central Province which accounts for 40% of the observations was attributed to the base category.

As explanatory variables 11 variables were selected as proxies for search and information costs, bargaining costs and transport related costs. In some cases it is possible to reason that one variable may serve as a proxy for more than one type of costs. Table 2.3 gives an overview which proxy variables can be assigned to which transaction costs category. There are however no suitable proxies in the sample that can be related to monitoring and enforcement. All observed transactions in our sample are spot market transactions that usually involve an instant exchange of goods and cash. However the absence of other channels may indicate that enforcement costs are too high to use other markets than spot markets.

Firstly we add a dummy that indicates whether the household possesses a mobile phone given that a mobile phone network has been available. This is the case for 45 % of the households in the sample. Phones are essential to gain market information, search for buyers and establish business connections. Next a dummy is added that takes the value 1 if the household used a phone to obtain market information. The substantial influence of price information on transaction costs has been discussed by Vakis, Sadoulet and Janvry (2003).

Personal attributes can influence the course of negotiations and its outcome. So we added the information whether the household head is female and the education of the household head in years. It is very likely that the household head is involved in price negotiations. The gender of the household head was included as a dummy that takes the value 1 if the gender of the household head is female. Depenbusch (2017) analyzes data from Kenya and finds that the gender of vegetable traders can influence price negotiations. The years of education can be related to several cost categories. On the one hand it will be easier for a more educated person to process market information on the other the ability to acquire and process information can influence the outcome of negotiations as well.

In general the place of the exchange as well as the trading partner will influence the distribution of the transaction costs. Associated variables may affect bargaining as well as transportation costs. A dummy was added that indicates whether the exchange took place at the farmgate. All off farm locations are attributed to the base. The counterpart in negotiations and their experience will influence the outcome of negotiations. The major counterparts involved were traders, consumers

and institutions such as schools. Since traders are the only group that negotiates food prices by profession a dummy is added that takes the value 1 if the counterpart was a trader. This allows inference on how a professional food trader fares compared to non professional one. In 75 % of all transactions the farmer negotiated with a trader.

The quantity involved in a transaction can influence the bargaining position of buyers and sellers. Typically the price per unit decreases in quantity. Quantity is hence related to bargaining as well as transport. However we may assume as well that traders try to exploit their transport capacities in which case the single sales quantity does not matter anymore while the trader relates the cost of transportation to its capacity and not the single sales quantity. On the other hand his bargaining power will increase with growing distance to a market since for the farmer alternatives to sell his goods decrease especially when missing adequate transport possibilities to a market. The sample provides drive time and distance to the next market. As these variables are highly correlated only drive time was considered in the model as it might better reflect the effort to reach a market which accounts for other factors like road condition beyond the mere distance. The road condition and the accessibility of a location can vary with the harvest season due to changing weather conditions. Consequently a dummy indicates the harvest season. Beyond that it can influence the risk of damage during transport for instance due to road flooding. The importance of the road infrastructure for agricultural production and markets has for instance been highlighted by Dorosh, Wang and Schmidt (2010). Further the distance to Nairobi is added, which is the capital of Kenya. Transport related variables do of course not only affect transport costs. They can also increase costs related to the search for buyers or market information. Kähkönen and Leathers (1999) additionally provide some examples on how factors that can affect physical handling of a good can be linked to institutional impediments.

The selection model contains characteristics of households that planted maize to estimate the probability of a household to participate in the market.

2.6 Empirical Results

Table 2.4 contains the results of the first stage probit that was estimated for the Greene selection model and the Battese and Coelli model. The model shows that the probability to participate in the market decreases in the household size. Since semi-subsistence farmers are subject to the

analysis we can assume that larger households are more likely to rely on subsistence and consume their produced crops. On the other hand it increases in the size of the land of a household. More land allows the household to produce food beyond subsistence needs which enables it to generate additional income. The age of the household head is negatively related to market participation. Older household heads are often less likely to adapt innovations. The existence of a market in the sub location has a significant positive effect. The existence of such a channel can be expected to reduce the costs related to the search for buyers significantly. The distance to the capital Nairobi has a negative effect. The distance is especially important for traders who want to sell to export firms or the industry since a larger distance increases their costs. The Greene selection model as well as the Battese and Coelli model show clear evidence of self selection. The selection parameter ρ that indicates that unobservable factors from the selection equation are correlated with the error term of the stochastic frontier model is highly significant. The same is true for the inverse mills ratio in the Battese and Coelli model so that the selection parameters in both models confirm that controlling for selection is required to avoid a selection bias.

From the estimation results for the Battese and Coelli model that are reported in Table 2.5, we can see that two of three province dummies are highly significant. We can thus infer that the location has a significant influence on the price, consistent with our theoretical framework. As expected damages in the area have a significant negative effect on the unit value which probably originates in loss of quality due to the damage. The amount of land has indeed a significant negative impact on the price, which supports the hypothesis that quality control becomes more difficult with increasing farm land size. For instance Yamano et al. (2005) find that semi-subsistence farmers in Kenya apply fertilizers more frequently to smaller plots. Owning a legal land title has a negative effect that probably originates from soil depletion due to long time cultivation. Surprisingly the years of education of the household head show a negative effect. One possible explanation is that better educated farmers face more alternatives to generate income and are therefore neglecting the cultivation.

The second stage models the mean of the transaction costs term μ . The presence of a market in the sub-location significantly increases transaction costs. A natural explanation for this effect is that this channel increases the competition among farmers. Price negotiations with a trader lead to significantly higher transaction costs. Hence there is significant positive difference in transaction costs when negotiating with a professional food trader as compared to a non professional one. The

drive time to the next market as well as the distance Nairobi increase costs. Both enhance the bargaining position of buyers. The years of education of the household head lead to a reduction transaction costs. Table 2.6 contains the results of the Greene's selection model. Only the dummy indicating that that household owns the land title has a significant effect as well as the province dummies. In the second stage of the model which was estimated by OLS only the trader dummy is significant. However according to the F-statistic the overall model relation is insignificant. It implies that Greene's selection model does not perform well in this study. Table 2.7 report the estimated relative and absolute transaction costs, respectively. Figure 2.2 shows the corresponding Kernel density estimates.

The estimates from the Battese and Coelli model are with 18.2 % substantially larger than the estimates from the Greene model with 12.1 %. The difference can occur due to the choice of the distribution as the Battese and Coelli model uses a truncated normal and the Greene model a half normal distribution. Our results on the magnitude of transaction costs can only be vaguely compared with other findings in the literature since they are calculated using a different base in the stochastic frontier approach as percentages relate to feasible values. However we can infer from a comparison on whether the results are reasonable. The estimate of 15% from Renkow, Hallstrom and Karanja (2003) is right in between ours. Since they concentrated on estimating fixed transaction that impose a market entry barrier it can be inferred that total transaction costs might be underestimated by this measure. Yamano and Arai (2010) calculated within part the same data as we used a price spread of 15% in the Kenyan maize market which is pretty close as well. Although pursuing a different aim than we did they also identified drive time as one of the major determinants of price spreads. Taking into consideration similar or in part comparable approaches to estimate transaction costs our estimates appear to be of a reasonable magnitude and comparable with the findings of other researchers.

2.7 Conclusions

This paper presents a theoretical framework and an empirical method to derive transaction costs within a stochastic frontier approach. This method can help to infer on transaction costs when dealing with heterogeneous goods as it is typically the case of agricultural products. It could be demonstrated that despite the unobservable nature of transaction costs the concept is well

operational at the example of Kenyan semi-subsistent farmers. The presented price decomposition techniques and estimation approaches in this paper can aid policy makers to assess transaction costs and set priorities among policies that aim to support semi-subsistence farmers. Such policies have substantial meaning for poverty reduction programs in developing countries.

The presented method can be used to estimate and simultaneously explain transaction costs by means of econometric models. Although the current state of development does not offer solutions to all econometric problems that are related to the presence of a selection bias while modeling a transaction costs term. In the proposed stochastic frontier approach we find transaction costs of 12.1-18.2 % in the Kenyan maize market. A comparison with other recent findings shows that the estimated magnitudes are of a reasonable level. Further information on the main determinants of transaction costs could be obtained. One hurdle to overcome appears to be the drive time to the next market and the distance to Nairobi. The corresponding conditions can change with improving road infrastructure which would allow farmers to receive better prices or gain access to markets in case the magnitude of the transaction costs still has a critical level that prevents market participation. Evidently negotiations with professional traders raise transaction costs significantly. The study also shows the importance of education. This has important implications for small hold farmers since better educated farmers do not only have more alternatives to generate income but can improve existing income sources. Although it was not possible to find adequate proxies for monitoring costs the persistent reliance of farmers on spot markets may be seen as an indicator in its own right. If monitoring and enforcement costs are too high to use other sales channels then the reasons may originate from the institutional setting.

It remains a challenge for empirical researchers to find adequate data for a transaction costs analysis. Institutions are difficult to assess and require plenty of proxy variables for a holistic assessment that covers the whole range of transaction costs. Currently very few data sources are available that allow an adequate analysis of the transaction costs that small hold farmers face in developing countries. Further research in this area should be concerned with elicitation of data and the identification of proper proxy variables for transaction cost analysis.

2.8 Appendix A

Table 2.1: Characteristics of transactions and products

Variable	Mean	Standard Deviation
Maize Price per kg	12.548	2.818
Chemical fertilizer per acre in kg	20.456	30.871
Organic fertilizer per acre in kg	72.206	317.819
Household owns parcel (dummy)	0.779	0.416
Improved hybrid seeds	0.74	0.439
Other sales party was a trader(dummy)	0.748	0.435
Quantity sold in kg	560.539	730.809
Sale took place at the farmgate	0.771	0.421

Observations: 258

Table 2.2: Household characteristics

Variable	Mean	Standard Deviation
Market participation	0.357	0.480
Household size	8.218	3.490
Damage (dummy)	0.175	0.380
Acres of land (owned and rented)	5.194	7.968
Education of household head in years	6.723	4.613
Age of household head	58.757	13.391
Hh posses mobile phone given a network(dummy)	0.449	0.498
Household head is female (dummy)	0.245	0.431
Drive time to next market in minutes	30.765	22.071
Distance to Nairobi in km	206.657	113.165
Hh obtained market information via phone (dummy)	0.045	0.201
Province 1(Nyanza)	0.284	0.452
Province 2(Western Province)	0.118	0.323
Province 3 (Rift Valley)	0.198	0.399
Province 4 (Central Province)	0.400	0.491
Observations: 510		

Table 2.3: Proxies for different types of transaction costs

Proxy/Type	Transport	Information	Bargaining
hh posses mobile phone given a network(dummy)		X	
hh obtained market information via phone (dummy)		X	
drive time to next market in minutes	X	X	
distance to Nairobi in km	X	X	
other sales party was a trader(dummy)			X
sale took place at the farmgate	X		X
quantity sold in kg	X		X
education of hh head (Years)		X	X
household head is female (dummy)			X
Season	X		
There is market in the area	X	X	

Table 2.4: Probit model on market participation

Variable	Coefficient	Standard Error	$P > z $
<i>P(Market Participation)</i>			
Ln Household size	-0.3881***	0.123	0.002
ln acres of land (owned and rented)	0.4835***	0.070	0.000
ln education of household head in years	0.0066	0.079	0.933
household head is female	-0.2087	0.147	0.155
ln age of household head	-0.7306***	0.266	0.006
there is a market in the area(dummy)	0.2504*	0.139	0.071
ln distance to Nairobi	-0.2727*	0.142	0.055
ln drivetime to next market	0.0749	0.076	0.324
Province 1 (Nyanza)	0.2339	0.280	0.404
Province 2 (Western Province)	-0.122	0.310	0.694
Province 3 (Rift Valley)	0.2473	0.196	0.208
Constant	3.9416***	1.380	0.004
Number of obs = 586			
Prob > χ^2 = .00000			
Log likelihood = -359.921			
Pseudo R^2 = .1046598			

Table 2.5: Stochastic frontier model for transaction costs (BC 1995)

Variable	Coefficient	Std. Error	$P > z $
<i>Ln Price per kg (Frontier Model)</i>			
Any Damage (dummy)	-0.0596*	0.0353	0.0908
ln Chemical fertilizer per acre in kg	0.0010	0.0097	0.9162
ln Organic fertilizer per acre in kg	-0.0099	0.0061	0.1034
household owns parcel (dummy)	-0.0590*	0.0339	0.0818
ln acres of land (owned and rented)	-0.1244***	0.0409	0.0024
ln education of household head in years	-0.0924**	0.0456	0.0428
ln mean education of hh members in years	0.0304	0.0410	0.4581
Ln Household size	0.0347	0.0424	0.4124
ln age of household head	0.1337	0.0907	0.1404
household head is female	0.3353	0.2993	0.2626
sold at farmgate(dummy)	-0.0123	0.0547	0.8216
improved hybrid seeds	-0.0073	0.0353	0.8371
Province 1(Nyanza)	0.2683***	0.0560	0.0000
Province 2(Western Province)	0.3197***	0.0747	0.0000
Province 3 (Rift Valley)	-0.0387	0.0419	0.3557
Inverse Mills Ratio	-0.4801***	0.1456	0.0010
constant	2.7158***	0.3145	0.0000
<i>μ(Transaction Costs Function)</i>			
hh posses mobile phone given a network(dummy)	-0.0419	0.0533	0.4313
hh obtained market information via phone (dummy)	0.0839	0.0918	0.3605
there is a market in the area(dummy)	0.1416**	0.0696	0.0418
drive time to next market in minutes	0.0028**	0.0012	0.0233
distance to Nairobi in km	0.0010**	0.0004	0.0126
other sales party was a trader(dummy)	0.2432***	0.0865	0.0049
sale took place at the farmgate	0.0148	0.0876	0.8659
quantity sold in kg	0.0000	0.0004	0.9392
education of hh head (Years)	-0.0290**	0.0131	0.0270
household head is female (dummy)	0.4010	0.3017	0.1838
Season(dummy)	-0.0427	0.0822	0.6033
Constant	-0.3388	0.2567	0.1870
σ^2	0.0394***	0.2125	0.0489
γ	0.4185**	0.0091	0.0000
Log likelihood	85.0905		
N	258		

Table 2.6: Stochastic frontier model for transaction costs (Green, 2010)

Variable	Coefficient	Std. Error	$P > z $
<i>Ln Price per kg (Frontier Model)</i>			
Any Damage (dummy)	-0.0543	0.0391	0.1645
ln Chemical fertilizer per acre in kg	-0.0074	0.0115	0.5207
ln Organic fertilizer per acre in kg	-0.0046	0.0061	0.4461
household owns parcel (dummy)	-0.0739**	0.0373	0.0476
ln acres of land (owned and rented)	-0.0495	0.0302	0.1011
ln education of household head in years	-0.0238	0.0272	0.3818
ln mean education of hh members in years	0.0365	0.0441	0.4082
Ln Household size	0.0162	0.0409	0.6923
ln age of household head	0.0190	0.0902	0.8276
household head is female	0.0177	0.0408	0.6719
sold at farmgate(dummy)	-0.0018	0.0357	0.9589
improved hybrid seeds	-0.0168	0.0349	0.6295
Province 1(Nyanza)	0.1439***	0.0438	0.0010
Province 2 (Western Province)	0.0592	0.0561	0.2915
Province 3 (Rift Valley)	-0.1382***	0.0400	0.0005
Constant	2.8083***	0.3769	0.0000
$\rho(w, v)$	-0.7919***	0.1949	0.0000
σ_u	0.1515	0.1370	0.2685
σ_v	0.2181***	0.0669	0.0011
Log likelihood	127.5835		
<i>OLS</i>			
hh posses mobile phone given a network(dummy)	-0.0072	0.0060	0.2343
hh obtained market information via phone (dummy)	0.0037	0.0106	0.7283
there is a market in the area(dummy)	0.0007	0.0060	0.9015
drive time to next market in minutes	0.0001	0.0001	0.2754
distance to Nairobi in km	0.0003	0.0003	0.2805
other sales party was a trader(dummy)	0.0190***	0.0064	0.0034
sale took place at the farmgate	0.0005	0.0067	0.9433
quantity sold in kg	0.0000	0.0000	0.7358
education of hh head (Years)	-0.0006	0.0007	0.3942
household head is female (dummy)	0.0022	0.0076	0.7719
Season(dummy)	-0.0076	0.0073	0.2979
constant	0.1049***	0.0130	0.0000
R^2	0.0605		
Prob F>F*	0.156		
N	258		

