UNIVERSITÉ DU QUÉBEC À MONTRÉAL

PRIX ET COÛTS AU COMMERCE DANS LES PAYS AFRICAINS

THÈSE

PRÉSENTÉE

COMME EXIGENCE PARTIELLE

DU DOCTORAT EN ÉCONOMIQUE

PAR

EWANE WILLIAM THEOPHILE

JUIN 2021

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

PRICES AND TRADE COSTS IN AFRICAN COUNTRIES

THESIS

PRESENTED

AS PARTIAL REQUIREMENT

TO THE PH.D IN ECONOMICS

BY

EWANE WILLIAM THEOPHILE

JUNE 2021

REMERCIEMENTS

Cette thèse a été un parcours long et difficile. J'ai pu terminer ce processus grâce à l'apport de plusieurs personnes. D'abord, mon directeur de thèse Julien Fréderic Martin qui a été exceptionnel tout au long de ma thèse. Il m'a tenu la main de bout en bout et m'a initié aux canons de la recherche. Je lui en suis infiniment reconnaissant. Je remercie également tous les professeurs du département d'économie, en particulier Kristian Behrens, Sophie Osotimehin, Florian Mayneris. Merci également à Mathieu Couttenier qui a accepté d'être l'évaluateur externe de cette thèse. Je voudrais également remercier l'équipe du sécretariat du département (Martine , Karine , Julie). Je ne saurai teminer sans remercier tous les collègues et amis du departement. En particulier Théophile Bougna pour son accompagnement et ses conseils. Enfin je remercie ma famille au Cameroun qui n'a cessé de me pousser vers l'avant. Je pense à Marie-Madeleine, Fany , Eugenie et Cynthia. Je vous aime.

TABLE DES MATIÈRES

LISTE DES TABLEAUX

Table Page

TABLE DES FIGURES

RÉSUMÉ

Cette thèse part de trois faits importants qui caractérisent la géographie de l'activité économique dans les pays africains. Il s'agit respectivement de : l'existence des coûts au commerce interne ; la dispersion spatiale des prix des biens et services ; et l'inégalité spatiale du coût de la vie.
La littérature économique qui s'est intéressée à ces faits nous donne l'opportunité de poser trois questions : (i) la méthode d'estimation des coûts au commerce interne à partir des différences spatiales de prix est-elle valide ? (ii) quelle est la taille des coûts au commerce interne au Rwanda et quels sont les effets d'une meilleure connectivité routière sur ces coûts ? (iii) qu'est-ce qui détermine la différence spatiale du coût de la vie en Éthiopie et quelles en sont les implications ?
Pour répondre à ces questions, cette thèse procède en trois chapitres. Le premier chapitre fait un exercice de validation de la méthode d'inférence des coûts au commerce interne à partir de la différence spatiale des prix. Une fois cette méthode validée, le deuxième chapitre l'utilise pour évaluer la taille des coûts au commerce au Rwanda, et analyser les effets sur ces coûts des infrastructures routières. Enrichi des leçons tirées des chapitres précédents, le troisième chapitre évalue les sources et les implications de la différence spatiale du coût de la vie en Éthiopie.

Le premier chapitre intitulé *"The measurement of trade costs in Ethiopia"* compare les coûts au commerce interne estimés et réels en Éthiopie. L'estimation des coûts au commerce interne est basée sur la méthode proposée par Atkin et Donaldson (2015). Cette méthode consiste à inférer les coûts au commerce interne à partir de la différence spatiale des prix corrigée de la variation spatiale des marges chargées par les vendeurs. Les coûts au commerce interne réels sont mesurés à l'aide des coûts encourus par le Programme Alimentaire Mondial (PAM) au cours de la distribution de l'aide alimentaire en Éthiopie. En utilisant la différence spatiale des prix des produits non alimentaires pour estimer les coûts au commerce interne, les résultats montrent que l'élasticité

estimée des coûts au commerce par rapport à la distance est trois fois plus faible que l'élasticité des coûts réels. Ce résultat s'explique par la différence de types de produits et de modes de transport utilisés entre les vendeurs et le PAM. En considérant la dispersion spatiale des prix des produits alimentaires pour estimer les coûts au commerce interne, les résultats indiquent une grande similarité entre les élasticités estimées et observées des coûts au commerce interne par rapport à la distance. Ce chapitre montre ainsi qu'utiliser la différence spatiale des prix corrigée des marges est une bonne méthode pour approximer les coûts au commerce interne. La contribution majeure de ce chapitre est qu'il permet de comparer des mesures directes et indirectes des coûts au commerce interne. Dans la littérature sur les coûts au commerce interne, qui est assez récente, cet exercice n'avait pas encore été effectué. La méthode de Atkin et Donaldson (2015) étant validée, elle peut être appliquée dans le deuxième chapitre de cette thèse.

Le second chapitre intitulé *"Trade costs, prices and connectivity in Rwanda"* mesure les coûts au commerce interne au Rwanda et évalue les effets d'une meilleure connectivité routière sur ces coûts. Les coûts au commerce interne sont mesurés à l'aide de la méthode proposée par Atkin et Donaldson (2015). Cette méthode consiste à inférer les coûts au commerce interne à partir de la différence spatiale des prix corrigée de la variation spatiale des marges chargées par les vendeurs. En ce qui concerne la connectivité routière, nous définissons trois niveaux de connectivité routière en fonction de la proximité de marchés avec des routes secondaires ou nationales. Ces trois niveaux de connectivité routière sont : faible connectivité, connectivité intermédiaire et forte connectivité. Trois principaux résultats émergent de notre analyse.
D'abord, nos estimations suggèrent que les coûts au commerce sont larges au Rwanda. En effet, définis comme fonction de la distance, les coûts au commerce au Rwanda sont 10 fois plus élevés qu'aux États-Unis, 3 fois plus élevés qu'en Éthiopie et deux fois plus élevés qu'au Nigéria. Le pouvoir de marché élevé des compagnies de transport, la faible qualité des infrastructures et le relatif enclavement du Rwanda expliquent ce résultat. Ensuite, nous démontrons que les coûts au commerce sont relativement plus élevés pour les produits manufacturés par rapport aux produits non manufacturés. La différence dans les modes de transport utilisés constitue une explication possible de ce

résultat. Enfin, s'agissant du rôle des infrastructures routières, nos estimations montrent qu'une meilleure connectivité routière des marchés se traduit par une réduction des coûts au commerce. En particulier, les coûts au commerce sont faibles lorsque les marchés sont connectés à la fois aux nouvelles routes secondaires construites par le gouvernement rwandais et aux routes nationales existantes. La connexion aux nouvelles routes secondaires est pertinente pour la réduction des coûts commerce, lorsque celles-ci permettent l'accès aux routes nationales. En somme, ce chapitre documente la littérature sur la taille des coûts au commerce interne dans les pays africains. Il complémente également la littérature sur les effets économiques des infrastructures de transport en mettant l'accent sur les effets de la connectivité en milieu rural.

Le dernier chapitre intitulé *"The price of remoteness : product availability and local cost of living in Ethiopia"* examine les déterminants et les implications de la différence spatiale du coût de la vie à travers les villes en Éthiopie. L'analyse est d'abord faite d'un point de vue désagrégé en considérant la dispersion spatiale des deux composantes du coût de la vie : les prix et la disponibilité des produits. Les résultats indiquent que les villes éloignées font face à des prix élevés et ont accès à une gamme restreinte de produits. De plus, les prix sont également élevés dans les grandes villes, mais ces dernières ont accès à plus de variétés. Armé de ces résultats, le chapitre analyse ensuite d'un point de vue agrégé la dispersion spatiale des indices du coût de la vie. Ces indices agrégés du coût de la vie prennent en compte à la fois les prix et la disponibilité des produits. Les résultats montrent que le coût de la vie est élevé dans les zones reculées, mais qu'il n'est pas systématiquement lié à la taille des villes. Le chapitre s'intéresse enfin à l'impact de la dispersion spatiale du coût de la vie sur les migrations. Les résultats montrent que les migrants quittent les villes où le coût de la vie est élevé pour s'installer dans les villes où il est plus faible. Ce chapitre contribue de plusieurs manières à la littérature. En premier lieu, il construit des indices de coûts de la vie prenant en compte à la fois le prix et la disponibilité des produits alors que les indices existants se concentrent sur les prix. En second lieu, il se départit de la distinction entre l'urbain et le rural qui prédomine dans la littérature, en analysant la dispersion spatiale du coût de la vie entre les villes. Enfin, il documente la littérature sur les déterminants des flux

migratoires internes dans un contexte de pays en développement.

Mots-clés : Coûts au commerce interne ; Coût de la vie ; Dispersion des prix ; Disponibilité des produits ; Distance ; Eloignement ; Ethiopie ; Infrastructures routières ; Marges ; Migration ; Rwanda.

ABSTRACT

The starting point of this thesis is three important facts about the reality of economic geography in Africa. These are respectively: the existence of internal trade costs, the spatial dispersion of prices, and the spatial differences in cost-of-living across cities.
The economic literature related to these facts gives us the opportunity to address three questions: (i) is estimating internal trade costs using spatial price differences a good metric for actual trade costs? ii) what is the size of internal trade costs in Rwanda and what are the effects of a better road connectivity on these costs? (iii) what determines the spatial differences in the cost of living across cities in Ethiopia and what are their implications?
To address these questions, this thesis proceeds along three chapters. The first chapter assesses the method of inferring internal trade costs from the spatial price differences. The second chapter uses this method to measure the size of internal trade costs in Rwanda and analyzes the effects of road infrastructure on these costs. Enriched with the implications drawn from previous chapters, the third chapter examines the sources and implications of spatial cost-of-living differences across cities in Ethiopia.

The first chapter entitled *"The measurement of trade costs in Ethiopia"* compares internal trade costs estimated using spatial price differences with the actual trade costs incurred by the World Food Program in Ethiopia. Based on Atkin and Donaldson (2015)'s method, I estimate the trade costs elasticity with respect to distance for both food and non-products. For food products, I find a similarity between the elasticities of estimated and actual trade costs with respect to distance. This result suggests that, inferring trade costs from spatial price dispersion is a good method to approximate actual trade costs. I find that the estimated trade costs elasticity is lower for non-food products compared to food products

The second chapter entitled *"Trade costs, prices and connectivity in Rwanda"* estimates the size of internal trade costs in Rwanda and assesses the effects of a better road connectivity on these costs. Internal trade costs are measured using the method proposed by Atkin and Donaldson (2015). This method consists to infer internal trade costs from the spatial difference in prices purged from markups charged by sellers. We first show that internal trade costs are high in Rwanda. Indeed, defined as a function of distance, internal trade costs in Rwanda are 10 times higher than in the USA, 3 times higher than in Ethiopia and 2 times higher than in Nigeria. Second, we find that these costs are particularly high for manufactured products. Last, our results suggest that a better road connectivity leads to a reduction of internal trade costs. On average, connecting a location to a feeder road reduces trade costs by approximately 2 cents per mile. This effect transits through the connection of this feeder road to a national road.

The third chapter entitled *"The price of remoteness: product availability and local cost of living in Ethiopia"* examines the determinants and implications of spatial cost-of-living differences across cities in Ethiopia. First, the empirical analysis uses the microdata underlying the Ethiopian CPI to examine the spatial dispersion in local prices and availability of 400 items across 110 cities. The results show that remote cities face higher prices and have access to fewer products. Furthermore, large cities also face higher individual prices but enjoy access to a wider set of products. Then, to assess the welfare implications of these patterns, we build aggregate cost-of-living indexes. The results imply that the cost of living is higher in remote areas, but not systematically related to population size. Finally, We then show spatial differences in the cost of living are a significant determinant of migration flows across Ethiopian regions.

Keywords : Internal trade costs; Cost-of-living, Spatial dispersion of prices, Availability, Distance, Remoteness, Road infrastructures, Markups, Migration, Ethiopia, Rwanda.

INTRODUCTION

Contexte. Depuis les travaux fondateurs de Paul Krugman (voir Krugman, 1991a,b), l'économie géographique occupe une place importante dans l'analyse économique. En effet, le nombre d'articles intégrant la dimension géographique de l'activité économique ne cesse de croître. Fort de cet intérêt grandissant, les économistes rejoignent les géographes autour d'un objet d'analyse commun : l'espace. Prendre en compte l'espace dans l'analyse économique est important car la plupart des faits marquants du paysage économique y sont liés. Ceci est valable pour les pays développés, mais aussi pour les pays en développement et en particulier les pays africains. Trois faits majeurs attirent l'attention dans la géographie de l'activité économique en Afrique.

Premièrement, l'existence des coûts au commerce interne. Dans son rapport annuel sur l'état du commerce dans le monde, l'Organisation Mondiale du Commerce indique que les coûts au commerce interne sont particulièrement larges dans les pays africains (voir WTO, 2004). Ces coûts incluent entre autres, les coûts de transport ou les pots-de-vin. En outre, dans une étude portant sur 42 pays d'Afrique subsaharienne, Porteous (2019) montre que les coûts médians du commerce interne dans les pays africains sont cinq fois plus élevés que dans le reste du monde. L'importance des coûts au commerce dans les pays africains interroge les chercheurs quant à leurs magnitudes, sources et implications potentielles.

Deuxièmement, la dispersion spatiale des prix des biens et services. En effet, aussi bien entre les pays qu'à travers les villes au sein d'un même pays, ces prix varient en fonction de la localisation. En Éthiopie, les données de l'agence centrale de la statistique révèlent que le prix des céréales est plus élevé dans les grandes villes et les zones reculées. Les céréales constituant la principale denrée alimentaire dans ce pays, cette hétérogénéité spatiale des prix a des enjeux en termes de sécurité alimentaire qui rendent impérieuse la compréhension de ses déterminants.

En dernier lieu, l'inégalité spatiale du coût de la vie. Le coût de la vie incorpore à la fois les prix et la disponibilité des biens et services. Dans le cadre du rapport annuel sur l'état de la sécurité alimentaire dans le monde, l'Organisation des Nations Unies pour l'alimentation et l'agriculture fait état de la disponibilité limitée voire de l'absence d'un ensemble de produits alimentaires de base dans certaines zones en Éthiopie (voir FAO et al., 2018). Cette absence de produits alimentaires se traduit par un coût de la vie élevé dans ces zones comparativement au reste du pays.

État de la littérature. La littérature économique s'est intéressée aux faits cités ci-dessus.

S'agissant de la littérature sur les coûts au commerce interne, les débats portent autant sur leur mesure que sur leurs déterminants. La littérature sur la mesure des coûts au commerce interne est assez récente et prend deux directions. La première avenue consiste à collecter des données sur les coûts du transport routier à travers des enquêtes auprès des compagnies de transport (voir Raballand et Teravaninthorn, 2009 ; de Rochambeau et Hjort, 2017 ; de Rochambeau, 2018). Ces mesures directes se caractérisent par leur rareté et leur représentativité limitée. La seconde avenue de mesure consiste à estimer les coûts au commerce interne à partir de la différence spatiale des prix (voir Atkin et Donaldson, 2015 ; Porteous, 2019). Cette procédure est prisée par les chercheurs du fait de la disponibilité des données de prix. Cependant, l'évaluation de sa capacité à approximer correctement les coûts réels n'a pas encore été abordée. En outre, le caractère récent de cette littérature offre l'opportunité de documenter la taille des coûts au commerce interne dans les pays africains. Les travaux sur les déterminants des coûts au commerce interne révèlent que la distance et la qualité des infrastructures en sont les principales sources (voir Rancourt et al., 2014 ; Storeygard, 2016 ; Jedwab et Storeygard, 2017 ; Donaldson, 2018). Il y a peu de travaux qui portent sur les pays africains. Cependant, la multiplication des projets d'investissements infrastructurels en Afrique, notamment dans un pays comme le Rwanda, donne la possibilité de documenter davantage cette littérature.

En ce qui concerne la dispersion spatiale des prix et du coût de la vie, la littérature s'attarde sur les différences de coût de la vie entre les zones urbaines et les zones rurales (voir Ravallion et Van De Walle, 1991 ; Deaton et Tarozzi, 2000 ; Muller, 2002). Cependant, la différence spatiale du coût de la vie entre les villes est moins abordée alors qu'elle permet de prendre en compte deux éléments importants de la géographie à savoir la taille des villes et leur éloignement. La question de la mesure du coût de la vie fait également l'objet de débat. Les principaux indicateurs du coût de la vie présents dans la littérature intègrent principalement les prix alors que la disponibilité des biens, qui constitue pourtant un aspect important du coût de la vie est négligée (voir Timmins, 2006 ; Ferré et al., 2012).

Questions de recherche. Compte tenu des trois faits évoqués précédemment et des opportunités qu'offre l'état de la littérature y afférente, cette thèse répond aux trois questions suivantes : (i) la méthode d'estimation des coûts au commerce interne à partir des différences spatiales de prix est-elle valide ? (ii) quelle est la taille des coûts au commerce interne au Rwanda et quels sont les effets d'une meilleure connectivité routière sur ces coûts ? (iii) qu'est-ce qui détermine la différence spatiale du coût de la vie en Éthiopie et quelles en sont les implications ?

Plan et contenu de la thèse. Pour répondre aux questions de recherche, cette thèse s'articule en trois chapitres. Le premier chapitre fait un exercice de validation de la méthode d'inférence des coûts au commerce interne à partir de la différence spatiale des prix. Une fois cette méthode validée, le deuxième chapitre l'utilise pour évaluer la taille des coûts au commerce au Rwanda, et analyser les effets sur ces coûts d'une meilleure connectivité routière. Enrichi des leçons tirées des chapitres précédents, le troisième chapitre évalue les sources et les implications de la différence spatiale du coût de la vie en Éthiopie.

Le premier chapitre intitulé *"The measurement of trade costs in Ethiopia"* compare les coûts au commerce interne estimés et réels en Éthiopie. L'estimation des coûts au commerce interne est basée sur la méthode proposée par Atkin et Donaldson (2015). Cette méthode consiste à inférer les coûts au commerce interne à partir de la différence spatiale des prix corrigée de la variation spatiale des marges chargées par les vendeurs. Les coûts au commerce interne réels sont mesurés à l'aide des coûts encourus par le

Programme Alimentaire Mondial (PAM) au cours de la distribution de l'aide alimentaire en Éthiopie. En utilisant la différence spatiale des prix des produits alimentaires pour estimer les coûts au commerce interne, les résultats montrent que l'élasticité estimée des coûts au commerce par rapport à la distance est assez proche de l'élasticité des coûts réels. Cela nous permet par la suite d'appliquer cette méthode pour les produits non-alimentaires. La contribution majeure de ce chapitre est qu'il permet de comparer des mesures directes et indirectes des coûts au commerce interne. Dans la littérature sur les coûts au commerce interne, qui est assez récente, cet exercice n'avait pas encore été effectué. Ce chapitre montre donc qu'utiliser la différence spatiale des prix corrigée des marges est une bonne méthode pour approximer les coûts au commerce interne. Cette méthode peut donc être appliquée dans le deuxième chapitre de cette thèse.