Table 2.7: Relative and absolute transaction costs

Total/Relative Tcs	Model	Mean	Standard error	Min	Max
Relative TCs	BC95	0.182	0.171	0.02	0.588
Total TCs	BC95	3.089	3.484	0.255	14.391
Relative TCs	Greene	0.121	0.043	0.045	0.27
Total TCs	Greene	1.93	0.759	0.706	5.141

Figure 2.1: Illustration of transaction costs as distance between observed and feasible price

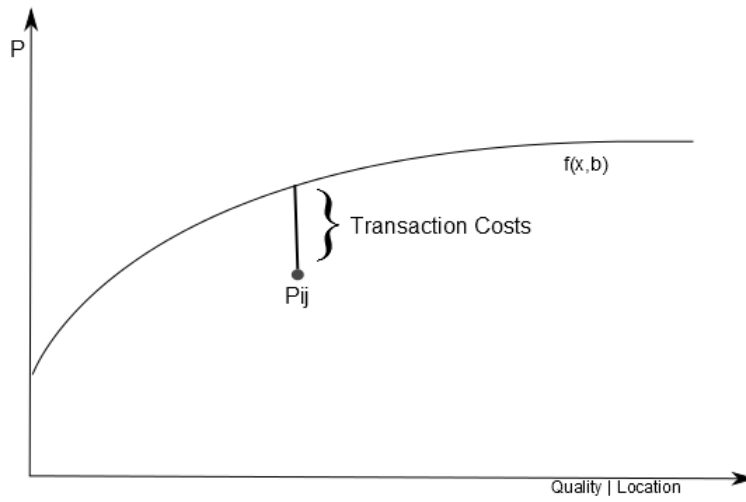
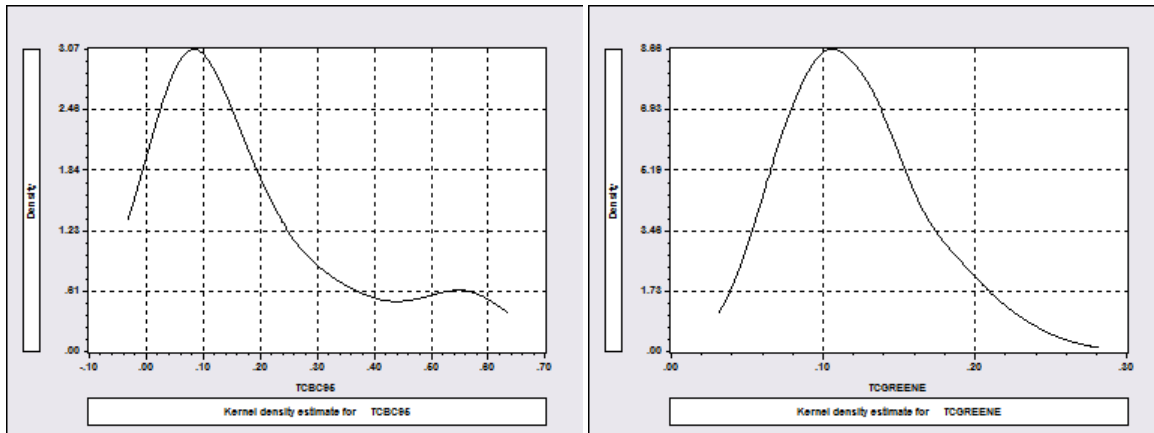


Figure 2.2: Kernel density estimates of transaction costs



Chapter 3

A Latent Class Analysis of the Demand for Food Diversity in India

Abstract

In this paper we study the latent demand structure for food diversity in India using data from the 2012 Consumer Expenditure Survey. We assume that consumers who have not yet attained calorie sufficiency favor calories over food diversity and once passing a threshold of subsistence substitute away from staples towards a more varied diet. This implies a latent demand pattern as calorie sufficiency depends on unobservable individual characteristics. Latent classes and demand patterns are identified by means of finite mixture models. Therefore we examine the link between food diversity indices and socioeconomic indicators, explain component memberships in order to characterize latent classes and evaluate nutritional implications. Two clearly distinct demand patterns for diversity could be identified, consistent with the initial assumptions. The identified classes differ substantially in income, household composition and nutritional adequacy.

Key Words: food diversity, India, finite mixture models, consumer demand

JEL: D12 I32

Introduction

Recent events such as the Agenda 2030 show that development goals are shifting beyond the mere elimination of hunger towards the improvement of food security and nutrition as an essential part of sustainable development. The Millennium Development Report 2015 accounts for the successful reduction in extreme poverty (UN, 2015) in India. However despite high growth rates in the last decade undernutrition levels in India remain higher than for most countries of Sub-Saharan Africa, even though those countries are currently much poorer than India (Deaton and Drèze, 2009).

A study that grew famous for unraveling the severity of this situation with respect to the nutritional status of children is the National Family Health Survey 2005-06 that showed that 48% of all under five year old children are chronically malnourished (stunted) and 54% of all death of under five year old children are related to malnutrition. While an adequate intake of calories is essential for survival it does not suffice to maintain health. In fact Deaten and Drèze (2009) find that in some regions of India malnutrition is worse although calorie intakes are higher.

Nutritionists seem to agree that the consumption of a variety of foods is a key feature in achieving a healthy diet that can translate into the sufficiency of a variety of nutrients including vitamins and minerals that are required to maintain health. In the light of prevailing nutrient deficiencies a better understanding of the demand structure for food diversity could serve as a powerful tool in fighting malnutrition or health issues related to dietary patterns like chronic diseases, the double burden of malnutrition and other non communicable diseases.

The current literature on the demand for food diversity mostly focuses on the inherent relation between food diversity and income within linear demand models. However especially for developing countries, where poverty and hunger are decisive when it comes to food choices, more complex demand structures should be taken into consideration. Bennett's Law (Bennett, 1941) states that consumers substitute away from starchy staples as their income increases. This implies that consumers add food items from food groups other than starchy staples to their diet, which usually leads to a higher level of food diversity and improved nutrition. Jensen and Miller (2011) extend this theory and assume that a consumer follows a different demand pattern depending on whether he faces subsistence concerns. Since a consumers status has a major influence on his marginal

utility of calories and other food attributes it can be reasoned that it will also affect his choice of food diversity. However the minimum calorie requirements for subsistence vary individually and are unobservable. Consequently the remaining challenge for demand analysis is that there are multiple observable demand behaviors based a consumer’s unobservable status. If members of subgroups cannot be distinguished based on their observable characteristics it is common to talk about latent classes.

While the literature on demand systems has already recognized the importance of identifying latent classes in demand systems (Zhou and Yu, 2014), mixed behaviors were not yet considered in the analysis of the demand for food variety. In traditional demand analysis it is usually assumed that all households belong to the same population and hence exhibit identical consumption patterns. However there may be several distinct patterns that consumers follow with mixed probabilities.

In this paper we use the basic intuitions from Jensen and Miller (2011) to reason observed demand for food variety is the outcome of two distinct demand patterns that consumers follow with mixed probabilities. As a measure for food diversity we use food item counts. We estimate the demand model by means of a finite mixture model using data from the 68th round of the CES consumer survey, analyze posterior component probabilities and evaluate nutritional implications of class membership. Results suggest that two latent demand patterns exist that are consistent with the theoretical assumptions. The identified latent classes differ substantially in income, household composition and nutrient adequacy.

3.1 Literature Review

To date more attention has been paid to the supply of variety than the demand (Gronau and Hamermesh, 2008). Beyond income and prices, diversity has been recognized as an important determinant of the variation in consumer demand. An early study to consider the consequences of consumers’ desire for variety in economic modeling is Dixit and Stiglitz (1977). They evaluate the implications of Chamberlin’s monopolistic competition model at the social optimum regarding the relation of market and resource allocation. However instead of modeling diversity directly they consider desire for variety as an inherent property of a weakly separable utility functions with convex indifference surfaces. Consequently if a consumer is indifferent between the quantities (0,1) and (1,0) of two different products he will prefer the set (1,1) to either (0,2) and (2,0). Their results contributed

to the "excessive diversity" vs. "excessive capacity" debate by showing that neither extreme is an outcome in the social optimum. Benassy (1996) uses the tool set proposed by Dixit and Stiglitz (1977) and concentrates on the derivation of a "taste of variety" parameter that expresses the utility gain from consuming a variety of goods instead of concentrating on a single one.

A study that focuses on the link between income and the demand for variety is presented by Jackson (1984). Jackson suggests a concept to model a hierarchy of purchases and proposes a modified utility function to adapt to such a hierarchy. If consumer demand follows a hierarchical structure consumers with a low income purchase a very limited set of items and will add more items as income increases leading to continuing growth in diversity at all income levels. Jackson emphasizes the gravity of non-negativity constraints in a demand system. His intuition was based on Houthakker (1954) who shows by means of quadratic utility functions that commodities can enter and leave the budget. By adding additional constraints to a general utility maximization problem Jackson shows that in his framework, at a given price, the number of items in the purchased set of commodities is an increasing function of income. For his analysis Jackson used the US Consumer Expenditure Survey which was conducted in 1972-1973. He confirms the strong link between income and variety using a count measure to represent diversity.

Another theoretical framework that aims to explain differences in consumers demand for variety comes from Gronau and Hamermesh (2001) who use a household production model in which households produce activities. They explain differences in demand through differences in the opportunity costs of activities in terms of time costs and pay especially attention to the positive correlation between the educational level and the demand for variety.

Most empirical studies on the demand for variety focus on developed countries. Thiele and Weiss (2002) analyze the demand for food diversity in Germany. As a measure of diversity they use the logit transformed Berry Index and the Entropy index. As the main positive drivers of the demand for food variety they identify income and the number of children between 7 and 17 years. Single male households demand significantly less variety. In Addition they find strong regional effects and food variety is higher in large East German cities. A significant influence of the educational level could not be confirmed.

Consumers in developing countries may however exhibit different pattern than those in developed countries. For instance due to differences in income, infrastructure or food supply. So far only a few studies that perform an analysis on a micro level are available for developing and transition

countries. Moon et al. (2002) presents a study on Bulgaria and find strong regional effects. High age, low incomes and low levels of education are associated with a less varied diet. Cupak et al. (2014) present a 2-stage least squares approach to estimate demand for diversity in Slovakia and find significant increases of income elasticities over time resulting in a convergence towards the levels of Western European countries. Thorne-Lymann et al. (2014) examine the link between dietary diversity scores and socioeconomic status in Bangladesh. They establish a strong link between diversity and income although a varied diet was low throughout all income groups. They also confirm that a low level of diversity increases the risk of malnutrition.

A diversified diet has important implication concerning the consumer's health and is hence of economic significance. The proposition that food diversity allows an inference on health implies a link between food diversity and nutrition or dietary quality. Hatloy et al. (1998) show that such an inference is possible. They compare Diet Diversity Scores, which are defined as the number of food groups consumed, with nutrient adequacy ratios that compare actual nutrient intakes with intake recommendations. The authors conclude that it is possible to infer on nutritional adequacy from dietary diversity.

3.2 Theory

The basic characterization of latent classes that is used here to model observed demand for food diversity as the outcome of at least two distinct behaviors is based on the revealed preference framework from Jensen and Miller (2010). They propose a new measure of nutritional sufficiency based on observed consumption. The proposition was motivated by the problem that whether an individual has achieved calorie sufficiency or overcome hunger cannot be adequately determined by a calorie threshold. However individuals reveal their nutritional status in their consumption pattern. In order to determine an individual's nutritional status they assume that a consumer exhibits different food consumption patterns when facing subsistence concerns and when he is outside of subsistence concerns. Under subsistence concerns a consumer suffers a significant disutility from calorie deprivation. The disutility can appear in the form of hunger, headaches and other side effects arising from calorie deprivation. In this state the marginal utility of calories is very high and the cheapest sources of calories will be demanded by a utility maximizing consumer. Consequently the proposed measure of nutritional sufficiency is the staple calorie share (SCS) that is defined as the

ratio between calories from staples and total calorie consumption. This ratio remains high as long as subsistence needs are not met and starts to decline once calorie sufficiency has been achieved. The cheapest sources of calories in India are typically rice and wheat that are also available via the public distribution system to which a consumer has access when he possess a ration card. Once the subsistence threshold is passed marginal utilities of calories decline the consumer starts substituting away from staples towards more expensive calories that he favors over other attributes such as taste or diversity. This model implies a latent demand structure for food diversity. Diversity will have a low marginal utility in contrast to calories as long as an individual faces the primal instinct of survival or seeks to avoid hunger. In the absence of hunger marginal utility of additional calories declines due to repletion and consumer may favor attributes like diversity over calories to realize further utility gains from food consumption.

The decisive connection between the present study that analyzes the latent demand structure for food diversity and the revealed preference framework is the assumption on how consumer preferences change with calorie sufficiency. In terms of calories there is no clear cut off that would separate these two behavioral patterns. Jensen and Miller emphasize that there is no clear consensus on a minimum calorie threshold. Consequently the subsistence threshold is individually varying and is hence unobservable since it depends on the absorption efficiency of the individual and thus represents a source of unobservable heterogeneity. Deaton and Drèze (2009) also doubt the usefulness of calorie norms since there are too many sources of variation such as health and activity levels. Further support for the relevance of the underlying demand structure comes from Zhou and Yu (2014) who are concerned with the resulting implications for calorie elasticities. They examine the relation between income and calorie consumption and verify the existence of a latent demand structure, which is consistent with the behavioral assumptions from Jensen and Miller (2010)

Similar to Drescher et al. (2009) who performed a demand analysis for Germany based on the healthy food diversity index, we assume a demand function of the form:

$$d = d(P, Y, K) \tag{3.1}$$

Here the demand for food diversity depends on prices, income and consumer characteristics. If we assume that observed demand for diversity is the result of C different demand behaviors that consumers follow with mixing probability π_j equation (3.1) can be rewritten as the weighted sum

of C distinct demand functions.

$$d = \sum_{j=1}^C \pi_j d_j(P, Y, K); \quad \sum_{j=1}^C \pi_j = 1 \quad (3.2)$$

The revealed preference framework suggests that consumers behave differently depending on their calorie sufficiency. Consequently two different classes are to be distinguished to model the demand for food diversity.

$$d = \pi_1 d_s(P, Y, K) + \pi_2 d_{ns}(P, Y, K), \quad \pi_1 + \pi_2 = 1 \quad (3.3)$$

While in concern about subsistence (d_s) calories are more important than diversity in order to avoid the disutility. A consumer with no subsistence concerns (d_{ns}) favors food diversity over calories as his basic needs are already satisfied. Disutility penalizes the process of substituting away from staples towards higher levels of food diversity if subsistence calories have not been attained yet. Consumers substitute away from staples if they do not experience a disutility in the form of hunger or other symptoms of deprivation. A consumer who is sufficiently wealthy to afford any food consumption bundle he desires will not have to face any disutility. A consumer's mixture probability is hence determined by the penalty he faces due to his individual distance to his calorie subsistence threshold. Ultimately each observed realization of this demand function is the outcome of both demand patterns with mixing probabilities π_1 and π_2 .

Figure 3.1 shows the assumed relationship between food diversity and income. The line below the subsistence threshold corresponds to the demand behavior d_s and above to d_{ns} . The income elasticity of food diversity increases after passing the unobservable subsistence threshold. However food diversity has a natural limit and will converge to certain limit for high income classes. On the one hand diversity is limited by supply on the other it is reasonable to assume that increasing efforts are required to further diversify an already well balanced diet. This circumstance would be sufficient to consider the existence of a potential third class. While this insight deserves testing it is beyond the here considered theoretical framework that concentrates on behaviors around a subsistence threshold.

Table 3.1 gives an overview on the expected differences between the latent demand patterns and consumer characteristics. Under d_s in contrast to d_{ns} the consumption basket is expected to have a lower level of food diversity and consumers are expected to earn more on average. The income

elasticity of d_s is lower than of $d_n s$ since consumers favor calories over diversity. This implies that the marginal utility of calories is higher than the marginal utility of diversity with d_s and the other way round under $d_n s$. Consequently the marginal utility of calories is higher under d_s than $d_n s$. Zhou and Yu (2014) test this implication of the revealed preference framework and find that calorie elasticities decrease in income.

Consumers reveal their preferences and with it their nutritional status in their consumption patterns. If they substitute away from staples it is a signal that the subsistence threshold has been surpassed and calorie elasticities are decreasing. Jensen and Miller(2010) suggest using the share of calories from staples in total calorie consumption as an indicator of nutritional sufficiency. In the empirical model this insight can be used to model mixture probabilities. Figure 3.2 shows the relation between the SCS and log income by means of a local polynomial regression. For India the two variables have a clear negative relation similar to the findings of Jensen and Miller (2010) for China. In the case of India however changes of calorie preferences with income appear to be smoother which renders the position of a subsistence threshold less obvious.

3.3 Data

Our analysis is based on the Consumer Expenditure Survey (CES) which has been conducted by the National Sample Survey Office (NSSO). The NSSO has been founded in 1950 by the government of India and belongs to the Ministry of Statistics and Programme Implementation. It regularly conducts consumer expenditure surveys nationwide throughout India. For our analysis we use the 68th round of the survey which was conducted in 2012. Over 4 sub rounds 101626 household were interviewed. The survey contains information on consumption expenditure over the last 30 days and provides a high level of detail on food expenditures. In total it provides quantity and expenditure information of 142 food items. After dropping tobacco and liquors 127 food items remain for the analysis.

Using a 30 day recall period for a dietary analysis is sometimes criticized as it is prone to memory errors. Ruel (2006) suggests that a 7 day recall might be optimal to minimize recall error. The NSSO (2002) found in a pilot study that frequently purchased items were recalled more precisely on a monthly than a weekly base while it was the other way round for infrequently purchased goods. While memory errors might represent a weakness in the assessment a longer period of observation

may reveal dietary patterns more clearly. A very interesting study that allows inference on the development of food diversity over time is presented by Drewnowski et al. (1997) who collected detailed food consumption data from individuals on 15 consecutive one day recalls. They find that food diversity is increasing over the whole period of 15 days although converging towards a steady level during the last days. Hence observations over a longer run can be helpful in dietary pattern analysis, especially since some consumers may repeat certain consumption patterns more frequently than others which is not necessarily observable in shorter periods.

A well accepted measure to evaluate of the nutritional status of the households within the latent classes would be the nutrient adequacy ratio. The calculation of nutrient adequacy ratios requires the construction of a conversion table to calculate nutrients from food item quantities and the recommended dietary allowances. Conversion tables were extracted from the database Nutrisurvey and dietary allowances are provided by the national dietary guidelines from the National Institute of Nutrition (2010). The guidelines contain nutrient intake recommendations for 12 groups that differ by age and gender. Since the CES has details on age and gender of all household members the information from the guidelines can be used for accurate calculations of dietary allowances on a household level.

Outliers were removed that were in terms of calories too high or low to be realistic. Following the example of Wiesmann et al. (2009), Foote et al. (2004) and Lovon and Mathiassen (2014) all observation below 500 calories and above 5000 calories per capita per day were removed. As a result 2151 observations were dropped.

3.4 Empirical Model

If there are classes to be distinguished by their demand patterns, then these would follow different probability distributions. In order to identify latent behaviors that consumers follow with mixed probabilities finite mixture models (fmm) are an adequate instrument of analysis. In a finite mixture model the population is assumed to consist of an additive mixture of C subpopulations that are mixed in proportions π_1, \dots, π_C . Each of the distinct probability distributions is assumed to arise from the same parametric family $(d_i|\theta)$, where the diversity index represents the latent variable and θ the set of parameters. The joint density function of the population is then given by:

$$f(d_i|\theta_j) = \sum_{j=1}^C \pi_j(z_i) f_j(d_i|\theta_j) \quad (3.4)$$

The mixture probabilities π_j satisfy the restrictions $\sum_{j=1}^C \pi_j = 1$ and $\pi_j \geq 0$. The component probability π_j can be modeled as a function of z_i . This can be especially useful in the presence of large overlaps within the mixture distribution. In this case we use the staple calorie share to model the component probabilities and the number of meals consumed outside the household as an additional control variable. In a two component model the prior component probabilities are then estimated as follows (Ayyagari et al., 2013):

$$\pi_{ij} = \text{logit}(z_i|\delta), \quad 0 \leq \pi_{ij} \leq 1, \quad \sum_{j=1}^C \pi_{ij} = 1 \quad (3.5)$$

As a measure of food diversity we choose food item counts. In the light of the present analysis this measure has several advantages. Count indices can be analyzed with standard econometric methods without the necessity of any transformation to meet distributional assumptions. This allows a straightforward interpretation of the results. Beyond that counts of food groups or items are the standard unit used in dietary recommendations. Another popular choice for a diversity index is for instance the berry index. However since this variable is bounded between zero and one it requires a log transformation prior to its inclusion in a model with normal distribution assumption or a model that can be fitted to these properties such as a beta model. While the berry index usually has a better correlation with nutritional adequacy count indices allow for a straight forward interpretation and can be analyzed with standard econometric methods within a finite mixture approach.

Count data follows a discrete distribution. It is usually assumed to follow a Poisson or negative binomial distribution. However the Poisson distribution, as a one-parameter distribution, tends to underestimate the variance while negative binomial models allow for over dispersion. For the analysis of the count measure for food variety we follow Moon et al. (2002) who use a negative binomial model. The probability to observe the consumption of n food items is given by the density function:

$$f(d_i|\theta) = \sum_{j=1}^C \pi_j \frac{\Gamma(d_i + \psi_{j,i})}{\Gamma(\psi_{j,i})\Gamma(d_i + 1)} \left(\frac{\psi_{j,i}}{\lambda_{j,i} + \psi_{j,i}} \right)^{\psi_{j,i}} \left(\frac{\lambda_{j,i}}{\lambda_{j,i} + \psi_{j,i}} \right)^{d_i} \quad (3.6)$$

In order to fit a model for the final analysis the Poisson, NB1 and NB2 models are estimated with two and three components. Ex post to the fmm estimation the posterior component probabilities can be estimated by means of Bayes rule:

$$Pr(d_i \in j|\theta, d_i) = \frac{\pi_j f_j(d_i|\theta)}{\sum_{j=1}^C \pi_j f_j(d_i|\theta)} \quad (3.7)$$

The estimated posterior probabilities allow the analysis of determinants of class memberships. This could be achieved by regressing a vector of socioeconomic variables on the probability of being member in component j . The posterior probabilities are however continuous on the open interval (0,1). An analysis by means of ordinary least squares would be flawed as the fact that the variates are bounded violates OLS assumptions (Kiesschnieck and McCullough, 2003). The error distribution of regression models using a bounded dependent variable is heteroskedastic as the conditional variance approaches zero when the mean approaches the boundaries. Kiesschnieck and McCullough (2003) analyze the fit of 7 different regression models for variables on the open interval (0,1) and recommend the use of a beta regression model for the analysis of this type of data. Smithson and Verkuilen (2006) agree with this conclusion. In order to explore the influence of household characteristics on posterior probabilities we hence use a beta regression model as proposed by Smithson and Verkuilen (2006) and Ferrari and Cribari-Neto (2004). With the component membership from the estimates it is possible to explore further nutritional implications of this class. An observation is attributed to a component if the respective component probability is greater or equal 0.5. The comparison of mean nutritional intakes and adequacy ratios across components yields information about distinctiveness of component members in terms of nutrients.

$$NAR_{li} = \frac{N_{li}}{\sum_{j=1}^{12} RDA_{lj} M_j} \quad (3.8)$$

The nutrient adequacy ratios are calculated by dividing all nutrients of nutrient l consumed by household i by the sum of dietary allowances over all household members. The sum of dietary allowances for a nutrient is here calculated by summing up the products of each of the 12 RDA groups with the number of household members in this group. The consideration of gender and age groups in the calculation of nutrient adequacy ratios gives a very precise picture of the household's nutritional status.

3.5 Variables

As a measure of food diversity we choose simple food item counts. The index is calculated by summing up the number of distinct food items that were consumed by the household over the 30 day recall period. The log of monthly per capita expenditure and the amount of land possessed were included as measures of income and capital. Due to the existence of zeros and ones we added one ha before taking logs of the amount of land. The 2015 Millennium Development Goals country report suggests that there are still large differences in the nutritional status between rural and urban regions. Hence a dummy was added to indicate whether the household is located in a rural area. Another dummy indicates whether the household is in possession of a ration card. A ration card gives access to very cheap calories via the public distribution system. Jensen and Miller (2011) find that subsidizing staples can result in consumers substituting away from these. Consequently ration cards are expected to have a positive effect on food diversity. Especially in rural areas of India we find many semi subsistence farmers. The survey contains the additional information if a consumed food item was purchased or produced at home. In order to capture the effect of home produced food consumption the ratio of the quantity of consumed home produced food and total consumed food was added as a variable. Characteristics on the household head include years of education. This variable was here derived from the stated general educational level. The relevance of the educational level is regularly discussed in the context of demand for diversity and is usually expected to have a positive impact. The food consumption data does not contain detailed information on consumption outside the household. However we can calculate the number of meals per household member outside the household that were consumed over the observed 30 day recall period. Additionally the average number of meals per household member was added for the same time period. Similar to Thiele and Weiss (2001) we count household members of different age groups. We thus count the number of children from 1-6 years, 7-13 and 14-17. In addition we add a count of elderly members with age 60 and above, the number of women and add the size of the household. Eating habits in India are often shaped by beliefs. For instance Hindus do not eat beef and many are vegetarians. Muslims avoid pork and Sikhs beef. We hence include religion dummies for Hinduism, Islam, Christianity, Sikhism, Jainism and Buddhism to control for possible influences on dietary diversity. In order to control for price differences across regions 35 state dummies were added. The smallest seven of these which make up for 2.3 % of the observations

were assigned to the base group. The assignment of too many state members of these small states in the latent class analysis resulted in the failure to calculate the respective coefficients and hence evaluate the model fit. As the survey was conducted in 4 sub rounds 3 sub round dummies were included to control for possible seasonal effects. Two variables were chosen to model the component probabilities. The first one is the number of meals outside the home. This part of nutrition is not captured by the data and hence needs to be controlled for. Households that eat out a lot may show lower levels of food consumption and diversity according to the data. Due to a supposedly large overlap in the distribution of dietary diversity such households may be attributed to a class with a lower mean of diversity than the one they rather belong to in terms of their actual level of diversity. The second variable to model the component probabilities is the staple calorie share. This indicator is here calculated as the ratio of calories obtained from rice and wheat in the total calorie consumption of the household. One of the basic insights from Jensen and Miller (2011) is that a consumer's subsistence threshold is unobservable as it depends on unobservable individual characteristics. However he reveals his status in his consumption pattern. In this case we can use the calorie staple share to model the mixture probabilities as a function of the consumer's revealed preferences.

3.6 Results

3.6.1 Model Selection

Among the single equation models the Akaike information Criteria the Bayesian Information Criteria and the log likelihood all favor unanimously the negative binomial 2 model over the Poisson and the negative binomial 1 model. The χ^2 based goodness of fit test for the Poisson model rejects the Null that the count index follows a Poisson distribution. The χ^2 test statistics that were estimated along with the NB 1 and NB 2 both reject the hypothesis that the data follows a Poisson distribution. This is consistent with the summary statistics of the count index from table 3.7 that clearly shows that the variance exceeds the mean greatly. Hence there is evidence of overdispersion which is more adequately modeled by means of a negative binomial model than a Poisson model which assumes equidispersion where the mean is identical to the variance.

In total six finite mixture models were estimated with two and three components for Poisson, NB1 and NB2 models. None of the three component models converged. Among the two component

models only the Poisson model converged. This circumstance could not be altered by changing model specifications or maximization algorithms. However all calculated information criteria show that the two component Poisson model fits the data better than the best fit among the single component models. This indicates that a mixture model might be better suited than a single equation models to analyze the demand for food diversity.

3.6.2 Characterization of Latent Classes

Before analyzing the demand model estimations differences between latent classes are compared in order to evaluate whether the class characteristics match the theoretical assumptions. Table 3.6 of Appendix B contains the summary statistics of the nutritional status and income for each component. Component 1 covers 28583 observations and 28.77 % of the data sample. Component 2 contains 70765 observations which make up for 71.23 % of the sample. Differences in income are striking. The mean per capita income in the first component is 1065 Rupees and the second one is with 2032 Rupees almost twice as large. On average the households in component 1 consume 30.44 distinct food items and the ones in component 2 consume 39.88 items. In terms of calories the classes are almost identical. This is also true for calorie adequacy and protein adequacy although fat adequacy of component 2 is with 96.81 % is clearly higher as compared to component 1 with 78.05 %. In component 2 adequacy ratios are higher for all tested micronutrients without any exceptions. The mean nutrient adequacy component 2 is with 79.48 % 12.74 percentage points higher than that of component 1. Figures that stick out are especially the adequacy of vitamin A that is critically low in component 1 with 50.12 %. The vitamin A adequacy of component 2 is significantly higher with 74.4 %.

So far the findings are consistent with theory. One class could be identified with lower income and a lower level of food diversity that also accounts for higher incidence of malnourishment. Additionally the level of calorie consumption is comparable in both classes indicating that members of component 2 have shifted their consumption towards more expensive calories while achieving higher levels of food diversity as compared to members of component 1. Therefore the probability to be a member of component 1 corresponds to π_1 of the theoretical model. As these characteristics are in line with the initial assumptions it is possible to match the empirical outcomes to the theoretical model. Thus component 1 refers to the class with subsistence concerns and component 2 to the class without subsistence concerns. Justified alternative labels in line with the identified characteristics for these

classes could also be deprived/well-nourished, poor/rich, hungry/sated or one-sided and diversifying food consumers.

Table 3.5 contains the results of the beta regression and the corresponding marginal effects. OLS results were included as well to be available for model comparison. According to the F and χ^2 statistic both models show overall a very good fit although the coefficients in the models differ considerably in terms of magnitude and significance. However the Null hypothesis that the dependent variable follows a normal distribution is strongly rejected for tests on skewness and kurtosis. Also the AIC and BIC indicate that the beta model has a better fit than the OLS model. The average marginal effects of the beta model show the probability of being a member of the class with subsistence concerns decreases in income and the amount of land. The signs of the coefficients are as expected since these are determinants of the household's purchasing power that enables the household to consume a higher level of food diversity. Further variables that have a significant negative effect are the share of home produced food, the number of meals outside the hh, years of education, the number of children between 1 to 6 years and 7 to 13 years as well as the number of women and elders. The probability is positively affected by a location in rural areas, the possession of a ration card, the household size and the number of older children between 14 and 17 years.

Figure 3.3 shows the local polynomial regression of the posterior probability of being a member of the class with subsistence concerns on mean nutrient adequacy. There is a clear negative relation between the two variables. So a high probability to be a member of the class with subsistence concerns is associated with a lower level of nutrient adequacy. Consequently, since there are only two components, it is the other way round for the probability of being a member of the class without subsistence concerns.

3.6.3 Demand Model Estimations

Table 3.3 contains the parameter estimates of the NB2 model and two component finite mixtures Poisson model. The Wald χ^2 statistic shows that models have an overall good fit. The results of the mixture model reveal that the staple calorie share is significantly related to the component probabilities while the number of meals consumed outside the house appears to be unrelated.