Le second chapitre intitulé *"Trade costs, prices and connectivity in Rwanda"* mesure les coûts au commerce interne au Rwanda, et évalue les effets d'une meilleure connectivité routière sur ces coûts. Les coûts au commerce interne sont mesurés à l'aide de la méthode proposée par Atkin et Donaldson (2015). Cette méthode consiste à inférer les coûts au commerce interne à partir de la différence spatiale des prix corrigée de la variation spatiale des marges chargées par les vendeurs. En ce qui concerne la connectivité routière, nous définissons trois niveaux de connectivité routière en fonction de la proximité de marchés avec des routes secondaires ou nationales. Ces trois niveaux de connectivité routière sont : faible connectivité, connectivité intermédiaire et forte connectivité. Trois principaux résultats émergent de notre analyse.
D'abord, nos estimations suggèrent que les coûts au commerce sont larges au Rwanda. En effet, définis comme fonction de la distance, les coûts au commerce au Rwanda sont 10 fois plus élevés qu'aux États-Unis, 3 fois plus élevés qu'en Éthiopie et deux fois plus élevés qu'au Nigéria. Le pouvoir de marché élevé des compagnies de transport, la faible qualité des infrastructures et le relatif enclavement du Rwanda expliquent ce résultat. Ensuite, nous démontrons que les coûts au commerce sont relativement plus élevés pour les produits manufacturés par rapport aux produits non manufacturés. La différence dans les modes de transport utilisés constitue une explication possible de ce résultat. Enfin, s'agissant du rôle des infrastructures routières, nos estimations montrent

qu'une meilleure connectivité routière des marchés se traduit par une réduction des coûts au commerce. En particulier, les coûts au commerce sont faibles lorsque les marchés sont connectés à la fois aux nouvelles routes secondaires construites par le gouvernement rwandais et aux routes nationales existantes. La connexion aux nouvelles routes secondaires est pertinente pour la réduction des coûts commerce, lorsque celles-ci permettent l'accès aux routes nationales. En somme, ce chapitre documente la littérature sur la taille des coûts au commerce interne dans les pays africains. Il complémente également la littérature sur les effets économiques des infrastructures de transport en mettant l'accent sur les effets de la connectivité en milieu rural.

Le dernier chapitre intitulé *"The price of remoteness : product availability and local cost of living in Ethiopia"* examine les déterminants et les implications de la dispersion spatiale du coût de la vie à travers les villes en Éthiopie. L'analyse est d'abord faite d'un point de vue désagrégé en considérant la dispersion spatiale des deux composantes du coût de la vie : les prix et la disponibilité des produits. Les résultats indiquent que les villes éloignées font face à des prix élevés et ont accès à une gamme restreinte de produits. De plus, les prix sont également élevés dans les grandes villes, mais ces dernières ont accès à plus de variétés. Partant de ces résultats, le chapitre analyse ensuite d'un point de vue agrégé la dispersion spatiale des indices du coût de la vie. Ces indices agrégés du coût de la vie prennent en compte à la fois les prix et la disponibilité des produits. Les résultats montrent que le coût de la vie est élevé dans les zones reculées, mais qu'il n'est pas systématiquement lié à la taille des villes. Le chapitre s'intéresse enfin à l'impact de la dispersion spatiale du coût de la vie sur les migrations. Les résultats montrent que les migrants quittent les villes où le coût de la vie est élevé pour s'installer dans les villes il est plus faible. Ce chapitre contribue de plusieurs manières à la littérature. En premier lieu, il construit des indices de coûts de la vie prenant en compte à la fois le prix et la disponibilité des produits alors que les indices existants se concentrent sur les prix. En second lieu, il se départit de la distinction entre l'urbain et le rural qui prédomine dans la littérature, en analysant la dispersion spatiale du coût de la vie entre les villes. Enfin, il documente la littérature sur les déterminants des flux migratoires internes dans un contexte de pays en développement.

CHAPTER I

THE MEASUREMENT OF TRADE COSTS IN ETHIOPIA

Abstract

This paper compares trade costs estimated using spatial price differences with the actual trade costs incurred by the World Food Program in Ethiopia. Based on Atkin and Donaldson (2015)'s method, I estimate the trade costs elasticity with respect to distance for food products. I find that the estimated trade costs elasticity is significantly similar than the actual trade costs elasticity. This result suggests that inferring trade costs from spatial price dispersion is a good method to approximate actual trade costs. Therefore, I propose an estimate of trade costs elasticity for non-foods products. I find that trade costs elasticity with respect to distance is lower for non-food products in Ethiopia.

Keywords: Estimated Trade costs; Prices dispersion; actual trade costs; World Food Program; Ethiopia.

JEL Classification: F14, L13, O18.

1.1 Introduction

The trade literature has extensively investigated the measurement and sources of trade costs. [1] While much work has focused on the measurement of international trade costs, internal trade costs have attracted less attention. Measuring and understanding internal trade costs is important for two reasons. First, domestic trade costs are an impediment for productivity and a drag for incomes (Sotelo, 2020). Second, an accurate metric of internal trade costs is important for welfare calculations (see e.g. Albrecht and Tombe, 2016). A recent literature has tried to estimate these costs using spatial dispersion of prices since direct measures are sparse, particularly in developing countries (see e.g. Atkin and Donaldson, 2015; Porteous, 2019). Is estimating internal trade costs using spatial price differences a good metric for actual trade costs?

To tackle this question, this paper estimates the elasticity of trade costs with respect to distance using spatial price differences across Ethiopia, and compares it with the actual trade costs elasticity.
I rely on the method proposed by Atkin and Donaldson (2015) to estimate the elasticity of trade costs with respect to distance. This method consists to infer trade costs from spatial dispersion of prices by respecting three conditions. First, collect high-frequency data on prices for narrowly defined products. Second, identify precisely the origin of products to ensure that spatial price differences correctly identify trade costs. Last, correct spatial price differences for the markups charged by intermediaries. This correction consists, on the one hand, to estimate the effect of a change in origin prices on destination prices; and on the other hand, to use this effect to correct the price difference.
I apply this method using micro-price data underlying the Ethiopian CPI and estimate the effect of distance on trade costs for both non-food and food products. Unlike to Atkin and Donaldson (2015) which estimate semi-elasticity of trade costs with respect to distance, this paper estimates elasticity.
Regarding the elasticity of actual trade costs with respect to distance, I identify these costs using data on costs incurred by the World Food Programme (WFP) while deliver-

1. See Anderson and Van Wincoop (2003) and the literature cited therein.

ing food aid across Ethiopia. I merged the WFP transport costs data with micro-price data. The regression of the log of the WFP transport costs on the log of distance allows me to recover the elasticity of actual trade costs relative to distance.

Starting with food products, I find a similarity between estimated and actual trade costs elasticities. These results hold when I control for the quality of the road by using travel time between locations instead of distance as determinant of trade costs. This results may be explained by the fact that intermediaries operate on the same origin- destination pairs as the WFP and transport same type of products. In a nutshell, my results show that using spatial price dispersion to infer trade costs is a good method.

Applying this method for non-food products, I find that the effect of distance on trade costs is lower than for food . More specifically, the estimated trade costs elasticity is about 3 times lower. This result may be explained by the fact that transport modes between these two types of goods are differents. In addition food products are perishable and require a specfic packaging , and therefore more costs.

This paper contributes to the literature on the measurement of trade costs. Some works in this literature have compared direct and indirect approaches to measure trade costs. However, their focus was on international trade costs and developed countries. For instance, using standards and technical regulations data in European Union, Chen and Novy (2012) show a contrasted effect of standards on bilateral trade frictions and actual trade flows. The present study differs due to its interest on assessment of indirect measures of internal trade costs. This paper also contributes by using an accurate direct measure of internal trade costs in a context of developing country where this type of measure is generally sparse and inaccurate.

Related literature. This paper pertains to two strands of the literature.
First, it relates to the recent literature that examines the size and the economic effects of internal trade costs in developing countries (see e.g. Raballand and Teravaninthorn, 2009; Atkin and Donaldson, 2015; de Rochambeau, 2018a; Sotelo, 2020). The message that emerges from this literature is that trade costs are high in developing countries and

constitute an impediment for productivity and welfare. Among others, Porteous (2019) shows that internal trade in Sub-Saharan Africa countries is five times higher than elsewhere. He also points out that reducing these costs leads to a reduction of food prices. Sotelo (2020) estimates trade costs in Peru using a model of agriculture and data on the transportation system. He shows that these costs are large and then evaluates their effects on productivity and welfare. By estimating the elasticity of trade costs relative to distance in Ethiopia, this paper fits with studies in this literature.

Second, this paper is also related to the role of geography and infrastructure on internal trade costs (see e.g., Brenton, Portugal-Perez and Régolo, 2014; Jedwab and Storeygard, 2017; Donaldson, 2018). Distance is the main geographical determinant of trade costs in the literature. Investigating the effects of distance on transport costs in Canada using trucking micro-data, Behrens and Brown (2016) report a monotically increasing relation between transport costs and distance. In the same vein, Rancourt et al. (2014) find that distance explains 27% of the variability in transport costs within Ethiopia using trade costs incurred by the World Food Program. In line with this literature, I use distance as the main determinant of internal trade costs. Many authors have also investigated the role of infrastructure on internal trade costs. Donaldson (2018) evaluates the effect of railroads on trade costs in India. He finds that railroads decrease trade costs and spatial price differences. This paper fits with this literature since I look at the effect of travel time on estimated and actual trade costs.