Table 3.4 shows the average marginal effects of the models. The effect of income on diversity is 21% higher for the class that has attained a subsistence level of nutrition. Below the subsistence threshold consumers are rather concerned with survival where the marginal utility of calories is still

very high. Consequently we observe a higher income elasticity above subsistence threshold where the marginal utility of calories is expected to be lower as compared to the marginal utility of other food attributes such as diversity. Living in a rural area has a negative effect only on the class with no subsistence concerns while the poorer class remains unaffected. The infrastructure and supply in rural areas appears sufficient to meet basic needs but impedes diets beyond that. Consumption from home produced food has a negative effect for both classes. Since the diversity in home production usually has sensible limits increasing this share in food consumption decreases the number of food items consumed. The possession of a ration card has a positive effect in both classes. In the class that struggles with subsistence the possession of a ration card increases the number of food items by 2.3 on average. The effect is slightly lower with 1.9 food items on average in the class that has attained subsistence calories. However these effects are to be related to the mean diversity of the respective classes. This confirms the finding of Jensen and Miller (2011) that subsidizing staples can result in consumers substituting away from these. This again results in a higher level of food diversity. The years of education have a significant positive effect in both classes. Moon et al. (2002) suggest that more educated consumers are more concerned about nutritional balance and demand more diverse diets. The effect of education is however more than 2 times larger in the class with no subsistence concerns. A possible explanation for this is that while higher education promotes better diets consumers still require the resources to realize better diets. Religions have no effect on food diversity in the poorer class however all considered religions except for Janism have a positive effect on diversity within the class without subsistence concerns. Culturally specific eating habits appear to require a certain level of food diversity that is not achieved as long as calorie sufficiency is the primary focus of a diet. The household composition reveals some interesting differences across classes. The number of children between 1 and 6 years as well as the number of children between 7 and 13 has significantly negative effect on diversity in the class with subsistence concerns. In the class with no subsistence concerns the effect is significantly positive. In developed countries we usually find that the number of children positively affect the households food diversity as for instance in the study of Thiele and Weiss (2001) who analyze German consumption data. One possible explanation for this effect is that parents want to offer their children healthy diets. The negative effect in the class that struggles with subsistence indicates that consumers are unable to do so and budget constraints force households to substitute towards cheaper calories if they need to provide for more children. The number of elder household members of age 65 and older has a small

negative effect in both classes. Moon et al. (2002) argue that older consumers are less likely to seek variety than younger generations. The coefficient on the number of women is insignificant in the poorer class but has significant positive effect at a 10 per cent level in the richer class. Lee and Brown (1989) also find that an additional female household member has a greater positive effect on the households food diversity than an additional male member and suspect that women have a greater interest in the preparation of various foods or nutrition. Most of the included state variables were highly significant suggesting strong regional effects. These may occur due to differences in supply or regional price differences.

3.7 Conclusions

In this paper some intuitions of Jensen and Miller (2011) are extended to a latent demand model for food diversity that can be estimated by means of a finite mixture model. The estimates from the demand model clearly show two distinct demand patterns that are consistent with our basic assumptions. Consumers that live near their individual subsistence threshold strive for survival resulting in lower income elasticity while the marginal utility of calories remains high. Once having passed that threshold the income elasticity of food diversity increases as consumers do not suffer disutility from deprivation anymore. The classes differ widely in terms of nutrition consistent with the theory. Under subsistence concerns more staples are demanded resulting in higher levels of malnutrition than the better off class that fares better with all micro nutrients while maintaining a comparable calorie intake. Beyond the confirmation of the underlying hypothesis the estimates reveal further interesting insights. It is often argued that better education could improve nutrition. While a significantly positive effect can be confirmed, the magnitude of the effect is much lower in the class with subsistence concerns. On the one hand this suggests that the effect of education might be overestimated when looking at the whole population on the other hand the effect might be constrained by income. The finding that children have a negative effect in the poorer class although they affect food diversity usually in a positive effect might be interpreted as a severe sign of neediness since households reduce their dietary quality to ensure the survival of all household members. This effect is easy to overlook since the one component NB2 model shows an overall positive effect. Beyond that the present analysis reveals some shortcomings of calorie based indicators as we find classed with vast differences in their nutritional status and identical calorie intake. The existence of

latent behaviors and unobservable heterogeneity in the demand for food diversity poses challenges not only for demand analysis but also for assessment of nutrition and poverty. The mere insight that such patterns exist can however aid to better identify households that are vulnerable to malnutrition.

3.8 Appendix B

Table 3.1: Expected class differences

	With Subsistence Concerns (d_s)		Without Subsistence (d_{ns})
Mean level of food diversity	Mean food diversity (d_s)	<	Mean food diversity (d_{ns})
Income elasticity	Income elasticity (d_s)	<	Income elasticity (d_{ns})
Marginal utility of calories	Marginal utility of calories (d_s)	>	Marginal utility of calories (d_{ns})
Mean income	Mean Income (d_s)	<	Mean Income (d_{ns})
$\frac{\text{Marginal utility of calories}}{\text{Marginal utility of diversity}}$	> 1		< 1

Table 3.2: Model selection criteria

Model	Df	ll	AIC	BIC
Poisson				
1 component*	52	-789890401	1579780906	1579781400
2 components	107	-776290487	1552581189	1552582207
Negativ Binomial 1				
1 component**	53	-789501912	1579003931	1579004435
Negativ Binomial 2				
1 component***	53	-789291816	1578583738	1578584242

*Goodness of Fit Test: $\chi^2 = 267000000$ $\bar{\chi}^2 = 0.000$

**Likelihood-ratio test of $\delta = 0$: $\bar{\chi}^2 = 780000$ $\bar{\chi}^2 = 0.000$

***Likelihood-ratio test of $\alpha = 0$: $\bar{\chi}^2 = 1200000$ $\bar{\chi}^2 = 0.000$

Table 3.3: Parameter estimates of the demand model

Component	NB2		FMM-Poisson			
	Coef.	$P > z $	Component 1		Component 2	
	Coef.	$P > z $	Coef	$P > z $	Coef	$P > z $
Log income	0.1674	0.000	0.1391	0.000	0.1408	0.000
Rural	-0.0322	0.000	0.0024	0.746	-0.0276	0.000
Log Land	0.0044	0.000	0.0026	0.051	0.0041	0.000
Home Produced Share	-0.0729	0.000	-0.0388	0.002	-0.0964	0.000
Ration card	0.0556	0.000	0.0729	0.000	0.0509	0.000
P.c. Meals out	-0.0037	0.000	-0.0008	0.696	-0.0026	0.000
Pc meals at Home	0.0004	0.000	0.0020	0.000	0.00002	0.880
Education(years)	0.0064	0.000	0.0028	0.042	0.0067	0.000
Hinduism	0.0273	0.113	-0.0257	0.445	0.0748	0.001
Islam	0.0319	0.068	-0.0055	0.873	0.0729	0.001
Christianity	0.0142	0.426	-0.0356	0.290	0.0690	0.003
Sikhism	0.0349	0.061	-0.0218	0.622	0.0819	0.000
Jainism	-0.0104	0.633	-0.0539	0.294	0.0354	0.169
Buddhism	0.0275	0.162	0.0218	0.594	0.0681	0.004
Household Size	0.0342	0.000	0.0403	0.000	0.0347	0.000
Children 1-6y	0.0063	0.000	-0.0157	0.000	0.0086	0.000
Children 7-13y	0.0086	0.000	-0.0088	0.029	0.0121	0.000
Children 14-17y	-0.0041	0.030	-0.0030	0.452	-0.0027	0.214
Elders $\geq 60y$	-0.0159	0.000	-0.0241	0.000	-0.0133	0.000
Women	0.0040	0.004	0.0040	0.164	0.0030	0.063
constant	1.3797	0.000	1.4980	0.000	1.7066	0.000
π_1 :						
			Staple Calorie Share		28.1565	0
			P.c. Meals out		-0.0755	0.157
			constant		-20.0602	0
Wald χ^2	42571		Wald χ^2		31892	
Prob $> \chi^2$	0		Prob $> \chi^2$		0	
N	99508		28583		70765	

Table 3.4: Average marginal effects of the demand model

Component	NB2		FMM-Poisson			
	-		Component 1		Component 2	
	Coef.	$P > z $	Coef	$P > z $	Coef	$P > z $
Log income	6.2307	0.000	4.5254	0.000	5.4580	0.000
Rural	-1.1993	0.000	0.0785	0.746	-1.0710	0.000
Log Land	0.1620	0.000	0.0845	0.052	0.1589	0.000
Home Produced Share	-2.7139	0.000	-1.2614	0.002	-3.7365	0.000
Ration card	2.0695	0.000	2.3704	0.000	1.9748	0.000
P.c. Meals out	-0.1374	0.000	-0.0265	0.695	-0.1000	0.000
Pc meals home	0.0148	0.000	0.0647	0.000	0.0007	0.880
Education(years)	0.2374	0.000	0.0914	0.046	0.2591	0.000
Hinduism	1.0144	0.113	-0.8369	0.444	2.8996	0.001
Islam	1.1860	0.068	-0.1780	0.873	2.8243	0.001
Christianity	0.5296	0.426	-1.1569	0.289	2.6760	0.003
Sikhism	1.2992	0.061	-0.7081	0.622	3.1754	0.000
Jainism	-0.3854	0.633	-1.7527	0.295	1.3743	0.170
Buddhism	1.0241	0.162	0.7091	0.594	2.6402	0.004
Hhsize	1.2718	0.000	1.3096	0.000	1.3435	0.000
Children 1-6y	0.2353	0.000	-0.5108	0.000	0.331	0.000
Children 7-13y	0.3193	0.000	-0.2871	0.029	0.4687	0.000
Children 14-17y	-0.1530	0.030	-0.0961	0.453	-0.1048	0.214
Elders ≥ 60 y	-0.5932	0.000	-0.7853	0.000	-0.5141	0.000
Women	0.1477	0.004	0.1303	0.163	0.1150	0.063

Table 3.5: Regression on posterior probabilities of component 1

	OLS		Beta Regression Model			
	Coef.	$P > z $	Coef	$P > z $	dy/dx	$P > z $
Log income	-0.1794	0.000	-0.8058	0.000	-0.1427	0.000
Rural	0.0555	0.000	0.2122	0.000	0.0376	0.000
Log land	-0.0053	0.000	-0.0275	0.000	-0.0049	0.000
Homeratio	-0.0621	0.000	-0.2830	0.000	-0.0501	0.000
Ration Card	0.0041	0.387	0.0796	0.000	0.0141	0.000
P.c. Meals out	-0.0003	0.588	-0.0088	0.000	-0.0015	0.000
Pc meals home	0.0012	0.000	0.0049	0.000	0.0009	0.000
Education(years)	-0.0029	0.000	-0.0115	0.000	-0.0020	0.000
Hinduism	0.0083	0.827	-0.0350	0.702	-0.0062	0.702
Islam	-0.0142	0.711	-0.0825	0.371	-0.0146	0.371
Christianity	0.0557	0.151	0.1409	0.135	0.0249	0.135
Sikhism	0.0003	0.994	-0.0507	0.610	-0.0090	0.610
Jainism	0.0260	0.517	-0.0149	0.897	-0.0026	0.897
Buddhism	0.0031	0.939	0.1004	0.326	0.0178	0.326
Hhsize	0.0127	0.000	0.0521	0.000	0.0092	0.000
Children 1-6y	-0.0253	0.000	-0.1212	0.000	-0.0215	0.000
Children 7-13y	-0.0029	0.254	-0.0198	0.000	-0.0035	0.000
Children 14-17y	0.0112	0.000	0.0302	0.000	0.0053	0.000
Elders	0.0013	0.625	-0.0217	0.000	-0.0038	0.000
Women	-0.0060	0.010	-0.0253	0.000	-0.0045	0.000
constant	2.2210	0.000	7.9380	0.000		
N	99348		N	99348		
R^2	0.4223		Wald χ^2	56827		
Prob >F	0		Prob > χ^2	0		
F(51, 99296)	1423		Log likelihood	194621		

Table 3.6: Characterization of latent classes

Model	FMM	
Component	Component 1	Component 2
P.C.Income (Rs)	1064.99	2032.15
Count Index	30.4426	39.8841
Staple Calorie Share	0.7739	0.5564
Nutrients		
Calories	1905.809	1937.234
NAR		
Calories	0.8132	0.8159
Protein	0.8487	0.8944
Fat	0.7805	0.9681
Vit. A	0.5097	0.7440
Magnesium	0.9185	0.9509
Zinc	0.7567	0.8123
Iron	0.4932	0.6318
Vit B1	0.6794	0.7922
Vit B2	0.5012	0.6745
Vit B6	0.6688	0.7182
Folate Acid	0.8070	0.9246
Vit. C	0.7179	0.8408
Calcium	0.3277	0.5864
Mean Adequacy	0.6674	0.7948
Observations	28583	70765

Table 3.7: Variable description

Variable	Description	Mean	SD
Count Index	Number of distinct food items consumed by the hh	37.4230	9.7778
Log income	Log of total per capita expenditure in Rupees (0.00)	11.8343	0.6258
Rural	Dummy that takes the value 1 if hh lives in a rural area	0.7008	0.4579
Log land	Log of the amount of land possessed in ha +1	3.7467	2.5895
Home ratio	Ratio of consumed home produced food quantity and total consumed food quantity	0.1518	0.2446
Ration Card	=1 if household posses a ration card	0.8229	0.3817
Education	Years of education of household head	4.7187	3.7626
Pc meals out	Number of meals per capita that were on average consumed within the 30 day recall period outside the household	1.9076	4.3334
Pc meals	Average pc number of meals within 30 day recall period	71.0242	13.8866
Hh size	Number of household members	4.5427	2.1584
Hinduism	= 1 if religious affiliation is Hinduism	0.8312	0.3746
Islam	= 1 if religious affiliation is Islam	0.1189	0.3236
Buddhism	= 1 if religious affiliation is Buddhism	0.0061	0.0781
Christianity	= 1 if religious affiliation is Christianity	0.0241	0.1532
Sikhism	= 1 if religious affiliation is Sikhism	0.0152	0.1225
Jainism	= 1 if religious affiliation is Jainism	0.0027	0.0515
Children 1-6y	Number of children in household age 1-6	0.5167	0.8325
Children 7-13y	Number of children in household age 1-6	0.6740	0.9445
Children 14-17y	Number of children in household age 1-6	0.3728	0.6466
Elders	Number of household members \geq 60 Years	0.3661	0.6402
women	Number of women in Household	2.2013	1.3131

Figure 3.1: Expected relation between income and food diversity

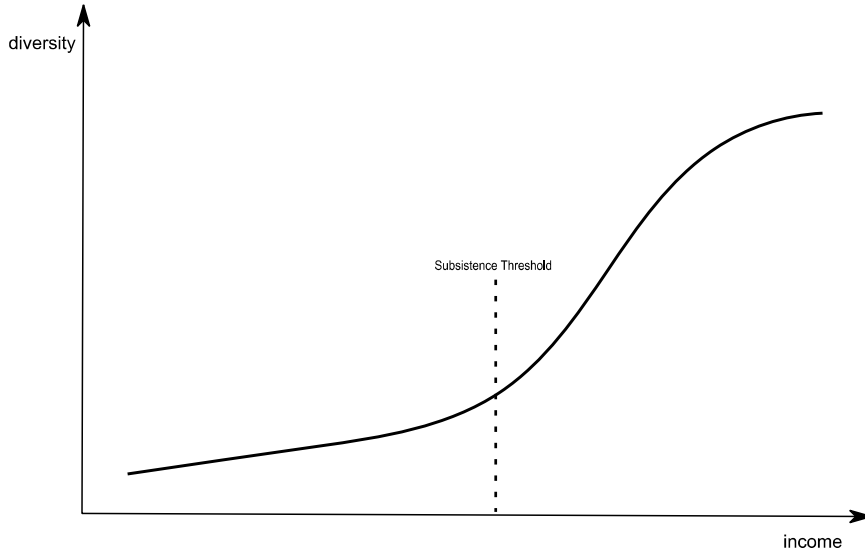


Figure 3.2: Local Polynomial Regression of the Staple Calorie Share on Log Income

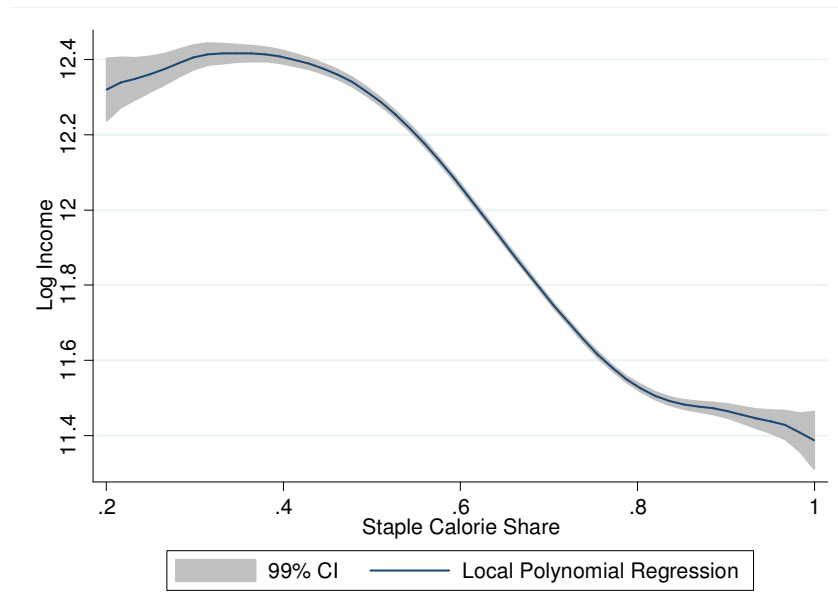


Figure 3.3: Local Polynomial Regression of Posterior Probabilities on Mean Nutrient Adequacy

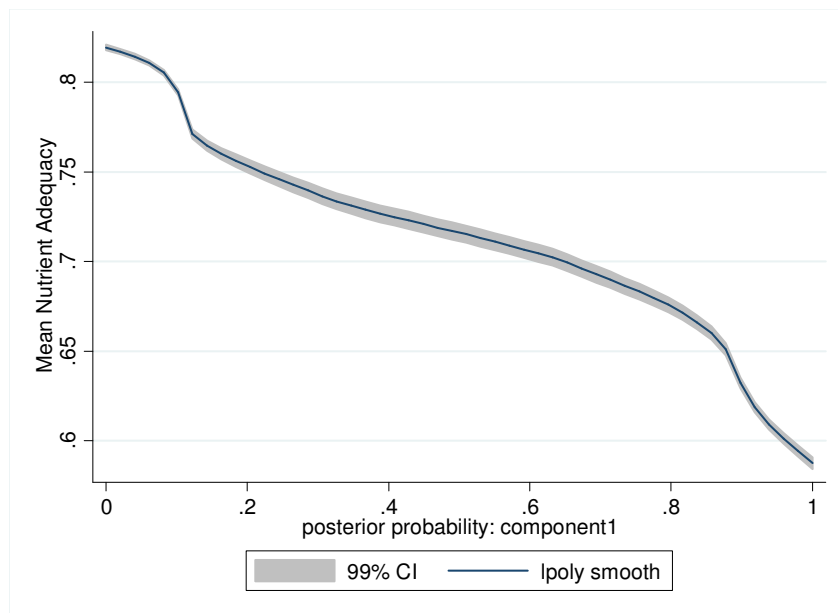


Figure 3.4: Histogram of the count index

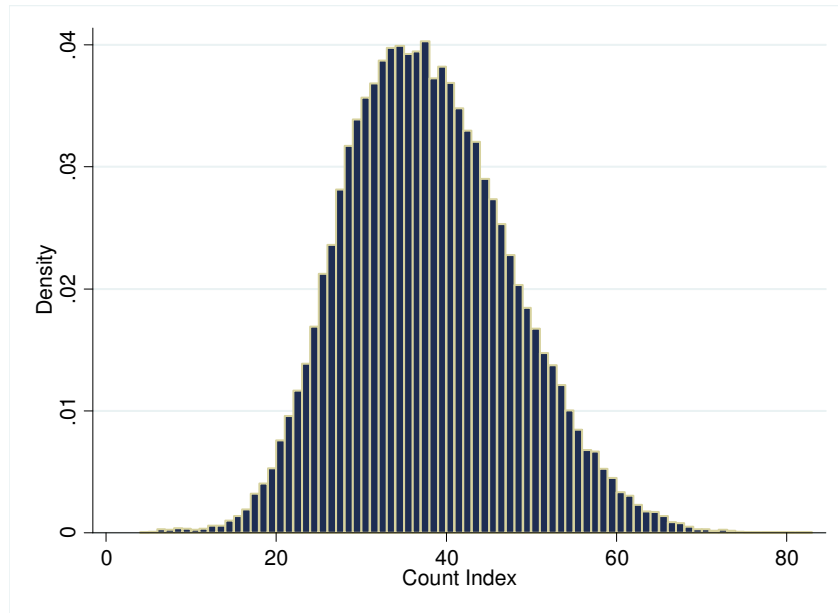
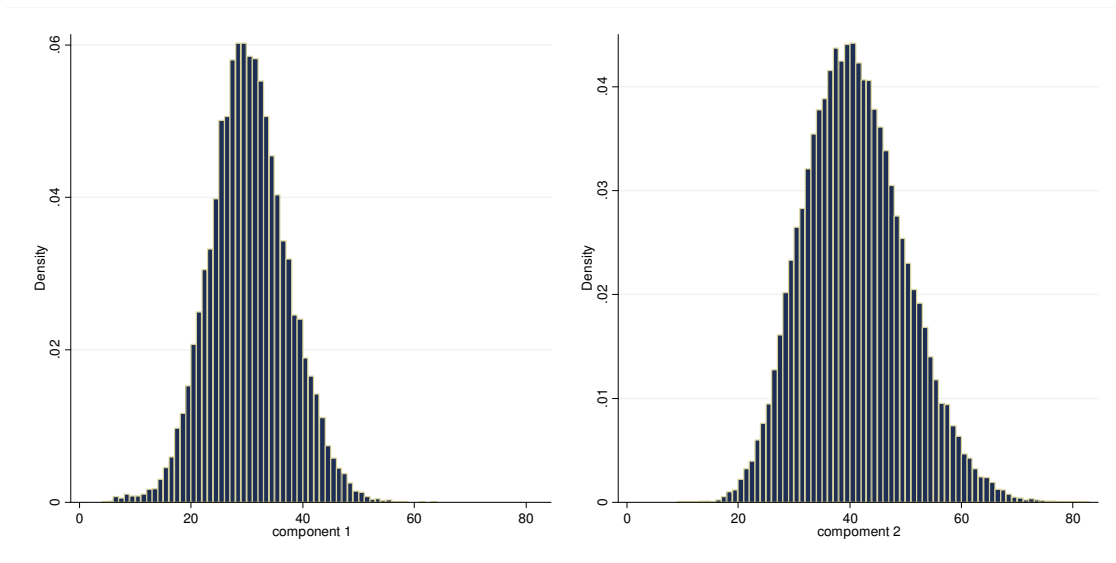


Figure 3.5: Histogram of the count index by component



Chapter 4

A Healthy Food Diversity Index for India

Abstract

Measures of diversity have become popular tools to infer on nutritional adequacy from observed consumption. However the most common measures do not consider that equal distribution of food consumption does not reflect an optimal diet. The proposed index in this paper adjusts the existing concept of the healthy diversity index so that it is applicable for Indian dietary analysis and extends it for analysis of household data. The results show that the modified index is a superior predictor of nutritional adequacy compared to common measures like the berry, entropy or count index.

Key Words: food diversity, nutrition, India

JEL: D12 C43 I19

Introduction

India is with 1.2 billion inhabitants the second most populous country in the world. Despite high growth rates in the last decade undernutrition levels in India remain higher than for most countries of Sub-Saharan Africa, even though those countries are currently much poorer than India (Deaton and Drèze, 2009). Malnourishment does not only persist in terms of calories but also a broad range of micronutrients. According to the National Family Health Survey 2005-06 48% of all under five year old children are chronically malnourished (stunted) and 54 % of all death of under five year old children are related to malnutrition, mainly vitamin A deficiency. While a lot of attention has been paid to calorie consumption, which is without doubt of crucial importance, an adequate calorie intake it does not suffice to maintain health. In fact Deaten and Drèze (2009) find that in some regions of India malnutrition is worse although calorie intakes are higher. Nutritionists seem to agree that the consumption of a variety of foods is a key feature in achieving a healthy diet that can translate into the sufficiency of a variety of nutrients including vitamins and minerals that are required to maintain health.

As the focus is shifting from calorie deprivation to malnutrition the situation demands the emergence of new measures that allow inference on the nutritional status from observed consumption. Calories have shown to be a poor measure of nutrient adequacy. Meanwhile diversity indices have established as popular indicators for food security and nutritional adequacy. A food diversity index describes how food consumption is distributed over different food items or groups. Dietary recommendations often emphasize the importance of a diverse diet as a key to a healthy nutrition since a diverse diet enables the intake of a variety of nutrients. Popular diversity indices like for instance the berry or the entropy index were however not designed for the analysis of diets. They are maximized when food consumption is distributed over all food items or groups in equal shares. While dietary guidelines support a diverse diet they do not refer to equal consumption shares. Optimally a diversity index should react to favorable or unfavorable redistributions that are in line with dietary recommendations. Drescher, Thiele and Mensink (2007) present a promising concept which they labeled healthy food diversity index. They suggest overcoming the weaknesses of current diversity indices by adding a weighting scheme to the Berry index that is derived from national dietary recommendations. While this adjustment of a diversity index is generally sensible the index must however be adjusted to the national eating habits and dietary guidelines of a country.

In this paper an adjustment of the healthy food diversity index (HFD) for the analysis of Indian diets is proposed along with further modifications of the index. Therefore weights are used from food group intake recommendations from the dietary guidelines of National Institute of Nutrition (2010). This method allows a more precise calculation food group weights as well as adjustment to individual characteristics than the original method which deducted from areas of the food pyramid triangle. Then the index is based on a food groups rather than food items. Food groups aggregate items that are similar in terms of their nutritional content. Therefore, if groups are defined sensibly in this way, group based indices are expected to be better suited for inference on the nutritional status. Lastly an adjustment to household data is proposed that considers the needs of its members given the households composition. This allows for the analysis of big data sources like the Consumer Expenditure Survey which has been conducted by the National Sample Survey Office. Such data sources often remain unexploited. However when concerned with food security problems it can be crucial to use representative data sources in order to identify vulnerable regions or households correctly.

For the analysis data from the 68th round of the CES Consumer Expenditure Survey is used along with conversion tables extracted from nutrisurvey and dietary recommendations from the National Institute of Nutrition (2010). The performance of the proposed index in contrast to other diversity indices is evaluated by means of Pearson correlation coefficients with nutrient adequacy ratios and local polynomial regressions on mean nutrient adequacy. The results reveal that the proposed HFD index has indeed a higher correlation with nutrient adequacy ratios and is better suited as a predictor of mean nutrient adequacy than other popular choices like the Berry or the Entropy Index. In the case of India it can be inferred that food group based indices outperform item based indices.

4.1 Developing a Healthy Food Diversity Index for India

Eating a variety of food items as key feature to an optimal diet has become a common recommendation in dietary guidelines. The official dietary guidelines in India are no exception in this case. The proposition that food diversity allows an inference on health implies a link between food diversity and nutrition or dietary quality. Hatloy et al. (1998) show that such an inference is possible. They compare Diet Diversity Scores, which are defined as the number of food groups consumed,

with a nutrient adequacy ratio, which is calculated based on recommendation on macro and micro nutrients. The authors conclude that it is possible to predict nutritional adequacy with dietary diversity scores.

Using diversity indices to analyze dietary patterns represents a possibility for dimensional reduction. The simplest measure of dietary diversity is represented by the counts of consumed food items or groups that capture the number dimension of diversity. The shortcoming of this measure is that it does not consider the distribution of food items in the food consumption bundle. Two indices that pay more attention to the distribution are the Berry Index and the Entropy index. The Berry Index is given by:

$$BI = 1 - \sum_{i=1}^N w_i^2$$

Where w_i is the share of product i in the total amount of food consumed. The index is bounded between 0 and $1 - \frac{1}{N}$. The upper limit of the index approaches 1 as N goes to infinity. If the index takes the value 0 then only one food product has been consumed. $BI = 1 - \frac{1}{N}$ refers to the situation in which equal shares of all available food products were consumed. The Berry index is directly related to the Herfindahl Index, that has been frequently used to measure concentration of an industry, which is represented by the sum of squared market shares $\sum s_i^2$. The Entropy index is derived from information theory and is well known in economics from the inequality literature. The first to use Entropy as a measure of diversity in demand analysis were Theil and Finke (1983) who estimate the demand for diversity for 30 countries. Meanwhile the Entropy index has become a common measure in the analysis of the demand for variety.

$$EI = - \sum_{i=1}^N w_i \log(w_i)$$

Unlike the Berry index the Entropy index is bounded between 0 and $\log(n)$ and takes the value 0 if only 1 food item has been purchased by the household (Weiss and Briglauer, 2002).

All the mentioned indices provide us with different information. The count index only shows us the number of distinct food items consumed and ignores any information on quantity or budget distribution.

Drescher, Thiele and Mensink (2007) pointed out that a diversity index should react to consump-

tion shares of a food item so that it reflects the distribution of consumption over all purchased food items. Hence if a consumer allocates his resources to two food items the index will give us the same value if the consumer decides to purchase equal shares of the food item or if he decides to allocate 99% to the first one 1% to the second one. On the upside the food count index gives us a very clear interpretation and is easy to communicate. The Berry Index is basically an inverse measure of concentration. If we calculate it with expenditure shares on food items in the food budget share then the index tells us how equally the budget is distributed over all purchased food items. It hence captures the distribution of the budget over the food items. The Entropy index has this characteristic as well. However in comparison with the Berry Index the Entropy Index places higher weights on smaller shares. This results in a higher sensitivity of the index to minor commodities. So if a consumer allocates a very small share of his budget to an additional commodity the increase in the Berry index will be smaller than in the Entropy index.

While the discussed indices are adequate tools to capture diversity they are not designed to capture the link between a diversified diet and a consumer's health. While suggesting a diverse diet has become popular in promoting healthy diets official dietary recommendations do not suggest equal consumption shares of different food groups or items for a balanced diet.

One approach to overcome this weakness is suggested by Drescher, Thiele and Mensink (2007) who emphasize that the shortcoming of the Berry index is the equal weighting of food items, which disregards their relevance for health.

They suggest a modification of the Berry Index. This is achieved by creating health values which are a product of recommended food group shares, which the authors label health factors, and the actual consumption shares of a food item. The Healthy Food Diversity Index is then constructed by multiplying the Berry Index with the health values. The approach is applied to a German Consumer Survey using DGE recommendations to calculate health values. A US counterpart is presented by Vadiveloo et al. (2014) who use the 2010 Report of the DGA to construct health factors and calculate a US counterpart to the index.

In order to construct a Healthy Food Diversity Index for Indian household data the index must be adjusted accordingly. The CES provides data on household composition and food consumption but not the individual consumption. As a source of dietary recommendation we use the dietary guidelines for Indians manual from the National Institute of Nutrition (2010) in India which represents the official dietary guidelines. The guidelines suggest dietary intake quantities for 9 different food

groups for 12 different sex-age groups. A summary of the guidelines can be found in table 4.1 of Appendix C. For men and women we assume a moderate working activity since the actual activity level is unobservable. We fully exploit this information as we expect to get a more precise picture of the dietary allowance of the household.

In the first step requires the calculation of the recommended consumption q_{ik}^* for each food group k in household i by summing up the products of the recommended consumption quantity q_{kj}^* of food group k for the sex-age group j with M_j , which represents the number of household members in sex-age group M_j .

$$q_{ik}^* = \sum_{j=1}^J q_{kj}^* * M_j$$

With the recommended food group quantities the health factor of each food group hf_{ik} can be calculated by means of the recommended food group shares.

$$hf_{ik} = \frac{q_{ik}^*}{\sum_{j=1}^J q_{ik}^*}$$

Unlike in previous versions of this index health factors are not a constant value assigned to each food group or sub group but individually calculated for each household based on its composition. This adjustment considers that optimal shares vary with gender and age. The shortcoming of this approach is of course that we can only observe whether the food basket of a household suffices to meet the needs of his members, but not whether the intra household allocation is optimal.