The remainder of the paper is organized as follows. Section 1.2 presents the data I used for the estimation of trade costs elasticities as well as descriptive statistics. Section 1.3 describes the methodology. Section 1.4 presents the baseline results and discusses their robustness. Last, Section 1.5 concludes.

1.2 Data and descriptive statistics

This section first provides a description of the various datasets used in the paper. Then, I present stylized facts on prices in Ethiopia.

1.2.1 Data

To estimate trade costs using spatial prices differences, highly frequency micro price data are crucial. I first present these data in the next section. Second, I describe data on observed trade costs. Third, I describe the geographic variables I used throughout the empirical analysis.

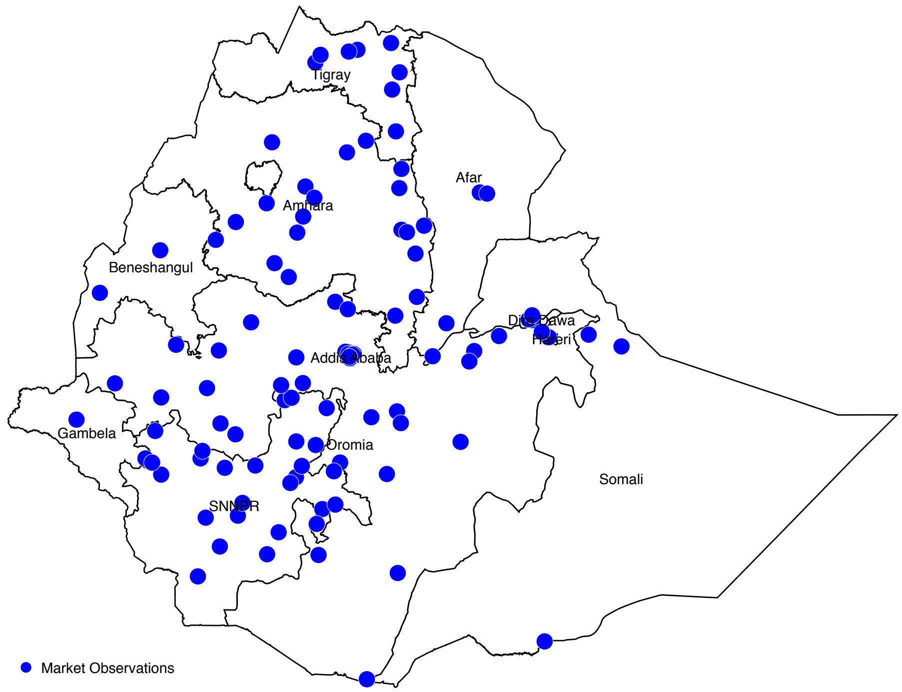
Micro-Price Data

Source. To build Consumption Price index, The Ethiopian central statistical agency (CSA) collects the data on retail prices in 117 markets across the country. These markets are urban centers located in each region of Ethiopia and are mapped in Figure 1.1. [2] In each market, enumerators collect price data on more than 400 products. Data collection activities are conducted from the first day to the 15th day of the month. I work with data from January 2011 to December 2015. Data for June, July and December 2013 were either incomplete or not available, so I excluded these months and ended up with a span period of 57 months. I use the same micro-price data as Atkin and Donaldson (2015) but for a more recent period of time. [3]

2. Ethiopia is divided into 9 regional states and 2 city administrations (Addis Ababa the capital city and Dire-Dawa).

3. Atkin and Donaldson (2015) use price data from September 2001 to June 2010.

Figure 1.1: ***Map of market observations***



Notes: This figure maps the 117 markets included in the CSA price survey. All regions are covered, but few markets are surveyed in Somali region due to armed conflicts.

Extraction and description. The CSA releases every month a 200-page document reporting the price of every product in each of the 117 markets.[4] For 57 months, I extracted this information from pdf files. The CSA price survey covers 427 products and services which include for instance food product or non-food products. The product description can be very precise with brandnames (e.g. "Detergents-Zahira-50 Gm", "Beer-Bedele- 300 Cc") or more generic (e.g. "Jeans Trouser", "T-shirt").

Selection of products. To correctly infer trade costs from prices, it is important to purge spatial price differences from the variation of quality between products. Therefore, I selected products of which description includes precisely the weight and the

4. Following Atkin and Donaldson (2015), a missing price in the price dataset means that enumerator was not able to find the product in that city during that month.

specific type of the product. I work with two main categories of products in the paper. In the empirical analysis, I use food and non-food products. Table 1.7 in appendix reports the list of non-food products.[5] This list includes a total of 15 products among which 6 are imported and 9 are locals. Imported products mostly come from England. Food products include a set of products similar to those distributed by the WFP across Ethiopia. As shown in Table 1.8 in appendix, I work with 118 food products. These products include for instance cereals, vegetables or fruits.

Outliers. Thanks to visual check, I have detected some outliers in the price data. A price might be abnormally high because the enumerator adds a figure by mistake when reporting the price. For instance, in Bedessa market, the reported price for the product "Motor Oil- Mobil Lt-1.2kg" in April 2015 is 1000 Birrs whereas all the reported prices in other months of 2015 are 100 or 110 Birrs. In Tercha market, the reported price for the product "Detergents-Zahira- 50 Gm" in June 2015 is 15.58 Birrs whereas the reported price for other months in 2015 range from 2.5 to 2,75 Birrs. I used a simple procedure to remove outliers. For each product, I dropped monthly observations for which the price was 5 times lower or 5 times higher than the median price across markets.[6] With this procedure, I dropped 51 observations out of more than 70000 for non-food products, and 544 observations out of more than 100000 for food products.

Origin locations. Infer trade costs from price gaps requires to know where the seller bought the product (origin location). Indeed, price gaps correctly identify the trade costs between two locations only when one location is the origin (location where the seller bought the product) and the other location is the destination (where the intermedi-

---

5. I use the same products as Atkin and Donaldson (2015) except of "Hard Soap (imported)-200 Gm".

6. This procedure is different from Atkin and Donaldson (2015). Atkin and Donaldson (2015)'s procedure consist to remove price quotes that lie more than 10 standard deviations away from the log mean price of a product.

ary sell the product to the consumer). [7] Consequently, I identified the production location of every product in the sample. For domestic products, I have identified firms that produce them and I consulted their websites to identify the production locations. For the products with many production locations, I chose for a given destination, the nearest production location. Each destination-product couple in the sample has a unique production location. Figure 1.2 shows these production locations in a map. There are 11 production locations: 10 for domestic goods and 1 for imported goods (Kombolcha). [8] Table 1.7 in appendix presents the origin locations for non-food products. Firms in food processing industry in Ethiopia are mostly installed in the capital (Addis-ababa) and particularly in Akaki market (see Sutton and Kellow, 2014). [9] Hence, I chose Akaki market as origin for local food products. For imported food products I chose Kombolcha, the main international trade corridor, as origin.

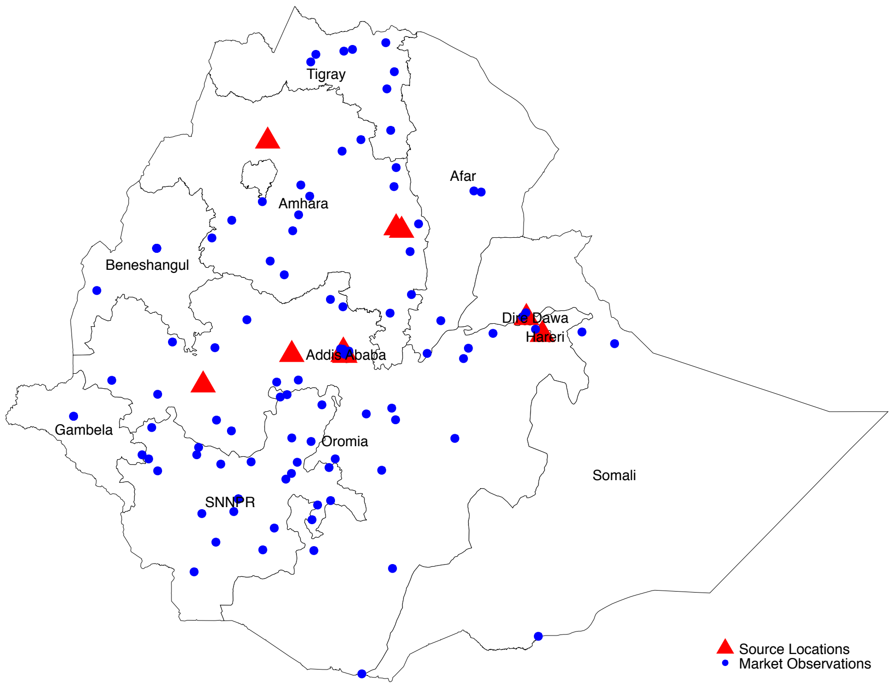
---

7. I consider in the empirical analysis only the origin-destination pairs for which the destination location is different from the origin location.

8. Kombolcha is the main international trade corridor through which shipment from and to Djibouti transit.

9. The CSA splits Addis ababa into 12 markets among which Akaki market.

Figure 1.2: ***Map of origin locations***



Notes: This figure maps the production locations of the products selected for this study.

Correction for inflation and exchange rates. In the empirical analysis, I work with inflation-adjusted prices in U.S. dollars. I deflated prices using inflation rates per month computed as follows. For every location-product-month triplet, I calculate the proportional change in price over the previous month. For each month, I compute the mean of these changes and used it as the monthly inflation rate. Then, I converted these prices using the Birr-U.S. dollar exchange rate prevailing in January 2011.[10] Finally, I end up with inflation-adjusted prices in U.S. dollars stripped of outliers.