The health value of a household is calculated by summing up all products of the health factors with the actual food group consumption shares w_{ik} .

$$hv_i = \sum_{k=1}^K h_{ik}^* * w_{ik}$$

In order to calculate the food group shares for each household we assigned each food item listed in the survey to a food group according to the food group description from the dietary guidelines. The survey contains information on the food quantities consumed that we used to calculate the shares. These quantities considered were sum of purchased foods and consumed home produced foods. In the last step we combine the Berry Index with the health values.

$$hfd_i = (1 - \sum_{i=1}^N w_i) * hv_i$$

Under this modification neither a high health value nor a high index value alone can lead to high HFD value. A rise of the Berry index can occur due to favorable or unfavorable redistribution from a health perspective. The health value imposes a punishment to unhealthy redistributions and allows it to rise under a healthier distribution. In this index version we preferred to use a food group based berry over an item based berry. Food groups aggregate items which are similar in terms of their nutritional content. An item based index can be high due to the consumption of many food items. If these items are however concentrated in very few food groups this might lead to deficiencies although diversity is high. A group based index can hence be a more robust indicator of a broad nutrient supply. In the context of our analysis this is an important feature. An item based index might react to a broader distribution on staples. Although this behavior will be recognized by the health value the index movement can be contradicted by the movement of the item based berry index while a group based berry will react more sensitive if consumers are shifting away from staples while it is insensitive to redistributions within a food group, which includes distributions over more staples. Although the index could be calculated with the above procedure, it does not capture one aspect of the dietary recommendation adequately. The guidelines were designed with respect to the Indian culture. Hence the general recommendation refers to vegetarians although leaving the possibility to exchange pulses for meat. However the recommendation on the pulses food group, that 30 gram of pulses can be replaced by 50 gram of meat, poses an additional challenge. Since portion sizes differ so will the recommended quantities of this group and consequently the recommended shares. Hence the above explained index requires a modification to take this food group adequately into account and to calculate the recommended quantity shares for all food groups correctly. One possible way to adequately calculate the recommended shares is to orientate on the actual composition of this food group within the household and to assign individual recommended quantity shares that lead as well to individual health factors based on the consumption pattern of the household. We can get the ratio with:

$$r_i = \frac{q_{i,meat}}{q_{i,meat} + q_{i,pulses}}$$

Based on the actual composition we can calculate the recommended quantities:

$$q_{i,pulses\&meat}^* = r_i * q_{i,meat}^* + (1 - r_i) * q_{i,pulses}^*$$

With the recommended quantity in hand we can proceed as described above with the calculation of the health factors. Generally a calibration of the index for a more convenient presentation is possible for instance by dividing all index values by the maximum value. However when comparing index values across time or space a sensible calibration choice will depend on the application which is why it is omitted here. In order evaluate the performance of the index we computed correlation coefficients and local polynomial regressions of different indices with nutrient adequacy ratios, which are discussed in the results section.

4.2 Data

The analysis is based on the Consumer Expenditure Survey (CES) which has been conducted by the National Sample Survey Office (NSSO). The NSSO has been founded in 1950 by the government of India and belongs to the Ministry of Statistics and Programme Implementation. It regularly conducts consumer expenditure surveys nationwide throughout India. For our analysis we use the 68th round of the survey which was conducted in 2012. Over 4 sub rounds 101626 household were interviewed. The survey contains information on consumption expenditure over the last 30 days and provides a high level of detail on food expenditures. In total it provides quantity and expenditure information on 142 food items

Household consumption is assessed using a 30 day recall period. The results of a 30 day recall are prone to a recall error especially for food item that are consumed irregularly. The NSSO tried to reduce recall errors by providing very detailed class descriptions to the surveyed households (NSSO, 2013). The recall period was chosen in line with a preceding study (NSSO, 2002) that tested the suitability of different reference periods to assess household consumption. They compared a 7 day and a 30 day recall period with a one day recall as reference. The findings concerning accuracy are mixed. Estimates for foods that are consumed more frequently such as cereals, milk and sugar appear more favorable with 30 day recalls. However estimates for foods that were more infrequently consumed were considered to be more appropriate with a 7 day recall period. For an estimate of dietary diversity a recall period of more than a week can still be preferable. Drewnowski et al.

(1997) collected detailed food consumption data from individuals on 15 consecutive one day recalls. This procedure allows inference on the development of diversity over time with a minimum recall error. In terms of food items diversity increased over the whole 15 days. However increases were very steep within the first 3 days and only small additions were made between day 10 and day 15. Since the variety curve flattens out after 10 days a shorter period might give a biased picture of a person's eating habits. Ignoring recall errors longer periods can hence be more adequate to capture a households overall variety. With reference to the same study Ruel (2006) concludes that good estimates for diversity can be inferred from a two week recall given an accurate assessment of intake while a 7 day recall might be preferable to minimize memory errors.

In order to remove extreme values such as calorie intakes that are unlikely to be observed in reality observations with less than 500 or more than 5000 calories per capita/day are dropped following Wiesmann et al. (2009). This procedure results in the loss of 2151 observations. The recommended food group quantities for different sex-age groups, which were used to construct the health factors of the healthy food diversity index, are extracted from the official dietary guidelines published by the National Institute of Nutrition (2010). This source additionally provides the dietary allowances for nutrients that are utilized in the calculation of nutrition adequacy ratios. In line with the dietary recommendations the food items from the CES are attributed to 9 distinct food groups, namely cereals and millets, pulses, milk and milk products, roots and tubers, green leafy vegetables, other vegetables, fat and oil and sugar. The conversion tables to convert the food items into nutrients were extracted by means of the software Nutrisurvey which is freely available at Nutrisurvey.de. In order to cover all food items we used Nutrisurveys standard data base as well as the India data base. Our resulting conversion table covered nutrients for 127 food items. Liquors and tobacco products were not considered. In combination with the recommended dietary allowances for Indians from the National Institute of Nutrition (2010) twelve nutrient adequacy ratios and the energy adequacy could be calculated.

4.3 Empirical Strategy

We chose our HFD index as concept of diversity from which we expect that it establishes an improved link between observed food consumption and nutrient supply to infer on the nutritional status. In order to evaluate the performance of the index we compare its relation to nutritional

adequacy to other popular indices. This includes the berry index, the entropy index and the count index. For all indices an item as well as a food group based version is provided with the exception of the count index. In this case there was hardly any notable variation in the number of consumed food groups due to the 30 day recall period. Considering also an item based version of the healthy food diversity index is obligatory since this modification of the index was not considered by previous authors. Performance evaluation is based on correlation coefficients with nutrient adequacy ratios and local polynomial regressions on mean nutritional adequacy. The nutrient adequacy ratios are calculated on a household base as the ratio of the actual households nutrient intakes divided the recommended intakes and are capped at 1 except for the energy intake.

$$NAR_{li} = \frac{N_{li}}{\sum_{j=1}^J RDA_{li}M_j} \quad (4.1)$$

Here N_{li} refers to quantity of nutrient l consumed in household i which is calculated by means of the household consumption quantities according to the consumer expenditure survey and the conversion table from Nutrisurvey. It is divided by the sum of recommended dietary allowances (RDA) of all household members as advised by the National Institute of Nutrition (2010). The recommended dietary allowances of nutrient l for sex-age group j are multiplied by the number of household members in this group following table 4.2 of appendix C. Note that in this case here the categorization of sex-age groups differs from the one used in table 4.1 for balanced diets. For adults a moderate level of work is assumed for the allowance. This procedure considers the actual household composition to calculate the needs of the household members. It is however not possible to control for the intrahousehold allocation of nutrients. Adequacy ratios were calculated for 12 nutrients and energy. The mean adequacy ratio is calculated as the arithmetic mean of all 12 NARs. The cap at one ensures that excess intakes of some nutrients cannot compensate the lag of others.

4.3.1 Correlation with Nutrient Adequacy Ratios

Table 4.3 contains Pearson correlation coefficients between nutrient adequacy ratios and diversity indices. Significance levels were as well calculated though with very few exceptions the coefficients are highly significant. The energy intake has close to zero correlation with diversity across all indices. Hence the calorie intake appears to be unrelated to food diversity. The most decisive value to look at is probably the correlation with the mean nutritional adequacy. Here the group based

HFD index shows the highest correlation with .46 although it closely followed by the group based berry with .42. In general all group based indices are without a single exception by far better correlated with nutrient adequacy than their item based counterparts. This result holds in all cases for all ratios. While it is an ongoing discussion (Drescher et al., 2007) whether group or item based indices should be favored to infer on the nutritional status the results here speak overwhelmingly clear and consistent in favor of group based indices. The item based HFD index that sticks to the original design is still the best performer among the item based indices. This indicates that the applied weighting procedure generally makes sense and leads to an improvement. However in terms of correlation it fails to compete with other group based alternatives like the entropy or the berry index. These show a correlation of .37 and .42 in contrast to .17. Looking at the 12 nutrient adequacy ratios can give an indication of how consistent the index performance is. The group based HFD index exhibits the highest correlation coefficients among all considered indices in 8 out of 12 cases. In the remaining 4 cases the Berry Index shows a slightly higher correlation.

4.3.2 Local Polynomial Regressions

Figure 4.1 and 4.2 show kernel weighted local polynomial regressions of the different indices on mean nutritional adequacy with 99% confidence bands. This yields an indication if a linear relation between nutrient adequacy and diversity exists and how reliable it is. Just as in the correlation analysis the item based indices perform very poorly. In all cases we can make out non linear relations. In addition we have much wider confidence bands than for the case of group based indices. The three graphs for group based indices show that for all three cases a linear relation seems to exist since in each case there is very little curvature. The confidence bands show that there remains some uncertainty for very low index values that is however reduced quickly with increasing values. The visual inspection reveals that the confidence bands of the HFD index are slightly smaller, which indicates that it is a more reliable predictor than its competitors.

4.3.3 Descriptive Statistics

The index itself ranges in the sample from 0.002 to a maximum of 0.177 with a mean of 0.129. Around two thirds of India's population lives in rural areas. Figure 4.3 shows kernel densities estimates separately for rural and urban areas. It reveals that malnutrition is more likely to occur

in rural areas. Rural areas exhibit a higher density at lower index level and a lower density at higher index levels. A great deal of this pattern can certainly be explained by differences in per capita income in rural an urban areas as food diversity tends to increase in income. Figure 4.4 shows the relation between diversity and income. As a proxy for income monthly total expenditure is used from the Consumer Expenditure Survey. In order to get a clear picture average Index values were calculated for each income decile. This graphical representation has been referred to as Diversity Engel Curve by Falkinger and Zweimüller(1996). It shows that average food diversity increases in each income decile while additional benefits from income are marginally decreasing. A reason for this particular shape can be found in the research of Bennet (1941) and Jensen and Miller (2010). At income levels that are close to subsistence, consumers mainly consume staples, as a cheap source of calories in order to survive. As income increases attributes other than calories become more important and consumers tend to shift away from staples towards a more diversified diet.

4.4 Conclusion and Discussion

The index presented in this paper relies on the basic concept of the healthy food diversity index from Drescher et al. (2007) but the concept is modified in several ways. Firstly it is adjusted to Indian dietary recommendations. The weights are here derived from actual intake recommendations rather than food pyramid graph. This procedure is less prone to errors since food pyramids are not designed to yield partial areas that consistently represent intake quantities. Secondly the concept was adjusted for the analysis of household data while considering the individual needs of the household members. This adjustment allows the use of big data sources like the Consumer Expenditure Survey which at the moment often remains unexploited for the analysis of food security and nutrition issues. Thirdly the basic concept was altered to rely on a group rather than an item based index. The results show that this allows a vastly improved inference on nutrient adequacy. The basic efforts of calculating the HFD for India are slightly higher than for the Berry Index. However due to mere consideration of intake recommendations it is as easy to communicate as a Berry Index. Especially since the implied weighting scheme renders positive reactions to unfavorable redistributions (and v.v.) more unlikely. This yields a more straightforward interpretation of the index and does not demand the degree of attention to potential weaknesses of standard indexes like the berry or the entropy index. Food group based indicators are better predictors of nutritional

adequacy than item based indexes. This insight is not a new one although it only been discussed in detail in a handful of papers so that the results in this paper can be considered an important contribution to reaching a consensus. Drescher (2007) points out that the results in this discussion so far are still contradicting. Besides it was also important to show that this is true for the here discussed index as the preceding two authors of this index did not provide such tests to determine an optimal index composition.

While solutions are here presented to infer on nutrition despite the aggregation over households the problem that food consumption cannot be observed in its entirety cannot be solved with this data alone. Future research could look into possibilities to predict intakes outside home which would allow a more robust inference on nutrition from big data sources like the CES.

4.5 Appendix C

Table 4.1: Balanced diet for gender and age groups (number of portions)

Foodgroup	Cereals & Millets	Pulses*	Milk & Milk Products	Roots & Tubers	Green leafy vegetables	Other vegetables	Fruits	Sugar	Fat &Oil
g/portion	30g	30g	100ml	100g	100g	100g	100g	5g	5g
Men	15	3	3	2	1	2	1	6	6
Women	11	2.5	3	2	1	2	1	6	5
Infants	0.5	0.25	4	0.5	0.25	0.25	1	2	4
children 1-3	2	1	5	0.5	0.5	0.5	1	3	5
children 4-6	4	1	5	1	0.5	1	1	4	6
children 7-9	6	2	5	1	1	2	1	6	7
Girls 10-12	8	2	5	1	1	2	1	5	8
Boys 10-12	10	2	5	1	1	2	1	6	7
Girls 13-15	11	2	5	1	1	2	1	5	8
Boys 13-15	14	2.5	5	1.5	1	2	1	4	9
Girls 16-18	11	2.5	5	2	1	2	1	5	7
Boys 16-18	15	3	5	2	1	2	1	6	10

*One portion of pulse may be exchanged with one portion (50 g) of egg/meat/chicken/fish

Source: National Institute of Nutrition (2010), "Dietary Guidelines for Indians: A Manual"

,

Table 4.2: Recommended dietary allowances for indians by gender and age

Group	energy	Protein	Fat	Calcium	Vit. A	Vit B1	Vit. B2
unit	kcal	g	g	mg	ug	mg	mg
man	2730	60	30	600	600	1.4	1.6
woman	2230	55	25	600	600	1.1	1.3
infants	634.8	10.23	9.5	500	175	0.25	0.35
children 1-3	1060	16.7	27	600	400	0.5	0.6
children 4-6	1350	20.1	25	600	400	0.7	0.8
children 7-9	1690	29.5	30	600	600	0.8	1
boys 10-12	2190	39.9	35	800	600	1.1	1.3
girls 10-12	2010	40.4	35	800	600	1	1.2
boys 13-15	2750	54.3	45	800	600	1.4	1.6
girls 13-15	2330	51.9	40	800	600	1.2	1.4
boys 16-17	3020	61.5	50	800	600	1.5	1.8
girls 16-17	2440	55.5	35	800	600	1	1.2

Group	Vit. B6	Vit. C	Vit. B12	Magnesium	Iron	Zinc
unit	mg	mg	ug	mg	mg	mg
man	2	40	1	340	17	12
woman	2	40	1	310	21	10
infants	0.25	25	0.2	37.5	1.932	0
children 1-3	0.9	40	0.6	50	9	5
children 4-6	0.9	40	0.6	70	13	7
children 7-9	1.6	40	0.6	100	16	8
boys 10-12	1.6	40	0.6	120	21	9
girls 10-12	1.6	40	0.6	160	27	9
boys 13-15	2	40	0.6	165	32	11
girls 13-15	2	40	0.6	210	27	11
boys 16-17	2	40	0.6	195	28	12
girls 16-17	2	40	0.6	235	26	12

Source: National Institute of Nutrition (2010) "Dietary Guidelines for Indians: A Manual"

Table 4.3: Pearson correlation coefficients between nutrient adequacy ratios and diversity indices

Nutrients/Index food group/item	Hfd India Group	Hfd Item	Entropy Group	Berry Group	Entropy Item	Berry Item	Count Item
Calories	0.0167***	-0.0477***	-0.0075**	-0.0023	-0.0755***	-0.0722***	0.0180***
Protein	0.1609***	0.0538***	0.1112***	0.1442***	-0.0111***	0.0433***	0.0308***
Fat	0.5647***	0.3060***	0.5421***	0.5489***	0.3161***	0.2869***	0.3066***
Vit. A	0.4649***	0.0876***	0.5002***	0.5129***	0.2362***	0.1361***	0.2793***
Vit. B1	0.2379***	0.1362***	0.1616***	0.2088***	0.0055*	0.1134***	-0.0866***
Vit. B2	0.5051***	0.1892***	0.2724***	0.3434***	-0.0670***	-0.0146***	0.0574***
Vit. B6	0.1696 ***	0.0266***	0.0853***	0.1469***	-0.0828***	0.002	-0.0946***
Folate Acid	0.3248***	0.1327***	0.3782***	0.3966***	0.2252***	0.2329***	0.1624***
Vit. C	0.3144***	0.0132***	0.4739***	0.4900***	0.2770***	0.2280***	0.1738***
Calcium	0.6867***	0.2595***	0.4092***	0.4822***	-0.001	-0.0085***	0.2053***
Magnesium	0.1014***	0.0567***	0.1223***	0.1345***	0.0546***	0.1101***	0.0135***
Iron	0.2641***	0.1670***	0.1647***	0.1909***	0.0238***	0.0838***	-0.0388***
Zinc	0.2099***	0.1306***	0.0698***	0.1093***	-0.0721***	0.0144***	-0.0643***
Mean adequacy	0.4586***	0.1731***	0.3703***	0.4180***	0.0953***	0.1254***	0.1070***

Coefficients were calculated under the consideration of sample weights

Figure 4.1: Kernel-weighted local polynomial regressions with 0.99 CI for group based indicators

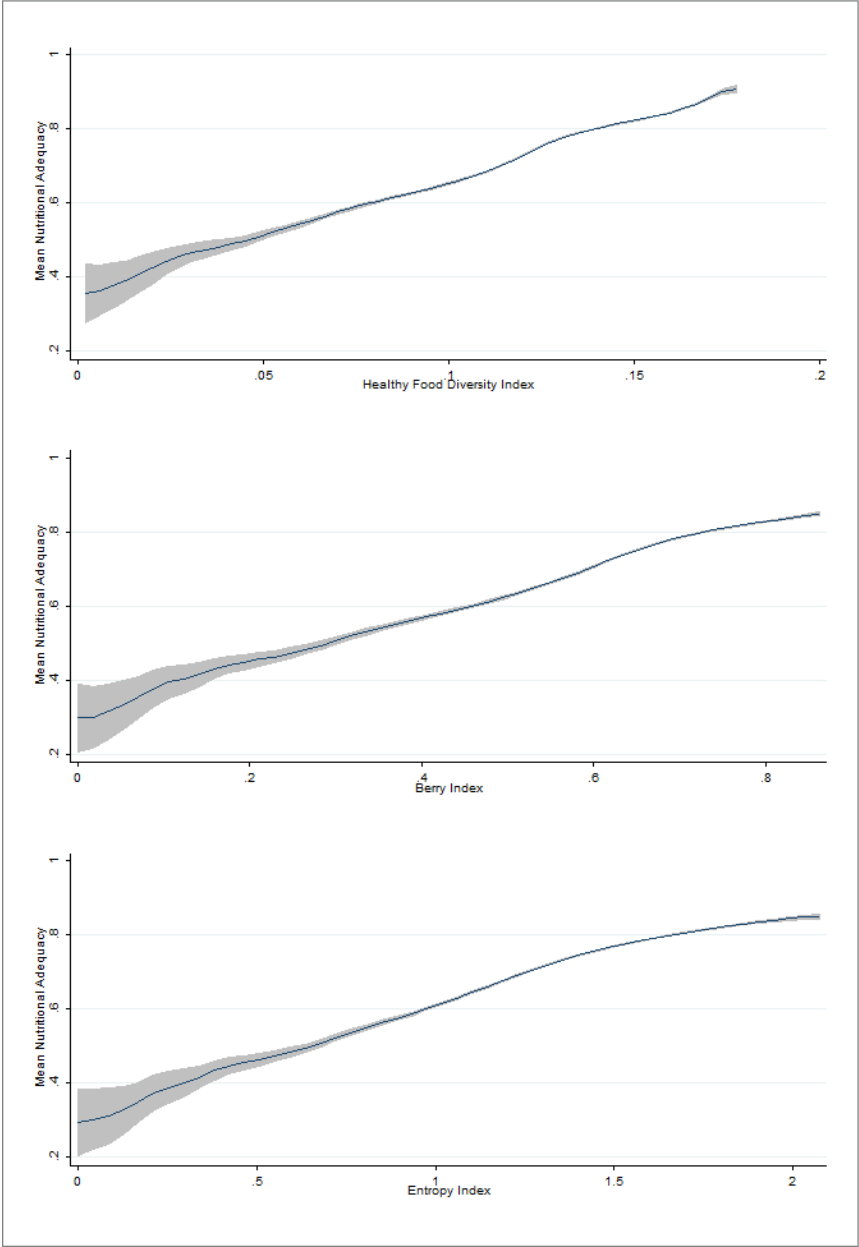


Figure 4.2: Kernel-weighted local polynomial regressions with 0.99 CI for item based indicators

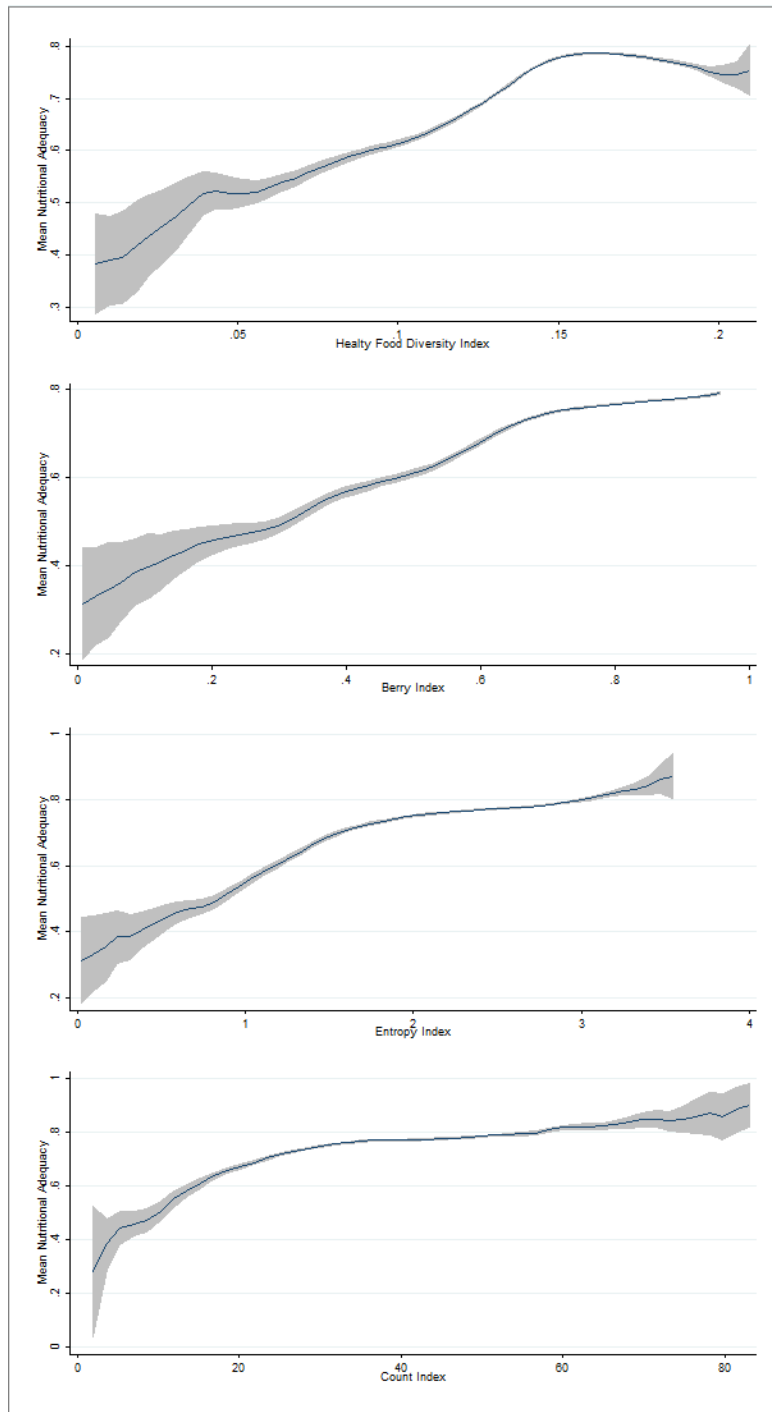


Figure 4.3: Income and food diversity

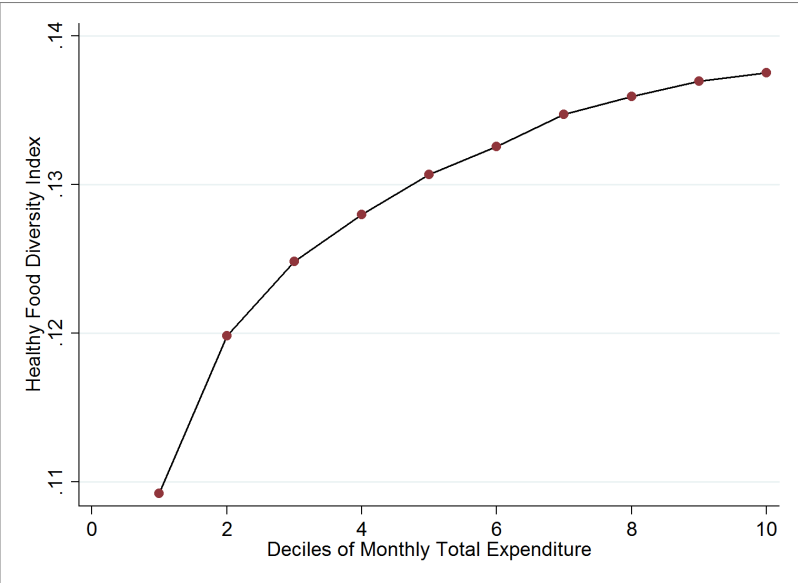
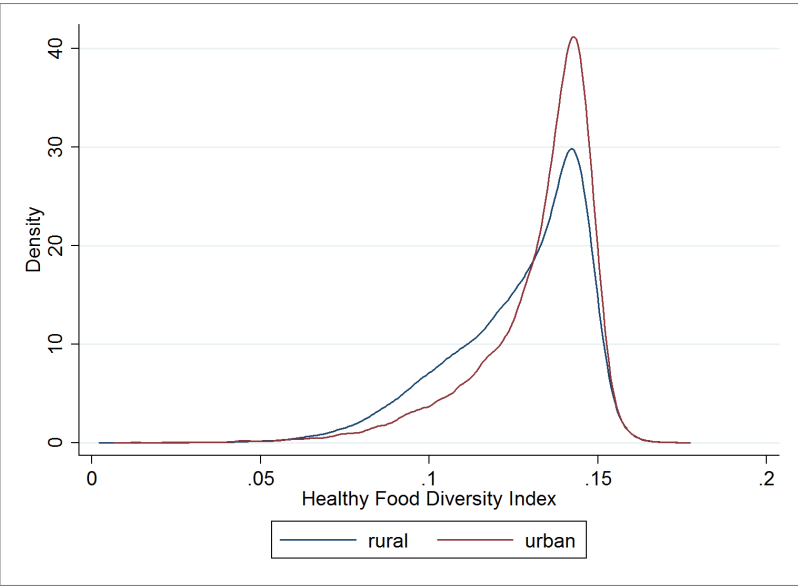


Figure 4.4: Kernel density estimates for rural and urban areas



Chapter 5

Conclusions

This final chapter briefly summarizes the results and provides some discussion. The three essays were concerned with different topics related to poverty and nutrition in developing countries.

The first essay proposes a simple method to evaluate transaction costs in agricultural exchange while considering the heterogeneous nature of agricultural goods that lead to differences in quality and hence value. The proposed stochastic frontier framework is able to estimate the magnitude of transaction costs and shed light on the determinants. Results indicate a magnitude of 12-18% for maize transactions from small scale farmers in rural Kenya and identify drive time, market distance, education and counterparts in negotiations as main determinants. These results suggests that conditions for small scale farmers in Kenya could be improved with better education and an enhanced infrastructure.

The essay also attempts to shed light on the empirical challenges of transaction costs analysis. Due to the mostly unobservable nature of transaction costs proxies need to be carefully selected in line with a basic concept of transaction costs. Kähkönen and Leathers (1999) provide a a very critical discussion which also shows how conflicted the transaction costs literature still is with respect to empirical application and that a consensus on how to measure it has yet to be reached.

Further the analysis of transaction costs in agricultural exchange requires very specific details on products and transactions that very few data sets provide so far. Additional data sources could aid policymakers to better identify bottlenecks and set investment priorities.

The second essay we estimated a latent demand model for food diversity based on the assumptions

that consumers exhibit different demand behaviors which depend on their unobservable nutritional status. The class with subsistence concerns has a lower income elasticity of food diversity as calories are more important than a diversified diet. The classes differ widely in terms of nutrient intake while the intake of many nutrients of the poorer class, which represents 29% of the population, is critically low for many essential nutrients. The regression on class membership probabilities show that besides income, education, location, wealth and household composition play a significant role. The basic motivation of Jensen and Miller (2012) to propose an alternative poverty index was the problem that whether an individual has achieved calorie sufficiency or overcome hunger cannot be adequately determined by a calorie norm. Hence indicators that aim to track the nutritional status of a consumer that he reveals in his consumption behavior can be a more reliable predictor of hunger and undernutrition. While the analysis in this essay is concerned with food diversity and thus with malnutrition rather than undernutrition it becomes clear that the same problems applies to tracking malnutrition. This insight is especially important in the light of recent events such as the Agenda2030 which shifted development goals beyond the mere elimination of hunger towards the improvement of food security and nutrition. In pursuit of these goals measured successes and related decisions in policy will be sensitive to the applied methods.

The third essay presents a modified version of the healthy food diversity index for India that is based on the original concept of Drescher, Thiele and Mensink (2007). The results show that the modified HFD index is a superior predictor of nutritional adequacy compared to common measures like the berry, entropy or count index and hence represents an interesting alternative to existing indices. The household adjustment of the index also offers interesting possibilities as it allows to analyze big data sources such as the NSS Consumer Expenditure Survey. Another interesting feature of the analysis is that sheds light on the performance of common diversity indices. For instance so far there is no consensus whether item based or food group based indices are better suited to infer on nutrition. However in this study group based indices perform without exception better than item based indices.