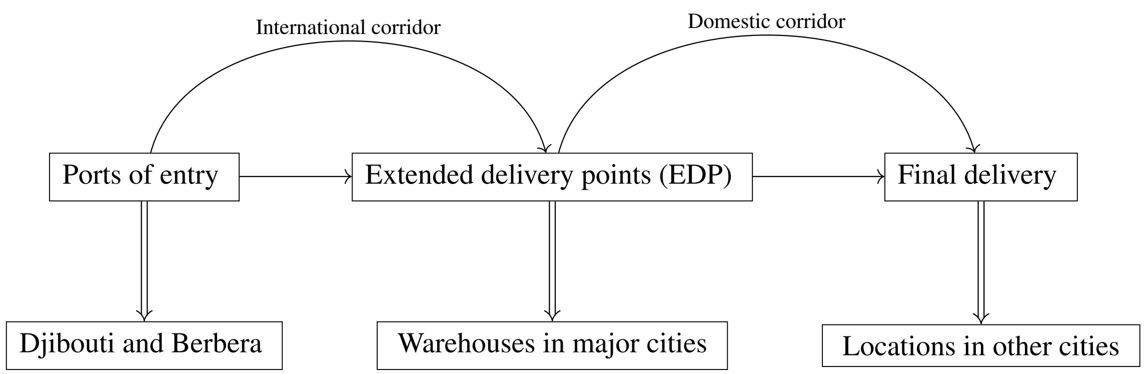
10. January 2011 is the first month of price collection. The Birr-U.S. dollar exchange rate in January 2011 was 1 Birr= 0.06 U.S. dollar. As the aim of this paper is to assess Atkin and Donaldson (2015) method, I use the same procedure they used to correct for inflation and exchange rates.

Actual trade costs

To compare estimated and actual trade costs, the availability of actual trade costs is critical. I use transport costs incurred by the WFP when providing food aid. [11]

The WFP transportation network in Ethiopia. The WFP provides food aid in Ethiopia since 1974. [12] Figure 1.3 describes the transportation network of food aid in Ethiopia. [13] This network includes two parts. The first part is an international corridor because Ethiopia is landlocked. Through this corridor, products are transported from Djibouti port (in Djibouti) or Berbera port (in Kenya) to warehouses in major Ethiopian cities such as Dire-dawa or Addis-Ababa. The WFP call these warehouses extended delivery points (EDP). The second part is a domestic corridor. Upon arrival at EDP, products are transported to various locations across the country based on food needs. These locations known as Final Delivery Points (FDP) are often remote from EDP. For this study, I only exploit data from domestic corridors because I am interested in trade costs within Ethiopia.

Figure 1.3: ***The WFP food aid corridors in Ethiopia***



11. The WFP transport costs data are not public. We obtained them thanks to Marie-Eve Rancourt, Professor at HEC Montreal who worked on food aid in Ethiopia. The costs only contain the transport costs. They do not contain distribution costs.

12. The WFP uses local transport companies to transport food within the country

13. See Rancourt et al. (2014) for more details on this transportation network.

How are the WFP transport costs determined? The WFP rely on Ethiopian transportation sector to deliver food aid across Ethiopia. Twice a year, the WFP launch a call for tender on lanes (EDP-FDP couples) that require transportation. For each lane, carriers make bids to the WFP which choose the lowest. Consequently, the WFP transport costs reflect correctly the structure of Ethiopian transportation market.

The WFP transport costs dataset. I have data on transport costs in Birrs incurred by the WFP on 957 lanes in 2011. [14] I converted these costs in U.S. dollars using the Birr-U.S. dollar exchange rate prevailing in January 2011. In the empirical analysis, I assume that these costs are constant over the period 2011 to 2015.

Distances and Travel times.

I complement my dataset with geographic variables. To assess how far a market is from another. I use geodesic distance in miles computed by the stata package Geodist. For robustness checks, I consider three other measures of distance. I first compute the great circle distance using the Stata package Geodist. Then, I calculate travel times and travel distance thanks to the Stata package Georoute (Weber and Péclat, 2016). Whereas travel distance measures the distance by car between two locations, travel time indicates the necessary amount of time to drive from one location to another.

1.2.2 Descriptive statistics

Summary statistics. Table 1.1 summarizes basic characteristics of my datasets. As shown in panel (a), the distribution of non-food products price at the production locations is positively skewed since the median price is about half of the average price. In addition, the price gap for non-food products shows an asymmetrical distribution, the average price gap being about three times greater than the median price difference. This asymmetry also characterizes both the distribution of food prices (panel (b)) and that of

14. In my dataset there are 35 extended delivery points (EDP) and 103 Final Delivery Points (FDP).

transport costs (panel (c)). However, food prices are more variable than non-food prices since the coefficient of variation of food prices is higher. Regarding geographic variables, the average distance from origin to destination is higher for WFP products. This is not surprising since the WFP provides the food aid in remote locations in Ethiopia.

Table 1.1: ***Descriptive statistics***

	Mean	SD	CV	P25	P50	P75	Observations
	Panel (a) : Non-food products dataset						
	Mean	SD	CV	P25	P50	P75	Observations
Origin price	0.939	1.170	1.246	0.300	0.499	0.780	74090
Price gap	0.055	0.234	4.231	-0.011	0.018	0.099	74090
Distance from origin to destination	211.721	110.280	0.521	140.235	201.723	286.819	74090
Travel time from origin to destination	6.280	3.274	0.521	4.183	6.184	8.416	74090
Weight	241.533	209.815	0.869	40	330	330	74090
	Panel (b) : Food products dataset						
	Mean	SD	CV	P25	P50	P75	Observations
Origin price	1.771	2.843	1.605	0.530	0.900	1.560	164341
Price gap	0.325	0.911	2.801	0.059	0.120	0.300	164341
Distance from origin to destination	180.774	107.691	0.596	105.837	198.732	249.147	164341
Travel time from origin to destination	5.172	3.281	0.635	2.688	5.317	7.231	164341
Weight	947.630	183.745	0.194	1000	1000	1000	164341
	Panel (c) : The WFP transport costs dataset						
	Mean	SD	CV	P25	P50	P75	Observations
The WFP transport costs	71.990	51.668	0.718	30	58.800	107.400	957
Distance from origin to destination	370.013	220.223	0.595	196.570	358.174	517.080	957

Notes: This table presents summary statistics of the data I use in the paper. Prices are expressed in U.S. dollars (January 2011) whereas transport costs are expressed in U.S. dollars (January 2011) per tone. Distances are in miles, travel time in hours and weight in grams. Price gap is the difference between destination price and origin price.

1.3 Methodology

This section presents the methodology I use in the paper. After presenting the model, I develop the empirical strategy.

1.3.1 Theoretical Framework

The estimation of trade costs elasticity in this paper relies on the model of intermediated intra-national trade developed by Atkin and Donaldson (2015). I only present here the intuition and main assumptions/equations of the model. The full framework and proofs can be found in the initial paper.

The idea of the model. The environment of the model considers multiple products j potentially sold in multiple markets d at multiple periods t. Each good j can be domestically produced at a factory location o , or imported into the country through a border o. Firms or intermediaries purchase goods from the origin o, transport these goods to the destination d and sell them to consumers in this location. The price at the origin is denoted by P^j_{ot} and the price in location d is P^j_{dt}. Equation 1.1 is the basic equation to infer trade costs from spatial price differences.

$$P^j_{dt} - P^j_{ot} = \tau(X^j_{odt}) + v^j_{dt} \tag{1.1}$$

In this equation, v^k_{dt} is the markup charged by intermediaries in location d and time t when selling product j. $\tau(X^j_{odt})$ is the trade costs incurred by the intermediary when buying product j in location o and selling it in location d. X^j_{odt} represents the set of cost-shifters. The main cost-shifter is the distance x_{od} between the origin o and the destination d. Equation 1.1 implies that to infer trade costs from price differentials, it is important to identify precisely the pairs of markets which actually trade by recovering the locations where intermediaries bought their products. Moreover, it is crucial to purge price gaps from markups such that these price differences only identify trade costs.

Main assumptions. Atkin and Donaldson (2015) make four assumptions to estimate the markup charged by firms and intermediaries.

Assumption 1 : An intermediary selling q_d^j unit of product j in location d faces fixed costs F_d^j and is subject to the following total costs:

$$C(q_{od}^j) = \left(P_o^j + \tau(X_{od}^j)\right) q_d^j + F_d^j \tag{1.2}$$

Assumption 2 : In market d, there are m_d identical intermediaries selling the product j and choosing the quantity q_d which maximise their profits given the competitiveness in the market d.

Assumption 3 : Consumer preferences take the form of constant pass-through inverse demand as developed by Bulow and Pfleiderer (1983).

Assumption 4 : The pass-through rate ρ_{odt}^j measures the effect on the destination price of a shock on the origin price. ρ_{odt}^j is fixed across time within a product-origin-destination triplet.

That is $\rho_{odt}^j = \rho_{od}^j$

Fundamental equation to infer trade costs. Combining all the previous assumptions lead to the following equation:

$$P_d^j - P_o^j = \rho_{od}^j \tau(X_{od}^j) + (1 - \rho_{od}^j)(b_d t^j - P_o^j) \tag{1.3}$$

Equation 1.3 shows that the pass-through ρ_{od}^j and the level of the inverse demand $b_d t^j$ are sufficient to control for markups, competitiveness and preferences. This equation is fundamental to infer trade costs.

1.3.2 How to estimate the elasticity of trade costs with respect to distance?

I estimate the trade costs elasticity in two steps. I first estimate pass-through rates ρ^j_{od} and then use these estimates to infer the effect of a change in distance on trade costs.