Bibliography

- Aigner, D., C.A.K. Lovell, and P. Schmidt (1977) "*Formulation and Estimation of Stochastic Frontier Production Function Models*," *Journal of Econometrics* 6, 21-37.
- Alene A. (2008) "*Market Participation under Transaction Costs: Maize Supply and Fertilizer Demand in Kenya*," *Food Policy* Vol.33(4):318-328.
- Arrow, K.J. (1969) "*The Organization of Economic Activity: Issues Pertinent to the Choice of Market versus Non-Market Allocation*," In *The Analysis and Evaluation of Public Expenditures: The PBB-System*, Joint Economic Committee, 91st Cong., 1st sess., vol. 1. Washington, D.C.: Government Printing Office.
- Ayyagari, P., P. Deb, J. Fletcher, W. Gallo, and S. L. Sindelar (2013) "*Understanding Heterogeneity in Price Elasticities in the Demand for Alcohol for older Individuals*" , *Health Econ.* 2013 January, 22(1) 98-105
- Barrett, C.B. (2008) "*Market Participation: Concepts and Evidence from Eastern and Southern Africa*," *Food Policy* 33:299–317.
- Battese, G.E.and T.J. Coelli (1993) "*A Stochastic Frontier Production Function Incorporating a Model for Technical Inefficiency Effects*," *Working Papers in Econometrics and Applied Statistics* 69/1993. Department of Econometrics, Univ. of New England, Armidale, Australia.
- Battese, G.E. and T.J. Coelli (1995) "*A Model for Technical Inefficiency Effects in a Stochastic Frontier Production Function for Panel Data*,"*Empirical Economics* 20, pp. 325-32.
- Benassy, J. (1996) "*Taste for Variety and Optimum Production Patterns in Monopolistic Competition*", *Economics Letters*, 52, 41-47

- Bennett, M. K. (1941) *"Wheat in National Diets"*, Wheat Studies of the Food Research Institute, Stanford University, California
- Coase, R.H. (1937) *"The Nature of the Firm,"* *Economica* 4 (16): 386–405.
- Cuevas, A.C. (2014) *"Transaction Costs of Exchange in Agriculture: A Survey,"* *Asian Journal of Agriculture and Development* Volume 11, Issue 1, Pages 21-38
- Deaton, A. (1988) *"Quality, Quantity and Spatial Variation of Price."* *American Economic Review.* 78(3):418-430.
- Deaton, A. and J. Drèze (2009) *"Food and Nutrition in India: Facts and Interpretations"* , *Economic & Political Weekly*, Vol XLIV No 7
- De Janvry, A. and E. Sadoulet (1998) *" Integration into Markets: Determinants of Entry and Supply Response,"* Paper presented at the Tercer Simposio Latinoamericano sobre Investigación y Extension en Sistemas Agropecuarios (IESA.AL III), Nuevos Enfoques para la Superación de la Pobreza Rural y para el Desarrollo de las Capacidades Locales., Lima, Peru.
- Deb, P. and P. K. Trivedi (2002) *"The Structure of Demand for Health Care: Latent Class versus Two-Part Models"* , *Journal of Health Economics* 21 601-625
- Dpenbusch, L. (2017) *Gender Price Gaps in Central Kenyan Vegetable Markets"*, Global Food Discussion Paper 108, University of Göttingen
- Dixit, A. and J. E. Stiglitz (1977) *"Monopolistic Competition and Optimum Product Diversity"*, *American Economic Review* 67, 292-308
- Drescher, L.S., S. Thiele and G.B. Mensink (2007) *"A New Index to Measure Healthy Food Diversity Better Reflects a Healthy Diet Than Traditional Measures"* , *Journal of Nutrition* 137:647-651
- Drescher, L.S. (2007) *"Healthy Food Diversity: Measurement, Determinants of Consumer Demand and Willingness to Pay"*, Cuvillier Verlag
- Drescher, L.S., S. Thiele, J. Rosen and G.B. Mensink (2009) *"Consumer Demand for Healthy Eating Considering Diversity- An Economic Approach for German Individuals"*, *International Journal of Consumer Studies* ISSN 147-6423

- Drewnowsky, A. , S. A. Henderson, A. Driscoll and B. J. Rolls (1997) "*The Dietary Variety Score: Assessing Diet Quality in Healthy Young and Older Adults*" , Journal of the American Dietetic Association Volume 97 Number 3
- Dorosh P., H. Wang, L. You and E. Schmidt (2010) "*Crop Production and Road Connectivity in Sub-Saharan Africa A Spatial Analysis,*" World Bank PRWP 5385
- Escobal, J.A. and D. Cavero (2012) "*Transaction Costs, Institutional Arrangements and Inequality Outcomes: Potato Marketing by Small Producers in Rural Peru,*" World Development, Volume 40, Issue 2, February 2012, Pages 329-341, ISSN 0305-750X
- Fafchamps, M. (2004) "*Market Institutions in Sub-Saharan Africa: Theory and Evidence,*" Cambridge, MA: MIT Press
- Falkinger, J. and J. Zweimueller (1996) "*The Cross-Country Engel Curve for Product Diversification*", Structural Change and Economic Dynamics
- FAO (1996) "*World Food Summit Plan of Action: Food and Agricultural Organization of the United Nations*", Documentation of World Food Summit 13-17 November 1996 Rome
- FAO (2010) "*Concept Note*", International Symposium on Food and Nutrition Security: Food-Based Approaches for Improving Diets and Raising Levels of Nutrition, FAO, Rome, 7–9 December 2010. Rome.
- FAO (2013) "*The State of Food and Agriculture*", Food and Agriculture Organization of The United Nations, Rome 2013
- Ferrari, S. L. P. and F. Cribari-Neto (2004) "*Beta Regression for Modelling Rates and Proportions*", Journal of Applied Statistics 31(7): 799-815
- Foote, J. A., S. P. Murphy, L. R. Wilkens, P. Basiotis and A. Carlson (2004) "*Dietary Variety Increases the Probability of Nutrient Adequacy among Adults*", American Society for Nutritional Sciences
- Furubotn, F. and R. Richter (2005) "*Institutions and Economic Theory: The Contribution of the New Institutional Economics,*" Michigan: University of Michigan Press

- Greene, W. (2009) "*A Stochastic Frontier Model with Correction for Sample Selection*" Journal of Productivity Analysis. doi:10.1007/s11123-009-0159-1
- Gronau, R. and D. Hamermesh (2008) "*The Demand for Variety: A Household Production Perspective*", The Review of Economics and Statistics August 2008, 90(3): 562-572
- Government of India (2015) "*Millennium Development Goals Report 2015*"
- Hatloy, A., L. E. Torheim and A. Oshaug (1998) "*Food Variety - a Good Indicator of Nutritional Adequacy of the Diet: A Case study from an Urban Area in Mali, West Africa*", European Journal of Clinical Nutrition 52, 891-898
- Headey, D and O. Ecker (2013) "*Rethinking the Measurement of Food Security: From the First Principle to the Best Practice*", Food Security (2013) 5:327-343
- Huang, C.J. and J.T. Liu (1994) "*Estimation of a Non-neutral Stochastic Frontier Production Function*," Journal of Productivity Analysis 5, pp. 171-80.
- International Institute for Population Sciences (IIPS) and Macro International (2007) "*National Family Health Survey (NFHS-3), 2005-06: India: Volume I*", Mumbai: IIPS
- Jackson, L.F. (1984) "*Hierarchic Demand and the Engels Curve for Variety*", The Review of Economics and Statistics, Vol.66, No 1 (Feb.1984) pp.8-15
- Jensen, R. T., and N.H. Miller (2010) "*A Revealed Preference Approach to Measuring Hunger and Undernutrition*", NBER working paper no. 16555, Cambridge, MA
- Jensen, R. T. and N. H. Miller (2011) "*Do Consumer Prices Subsidies Really Improve Nutrition?*", The Review of Economics and Statistics, November 2011, 93(4) 1205-1223
- Key, N., E. Sadoulet and A. de Janvry (2000) "*Transactions Costs and Agricultural Household Supply Response*" American Journal of Agricultural Economics 82 (May 2000): 245-259
- Kenya National Bureau of Statistics (KNBS) (2006) "*Kenya Integrated Household Budget Survey*" (KIHBS) 2005/06 (Revised Edition) Basic Report Ministry of Planning and National Development, Government of Kenya. Nairobi: 300 pp)

- Kumbhakar, S. C., S. Ghosh, and J. T. McGuckin (1991) "A Generalized Production Frontier Approach for Estimating Determinants of Inefficiency in U.S. Dairy Farms," *Journal of Business and Economic Statistics* 9, pp. 279-86
- Lancaster, K. (1990) "The Economics of Product Variety: A Survey", *Marketing Science* 9, 189-206
- Lee, J. and M. G. Brown (1989) "Consumer Demand for Food Diversity", *Southern Journal of Agricultural Economics* Volume 21, Number 02, December 1989 pp47-52
- Li Z., X. Yu, Y. Zeng and R. Holst (2012) "Estimating Transport Costs and Trade Barriers in China: Direct Evidence from Chinese Agricultural Traders," *China Economic Review*, Vol.23(4):1003-1010.
- Lovon, M. and A. Mathiassen (2014) "Are the World Food Programme's Food Consumption Groups a Good Proxy for Energy Deficiency?," *Food Security* (2014) 6:461-470
- Scott E. Masten, S.E. (2000) "Transaction-Cost Economics and the Organization of Agricultural Transactions", *Industrial Organization*, Volume 9, pages 173-195
- Matsumoto, T., and T. Yamano (2009) "Soil Fertility, Fertilizer, and the Maize Green Revolution in East Africa", *World Bank Policy Research Paper*. Washington, World Bank
- McLachlan, G. J. and D. Peel (2000) "Finite Mixture Models", New York: Wiley
- Meyer, S. and X. Yu (2013) "The Impacts of Production Uncertainties on World Food Prices", *China Agricultural Economic Review*. Vol.5(4):459-472
- Mkenda, B.K. and B.van Campenhoutm (2011) "Estimating Transaction Costs in Tanzanian Supply Chains," *IGC Working Paper* 11/0898
- Ministry of Health and Family Welfare, Government of India (2006) "National Family Health Survey (NFHS-3) India 2005-06", International Institute for Population Sciences Deonar, Mumbai
- Moon, W., W. J. Florkowski, L. R. Beuchat, A. V. Resurreccion, P. Paraskova, J. Jordanov and M. S. Chinnan (2002) "Demand for Food Variety in an Emerging Market Economy", *Applied Economics* 34: 573-581

National Institute of Nutrition (2010) *"Dietary Guidelines for Indians: A Manual"*, Hyderabad - 500 007, India

National Institute of Nutrition (2009) *"Nutrient Requirement and Recommended Dietary Allowances for Indians"*, Hyderabad - 500 007, India

National Sample Survey Organization Ministry of Statistics and Program Implementation Government of India (2002) *"Key Indicators of Household Consumer Expenditure in India"*,

National Sample Survey Organization Ministry of Statistics and Program Implementation Government of India (2002) *"Results of a Pilot Survey on Suitability of Different Reference Periods for Measuring Household Consumption"*,

Ndulu B.J. (2006) *"Infrastructure, Regional Integration and Growth in Sub-Saharan Africa: Dealing with the disadvantages of Geography and Sovereign Fragmentation"*, Journal of African Economies, Volume 15, AERC Supplement 2, PP. 212–244

Niaz,A. ,M.Yaseen, M.Arshad and R.Ahmad (2015) *"Response of Maize Yield, Quality and Nitrogen Use Efficiency Indices To Different Rates and Application Timings,"* The Journal of Animal and Plant Sciences, 25(4): 2015, Page:1022-1031 ISSN: 1018-7081.

North,D.C. (1991) *"Institutions,"* Journal of Economic Perspectives Volume 5, Number 1 Winter 1991 Pages 97-1

Olwande, J., M. Smale, M.K. Mathenge, F. Place, D. Mithöfer (2015). *"Agricultural Marketing by Smallholders in Kenya: A Comparison of Maize, Kale and Dairy,"* Food Policy Vol.52: 22–32

Omamo S.W. (1998) *"Transport Costs and Smallholder Cropping Choices: An Application to Siaya District, Kenya,"* American Journal of Agricultural Economics, Vol. 80(1): 116–123

Osebeyo, S.O. and G.C. Aye (2014) *"Transaction Costs and Marketing Decision: a Case Study of Smallholder Tomato Farmers in Makurdi, Nigeria,"* Urban Planning and Transport Research, Volume 2 2014-Issue 1 Pages 333-340

Rapsomanikis, G (2015) *"The Economic Lives of Smallholder Farmers: An Analysis Based on Household Data from Nine Countries,"* Food and Agriculture Organization of the United Nations Rome, 2015

- Renkow, M., D. Hallstrom, and D. Karanja. (2004) "*Rural infrastructure, Transaction Costs and Market Participation in Kenya*". *Journal of Development Economics* 73 (1): 349-367
- Ruel, M. (2006) "*Operationalizing Dietary Diversity: A Review of Measurement Issues and Research Priorities*", *The Journal of Nutrition*
- Schmidt, P. (2011) "*One-Step and Two-Step Estimation in SFA Models*," *Journal of Productivity Analysis*.
- Shonkwiler, J. S., J. Lee and T. G. Taylor (1987) "*An Empirical Model of the Demand for Varied Diet*", *Applied Economics*, 19:10, 1403-1410
- Smithson, M. and J. Verkuilen (2006) "*A Better Lemon Squeezer? Maximum Likelihood Regression with Beta-Distributed Dependent Variables*", *Psychological Methods* 11(1): 54-71
- Subramanian, S., and A. Deaton (1996) "*The Demand for Food and Calories*", *Journal of Political Economy* 104 (1): 133-162.
- The World Bank (2008) "*Agriculture for Development*", The World Bank, Washington DC
- The World Bank (2016) "*Taking on Inequality*", The World Bank, Washington DC
- Thiele, S and C. Weiss (2002) "*Consumer Demand for Food Diversity: Evidence for Germany*", *Food Policy* 28 (2003) 99-115
- Theil, H. and R. Finke (1983) "*The Consumer's Demand for Diversity*", *European Economic Review* 23 (1983) 395-400 North Holland
- Thompson, B. and L. Amoroso (2014) "*Improving Diets and Nutrition - Food-Based Approaches*", The Food and Agriculture Organization of the United Nations and CABI
- Thorne-Lyman, A.L., N. Valpiani, K. Sun, R. D. Semba, C. L. Klotz, K. Kraemer, N. Akhter, S. de Pee, R. Moench-Pfanner, M. Sari and M.W. Bloem (2010) "*Household Dietary Diversity and Food Expenditures are Closely Linked in Rural Bangladesh Increasing the Risk of Malnutrition due to the Financial Crisis*", *J Nutr* 140(1): 182S-8S
- United Nations (2015) "*Millennium Development Goals Report 2015*"
- United Nations (2018) "*The Sustainable Development Goals Report 2018*"

- Vadiveloo, M. , L.B. Dixon , T. Mijanovich, , B. Elbel and N. Parekh (2014) ” *Development and Evaluation of the US Healthy Food Diversity Index* ”,British Journal of Nutrition
- Vakis, R., E. Sadoulet and A. de Janvry (2003) ” *Measuring Transaction Costs from Observed Behaviour: Market Choices in Peru*”, CUDARE Working Paper Series, Department of Agricultural and Resource Economics, University of California, Berkeley.
- Wang, H. (2002) ” *Heteroscedasticity and Non-Monotonicity Efficiency Effects of a Stochastic Frontier Model*”,Journal of Productivity Analysis 18: 241-253.
- Wang, H., and P. Schmidt (2002) ” *One-Step and Two-Step Estimation of the Effects of Exogenous Variables on Technical Efficiency Levels*”,Journal of Productivity Analysis 18,pp.129-144.
- Weiss, C. and W. Briglauer (2002) ” *Determinants and Dynamics of Farm Diversification* ”, Paper prepared for the Xth EAAE Congress 2002
- WHO (2006) ” *Obesity and Overweight. Fact Sheet No. 311,*”
- Wiesmann, D., L. Bassett, T. Benson and J. Hoddinott (2009) ” *Validation of the World Food Programme’s Food Consumption Score and Alternative Indicators of Household Food Security*”,IFPRI Discussion Paper 00870
- Wollni, M and B. Brümmer ” *Productive efficiency of specialty and conventional Coffee Farmers in Costa Rica: Accounting for technological Heterogeneity and Self-Selection,*” Food Policy 37 (2012) 67-76
- Yamano, T. and A. Arai (2010) ” *The Maize Farm-Market Price Spread in Kenya and Uganda*”, National Graduate Institute for Policy Studies
- Yamano, T, K. Otsuka, F.Place, Y.Kijima and J.Nyoro (2005) ” *The REPEAT Survey in Kenya (First Wave):Results,*” National Graduate Institute for Policy Studies (GRIPS)
- Yu, X. (2014) ” *Raising Food Prices and Welfare Change: A Simple Calibration*” Forthcoming in Applied Economics Letters.
- Yu, X. and D.Abler (2009) ” *The Demand for Food Quality in Rural China,*” American Journal of Agricultural Economics, Vol. 91(1): 57-69.

Zhou, D and X. Yu (2014) " *Calorie Elasticities with Income Dynamics: Evidence from the Literature*", Global Food Discussion Paper N0.35

Declarations:

1.) I, hereby, declare that this Ph.D. dissertation has not been presented to any other examining body either in its present or a similar form. Furthermore, I also affirm that I have not applied for a Ph.D. at any other higher school of education.

Göttingen, 11.5.2017

.....

Christoph Steffen

2.) I, hereby, solemnly declare that this dissertation was undertaken independently and without any unauthorized aid.

Göttingen, 11.05.2017

.....

Christoph Steffen