Step 1: recover estimates $\widehat{\rho^j_{od}}$ of pass-through rates ρ^j_{od}. Combining equation 1.3 with the assumptions of the model leads to:

$$P^j_{dt} = \rho^j_{od} P^j_{ot} + \rho^j_{od} \tau(X^j_{od}) + (1 - \rho^j_{od})(b^j_{dt}) \tag{1.4}$$

Applying a fixed effects approach to decompose $b_d t^j$ (the level of inverse demand) and $\tau(X^j_{od})$ leads to the following equation:

$$P^j_{dt} = P^j_{ot} \rho^j_{od} + \mu^j_{odt} + \mu^j_{odt} t + \chi^j_{dt} \tag{1.5}$$

Where μ^j_{odt} is a destination-product fixed effect, $\mu^j_{odt} t$ is a destination-product linear time trend and χ^j_{dt} is an error term.

Therefore, pass-through rates can be estimated by a regression of the destination prices P^j_d on the origin prices P^j_o and adequate fixed effects.

Step 2: recover estimates of the elasticity of trade costs with respect to distance $\frac{\partial Log\left[\tau(X^j_{odt})\right]}{\partial Log(x_{od})}$. After recovering pass-through rates estimates $\widehat{\rho^j_{od}}$, I can estimate the elasticity of trade costs with respect to distance.

I make a difference here with the work of Atkin and Donaldson (2015) since my point of interest is the elasticity of trade costs with respect to distance $\frac{\partial Log\left[\tau(X^j_{odt})\right]}{\partial Log(x_{od})}$ and not the semi-elasticity $\frac{\partial \left[\tau(X^j_{odt})\right]}{\partial Log(x_{od})}$.

Recall equation 1.4, rearrange and apply logarithm leads to:

$$Log\left[\frac{P^j_{dt}-P^j_{ot}\rho^j_{od}}{\rho^j_{od}}-\frac{(1-\rho^j_{od})}{\rho^j_{od}}b^j_d\right]=Log\left[\tau(X^j_{od})\right] \tag{1.6}$$

I consider a fixed effects approach to decompose b^j_d (the level of the inverse demand) as a sum of fixed effects. In addition, I assume a log-linear trade costs function (see, e.g., Donaldson, 2018) and use the estimates of pass-through rates. I thus obtain the following specification:

$$Log\left[\frac{P^j_{dt}-P^j_{ot}\widehat{\rho^j_{od}}}{\widehat{\rho^j_{od}}}-\frac{(1-\widehat{\rho^j_{od}})}{\widehat{\rho^j_{od}}}\lambda^j_t-\frac{(1-\widehat{\rho^j_{od}})}{\widehat{\rho^j_{od}}}\lambda_d\right]=\beta_1 Log(x_{od})+\varepsilon^j_{dt} \tag{1.7}$$

In this equation λ^j_t is a time-product fixed effect, λ_d a destination fixed effect. x_{od} is the distance between o and d. The error term ε^j_{dt} captures all the other determinants of trade costs and measurement error in $Log\left[\frac{P^j_{dt}-P^j_{ot}\rho^j_{od}}{\rho^j_{od}}-\frac{(1-\rho^j_{od})}{\rho^j_{od}}b^j_d\right]$. $\widehat{\rho^j_{od}}$ represents the estimates of pass-through rates. Finally, β_1 is the elasticity of trade costs with respect to distance, the parameter of interest. [15]

To estimate equation 1.7, I need to empirically recover the expression

$$\left[\frac{P^j_{dt}-P^j_{ot}\widehat{\rho^j_{od}}}{\widehat{\rho^j_{od}}}-\frac{(1-\widehat{\rho^j_{od}})}{\widehat{\rho^j_{od}}}\lambda^j_t-\frac{(1-\widehat{\rho^j_{od}})}{\widehat{\rho^j_{od}}}\lambda_d\right]$$

In the rest of the paper, I refer to the expression $\left[\frac{P^j_{dt}-P^j_{ot}\widehat{\rho^j_{od}}}{\widehat{\rho^j_{od}}}-\frac{(1-\widehat{\rho^j_{od}})}{\widehat{\rho^j_{od}}}\lambda^j_t-\frac{(1-\widehat{\rho^j_{od}})}{\widehat{\rho^j_{od}}}\lambda_d\right]$ as "purged price gap".

15. If $\rho^j_{od}=1$ (the case of perfect competition or no-markup), the elasticity of trade costs with respect to distance is simply $\beta_1=\frac{\partial Log\left[P^j_{dt}-P^j_{ot}\right]}{\partial Log(x_{od})}$.

To identify empirically the purged price gap, I first regress $\left[\frac{P^j_{dt} - P^j_{ot}\widehat{\rho^j_{od}}}{\widehat{\rho^j_{od}}}\right]$ on $\left[\frac{(1-\widehat{\rho^j_{od}})}{\widehat{\rho^j_{od}}} \times \text{ product-time-FE}\right]$ and $\left[\frac{(1-\widehat{\rho^j_{od}})}{\widehat{\rho^j_{od}}} \times \text{ destination-FE}\right]$.

Then, I recover the residuals of this regression. These residuals identify empirically the purged price gap. A regression of the logarithm these residuals on the log of distance allows me to estimate the trade costs elasticity (β_1).

1.3.3 Strategy to compare estimated trade costs elasticity and actual trade costs elasticity.

Here, I use a two-steps strategy.

Step 1: Combine micro-price data with the WFP transport costs data

Using the WFP transport costs data, I identify all common origin-destination pairs with the micro-prices datasets (non-food products dataset and food-products dataset). I obtain two new datasets that give for each pair of origin-destination and on a monthly basis : price gaps and the costs paid by the WFP. In the rest of the paper, I refer to the dataset obtained by merging the non-food products dataset and the WFP transport costs dataset as "non-food merged dataset". Similarly, I refer to the dataset obtained by merging the food product dataset and the WFP transport costs dataset as "food merged dataset". The "non-food merged dataset" includes 36 origin-destination pairs whereas the "food merged dataset" includes 50 origin-destination pairs.

Step 2: Compare estimated and actual trade costs elasticities

Here, I use "food merged dataset" and "non-food merged dataset" to compare estimated and actual trade costs elasticities.

Weight adjustment. The WFP transport costs measure the cost to transport a given weight of products from an origin to a given destination and are expressed in dollars per tone. However, prices data give the price per product (and not per weight) which are expressed in dollars. Then, for an accurate comparison between estimated and actual trade costs elasticities, I corrected prices by the weight of products. More specifically, for each product, I divided the price by the weight of the product in grams. I also converted the WFP transport costs in dollars per gram. Consequently both the WFP transport costs and prices are expressed in dollars per gram.

The following equation presents the relationship between cost per weight and cost per product :

$$\frac{C}{Q} = \frac{C}{W} \times \frac{W}{Q} \tag{1.8}$$

Where $\frac{C}{Q}$ is the cost per product , $\frac{C}{W}$ the cost per weight and $\frac{W}{Q}$ the weight per product

Comparison of elasticities. After correcting prices by the product weights, I first estimate the trade costs elasticity using the strategy presented in section 1.3.2. Then, I compare this elasticity with observed trade costs elasticity obtained by regressing the logarithm of the WFP transport costs on the log of distance.

1.4 Results

1.4.1 Estimation of pass-through rates

Figure 1.4 plots the estimated pass-trough rates (obtained by running the regression 1.5) against the logarithm of distance to origin respectively for non-food products (panel a) and food products (panel b). It reveals that pass-through rate estimates mostly lie below one, regardless of the type of product [16]. This result means that shocks to origin prices are not completely transmitted to destination prices by the intermediaries. It can there-

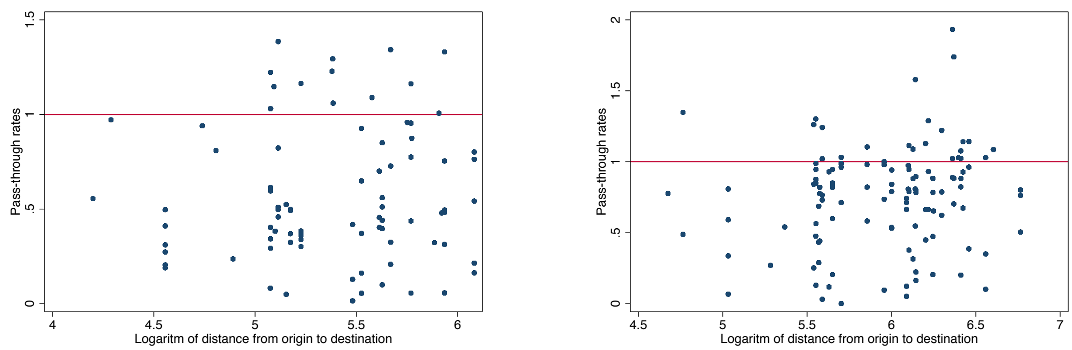
16. More than 75% of estimated pass-through rates are below 1 for both food and non-food products. 5% of estimated pass-through rates are below 0 for non-food products.

fore be assumed that there is imperfect competition in the intermediaries market. To complement Figure 1.4, Table 1.2 summarizes basic characteristics of estimated pass-through rates.[17] On average, non-food products have a lower estimated pass-through rate than non-food products (0.5 for non-food products and 0.7 for food products). A greater estimated pass-through for food products than for non-food products suggests a change in distance traveled (and therefore marginal cost) has more impact on markups and therefore on prices for food products. Thus, in remote areas, food traders are expected to charge higher margins than non-food traders. In addition, Table 1.2 shows that the distribution of pass-through estimates for non-food products is slightly right-skewed since the average pass-through is slightly greater than median pass-through. In contrast, the distribution of estimated pass-through for the food products is slightly negatively skewed since the median pass-through exceeds a little bit the mean pass-through. Turning to the relationship between pass-through and distance, Figure 1.5 in appendix shows that pass-through are lower when destinations are remote. Therefore, intermediaries tend to charge lower markups in remote areas.

---

17. The pass-through is a sufficient statistic to assess the distribution of margins in the markets. It is defined as the effect on destination prices of a change in the marginal cost of intermediaries. (In this chapter, this variation in marginal cost is caused by a variation in the origin-destination distance traveled by intermediaries). It can be interpret as the effect of a variation in the distance traveled by intermediaries on the markups they charge (this effect of the distance passes through an effect on the marginal cost of intermediaries). A greater estimated pass-through for food products than for non-food products suggests that a change in distance traveled (and therefore in marginal cost) has more impact on markups and therefore on prices for food products.Thus, in remote areas food traders are expected to apply higher margins than non-food traders.

Figure 1.4: ***Estimated Pass-through rates***



(a) Non-food merged dataset (b) Food merged dataset

Note: These graphs present respectively the estimated pass-through rates using the non-food merged dataset (panel a) and the food merged dataset (panel b). These estimated pass-through rates are obtained by running the regression 1.5.

Table 1.2: ***Statistics on estimated pass-through rates***

	Panel (a) : Non-food merged dataset					
	Mean	SD	CV	P25	P50	P75
Estimated pass-through rates	0.507	0.391	0.770	0.273	0.441	0.775
	Panel (b) : Food merged dataset					
	Mean	SD	CV	P25	P50	P75
Estimated pass-through rates	0.771	0.373	0.484	0.540	0.808	0.974

Notes: This table presents basic statistics on the estimated pass-through rates. These estimated pass-through rates are obtained when running the regression 1.5 using the non-food merged dataset (panel a) and the food merged dataset (panel b).

1.4.2 Estimated and actual trade costs elasticities: Baseline Results

In this subsection, I first compare estimated and actual trade costs elasticities for food products. Then, I propose an estimate of trade costs elasticity for non-foods products. Table 1.3 presents these results. Columns (1)-(2) present the results using food merged dataset whereas columns (3) report the results using the non-food merged dataset.

Starting with the non-food merged dataset, the following picture emerges. The estimated trade costs elasticity is quite similar to actual trade costs elasticity. A test of equality of the two coefficients shows that they are significantly equivalent. From a

quantitative point of view, a 10% increase in distance between origin and destination implies a 7.5% raise of actual trade costs and a 7.6% raise of estimated trade costs. This results may be explained by the fact that intermediaries operate on the same origin-destination pairs as the WFP and transport same type of products. Given that direct measures of trade costs are rarely available, most researchers in trade literature estimate trade costs using spatial price dispersion. The results in columns (1)-(2) validate this method since estimated and actual trade costs elasticities with respect to distance are quite similar.
Since Atkin and Donaldson (2015) 's method works well with for foods products, I apply it for non-foods products. Column (3) displays the results and shows that a 10% increase in distance between origin and destination implies a 2.6% raise of estimated trade costs for non-foods products. Hence, estimated trade cost elasticity is almost three times higher for food products than non-food products. This result may be explained by the difference in terms of products'type and transport modes. [18] The packaging of products for transport is different depending on the type of product. For example fruits (food products) require a refrigerated environment while cigarettes do not require this. This difference in packaging generates a difference in terms of the cost of transporting the two types of products. Thus, the difference in the type of products transported induces a difference in the cost of transport. In addition, the transport of certain types of products is subject to special regulations that can also have an impact on the cost of transport.

Table (1)-(3) also show that the R2 picks up much more of the variation for the actual trade costs than for purged price gaps. This results means that actual costs should be more accurate in terms of predictions.

18. These two explanations are linked because transport modes could vary by products.

Table 1.3: ***Baseline results***

	Food merged dataset		Non-food merged dataset
	Log of purged price gap (1)	Log of actual trade costs (2)	Log of purged price gap (3)
Log distance to origin	0.757^a	0.748^a	0.264^a
	(0.020)	(0.013)	(0.054)
Observations	3275	3275	2012
R-squared	0.24	0.60	0.36

Notes: Column (1) and (3) present the regressions estimating equation 1.7 using respectively the food merged dataset and the non-food merged dataset. Column (2) displays the regressions of actual trade costs on distance using the food merged dataset. All these regressions include product-time fixed effects. Standard errors clustered at product-time level are in parentheses.^{a}, ^{b} and ^{c} respectively denote significance at the 1, 5, and 10% levels.

1.4.3 Robustness checks

In this section, I check the robustness of my baseline empirical results. More specifically, I propose two types of checks. First, the quality of roads is an important determinant of trade costs (Limao and Venables, 2001), particularly in a context of developing country such as Ethiopia. To take this fact into account, I consider alternative distance metrics that adjust for the road quality such as travel time or travel distance between locations. Second, in the methodology, I supposed that the level of inverse demand can be decomposed as a sum of of product-time fixed effects and destination fixed effects. I test the sensitivity of this assumption by replacing destination fixed effects by destination-year fixed effects.

Tables 1.4 presents the comparison between estimated and actual trade costs elasticities with respect to distance using alternative measures of distance. Changing the distance metrics do not change the baseline empirical regularities Indeed, estimated and actual trade costs elasticities with respect to travel distance or great circle distance are similar. However, There is a slight difference between these elasticities when travel time is used as distance metric.

Tables 1.5 presents the estimated trade costs elasticities for non-foods products using alternative measures of distance. The results are similar to the the baseline results when considering travel distance (columns (2)) or great circle distance (columns (3)) as distance metrics. Table 1.6 shows the results when destination-year fixed effects are included in the level of inverse demand. My baseline results are not too sensitive to this change.

Table 1.4: ***Observed versus estimated trade costs elasticity : alternative measures of distance***

	Food merged dataset					
	Log of Purged price gap			Log of actual trade costs		
	(1)	(2)	(3)	(4)	(5)	(6)
Log of travel time to origin	0.839^a			0.731^a		
	(0.027)			(0.005)		
Log of travel distance to origin		0.649^a			0.637^a	
		(0.028)			(0.004)	
Log of great circle distance to origin			0.758^a			0.748^a
			(0.020)			(0.012)
Observations	3,275	3,275	3,275	3,275	3,275	3,275
R-squared	0.25	0.23	0.24	0.52	0.48	0.60

Notes: Column (1) to (3) present the regressions estimating equation 1.7 using the food merged dataset and other distance metrics. Column (3) to (6) display the regressions of actual trade costs on distance using the food merged dataset and other distance metrics. All these regressions include product-time fixed effects. Standard errors clustered at product-time level are in parentheses.[a], [b] and [c] respectively denote significance at the 1, 5, and 10% levels.

Table 1.5: ***Trade costs elasticity for non-food : alternative measures of distance***

	Non-food merged dataset		
	Log of Purged price gap		
	(1)	(2)	(3)
Log of travel time to origin	0.176^a		
	(0.046)		
Log of travel distance		0.269^a	
		(0.053)	
Log of great circle distance			0.265^a
			(0.054)
Observations	2,102	2,102	2,102
R-squared	0.35	0.36	0.36

Notes: Column (1) to (3) present the regressions estimating equation 1.7 using the non-food merged dataset and other distance metrics. . All these regressions include product-time fixed effects. Standard errors clustered at product-time level are in parentheses.^{a}, ^{b} and ^{c} respectively denote significance at the 1, 5, and 10% levels.

Table 1.6: ***Including destination-year fixed effects in the level of inverse demand***

	Food merged dataset	
	Log of Purged price gap	Log of actual trade costs
	(1)	(2)
Log of distance to origin	0.729^a	0.795^a
	(0.055)	(0.017)
Observations	3,185	3,185
R-squared	0.23	0.60

Notes: Column (1) presents the regressions estimating equation 1.7 while replacing in the left-side the destination fixed effects by a destination-year fixed effects. Column (2) represents the regression using the food merged dataset. All these regressions include product-time fixed effects. Standard errors clustered at product-time level are in parentheses.^{a}, ^{b} and ^{c} respectively denote significance at the 1, 5, and 10% levels.

1.5 Conclusion

Measuring trade costs is an important issue in both trade literature and public policy. Indeed, recent models of trade include trade costs and these costs are important for welfare calculations. The trade literature has focused on the measurement of international trade costs whereas the trade costs within countries have received little attention, particularly in developing countries.

In a seminal work, Atkin and Donaldson (2015) propose to measure intra-national trade costs using spatial prices differences that are corrected for spatial variation of markups. Is estimating internal trade costs using spatial price differences a good method to approximate actual trade costs? To address this question, this paper uses micro-data underlying the CPI and the trade costs incurred by the World food program in Ethiopia to compare estimated and actual trade costs elasticities. The picture that emerges is that estimated and actual trade costs elasticities are significantly close for foods products. This result shows that spatial dispersion of prices accurately estimates trade costs within a country. As this method works well for foods products, I applied it to non-foods products and I found that in Ethiopia, the trade costs elasticity relative to distance is lower for non-foods products.

1.6 Appendix to Chapter 1

1.6.1 Additional tables on the data

Table 1.7: ***Non-food products list***

Product	**Source Locations**	**Origin Country**
Batteries-Eveready Drycell No-40gm	Kombolcha	Kenya
Cigarettes-Rothmans- 20pack	Kombolcha	England
Detergents-Zahira- 50 Gm	Kombolcha	Yemen
Electric Bulb-Philips 40/60 Watt No- 20 Gm	Kombolcha	European union
Hard Soap (imported)- 200 Gm	Kombolcha	England
Pen Ball Point-Bic England No-20 Gm	Kombolcha	England
Beer-Bedele- 300 Cc	Bedele	Ethiopia
Beer-Harar- 330 Cc	Harari-Town	Ethiopia
Beer-Meta Abo- 330 Cc	Merkato	Ethiopia
Hair oil-Zenith Non-Liquid Form- 330 Cc	Akaki	Ethiopia
Motor Oil- Mobil Lt- 1.2kg	Kera	Ethiopia
Soft Drinks-Coca Cola - 300 Cc	Dire-dawa & Merkato	Ethiopia
Soft Drinks-Pepsi Cola Mirinda- 300 Cc	Dessie & Gondar	Ethiopia
Water Bottled-Ambo Mineral Water- 500 Cc	Ambo	Ethiopia
Wine-Saris Normal- 750 Cc	Saris	Ethiopia

Notes: This table presents the list of non-food products and the place where they are produced.For imported products, I consider Kombolcha as the port of entry because it is the main city on the corridor to djibouti.

Table 1.8: ***Food products list***

Product	Weight	Product	Weight	Product	Weight
’Bula’	Kg	Cow Milk (Unpasteurized)	Lt	Peas Mixed	Kg
’Dabo’ (Traditional Ambasha)	350gm	Cow Milk (pasteurized)	Lt	Peas Split	Kg
’Dube’ Flour	Kg	Dry Yeast(Baking powder)	350gm	Peas Split(Roasted)	Kg
’Enjera’ (’Teff’ Mixed)	325gm	Durrah	Kg	Peas White	Kg
’Fafa’ Flour	Kg	Ethiopian Kale	Kg	Pepper Green	Kg
’Furno Duket’ Locally Processed	Kg	Fenugreek Milled	Kg	Pepper Whole	Kg
’Kocho’ (Unprocessed)	Kg	Fenugreek(Green)	Kg	Potato	Kg
’Teff’ Black (Red)	Kg	Fish Fresh	Kg	Powdered Milk (Me&My)	450gm
’Teff’ Mixed	Kg	Garlics	Kg	Pumpkin	Kg
’Teff’ White	Kg	Ginger Dry(Local)	Kg	Rape Seed	Kg
African Millet	Kg	Ginger Wet(Local)	Kg	Rice (Imported)	Kg
Avocado	Kg	Grapes	Kg	Salt	Kg
Banana	Kg	Green Peas	Kg	Sardines (Imported)	125gm
Barley Black	Kg	Ground Nut Shelled	Kg	Sesame Seed Red	Kg
Barley Mixed	Kg	Haricot Beans	Kg	Sorghum	Kg
Barley White	Kg	Honey	Kg	Sorghum Red	Kg
Barley for Beer	Kg	Horse Beans	Kg	Sorghum White	Kg
Basil Dry	Kg	Horse Beans Milled	Kg	Sorghum Yellow	Kg
Beet Root	Kg	Horse Beans Split(Roasted)	Kg	Soya Beans	Kg
Biscuits	150gm	Hulled Barley	Kg	Spaghetti (Local) Without Egg	Kg
Black Cumin(Local)	Kg	Leaks	Kg	Spinach	Kg
Black Pepper(Local)	Kg	Lemon	Kg	Sugar	Kg
Bread Wheat (Bakery)	350gm	Lentils	Kg	Sunflower	Kg
Butter Unrefined	Kg	Lentils Split	Kg	Sweet Potato	Kg
Cabbage	Kg	Lettuce	Kg	Tangerine	Kg
Cactus	Kg	Lima Beans	Kg	Tomatoes	Kg
Camel Milk	Lt	Linseed Red	Kg	Tumeric Flour(Local)	Kg
Canned Tomato (Local)	410gm	Linseed White	Kg	Vegetable Butter(Sheno & Shady)	Kg
Cardamon(Local)	Kg	Long Pepper(Local)	Kg	Vetch	Kg
Carrot	Kg	Macaroni (Local) Without Egg	Kg	Vetch Milled	Kg
Castor Beans	Kg	Maize (White)	Kg	Vetch Split(Roasted)	Kg
Cauliflower	Kg	Mango	Kg	Wheat Black (Red)	Kg
Cheese Cottage	Kg	Mixed Pulses Milled	Kg	Wheat Mixed	Kg
Chick Peas	Kg	Niger Seed	Kg	Wheat White	Kg
Chick Peas Milled	Kg	Oats	Kg	White Cumin(Bishop’s Weed)Local	Kg
Chick Peas Split(Roasted)	Kg	Onions	Kg	Yoghurt (Traditional)	Lt
Chillies Whole	Kg	Orange	Kg		
Cinnamon(Imported)	Kg	Papaya	Kg		
Cloves(Imported)	Kg	Pastini	Kg		
Cooking Oil (Imported)	Lt	Peas Green(dry)	Kg		
Cooking Oil (Local)	Lt	Peas Milled	Kg		

Notes: This table presents the list of food products and their weights.

Heterogeneity in inflation rates. I now turn to stylized facts about inflation rates. To do this exercise, I compute inflation rates by product as follows. For every product-location-month triplet, I calculate the price-change over the previous month. For each product, I compute the mean of these changes and use it as the average inflation rate for the product. Table 1.9 and Table 1.10 present respectively the inflation rates per product for non-food and food products. Two observations emerge form these tables. First, prices rise faster for food products than non-food products. Second, inflation rates are more heterogeneous for food products than non-food products.

Table 1.9: ***Inflation heterogeneity for non-food products***

Product	Inflation rate
Batteries-Eveready Drycell No-40gm	1.49%
Cigarettes-Rothmans-20packs	1.66%
Detergents-Zahira	1.50%
Electric Bulb-Philips 40/60 Watt No-20 Gm	7.83%
Hard Soap (imported)-200 Gm	1.90%
Pen Ball Point-Bic England No-20 Gm	1.62%
Beer-Bedele-300 Cc	1.18%
Beer-Harar-330 Cc	1.38%
Beer-Meta Abo-330 Cc	1.29%
Hair oil-Zenith Non-Liquid Form-330 Cc	1.45%
Motor Oil- Mobil Lt-1.2kg	1.02%
Soft Drinks-Coca Cola -300 Cc	1.54%
Soft Drinks-Pepsi Cola Mirinda-300 Cc	1.55%
Water Bottled-Ambo Mineral Water-500 Cc	1.75%
Wine-Saris Normal-750 Cc	1.67%

Notes: this table reports the average inflation rates by product for non-food products. For imported products, I consider Kombolcha as the port of entry in the country because it is the main city on the corridor to djibouti.

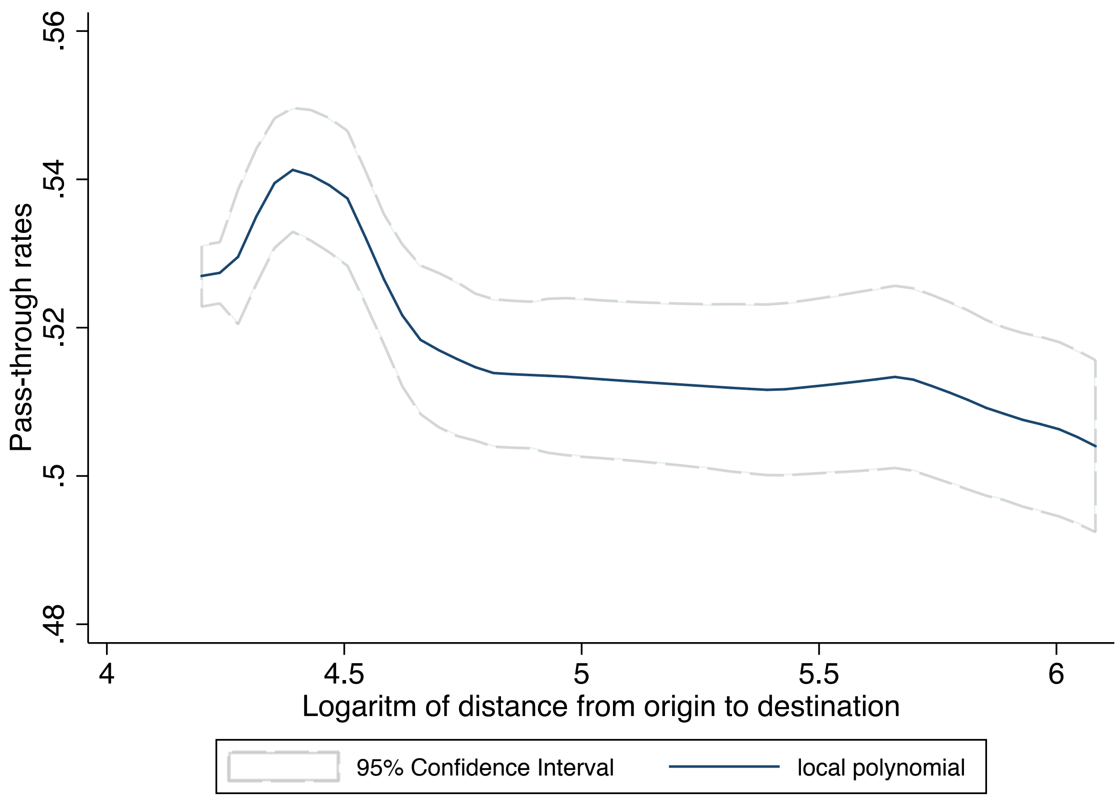
Table 1.10: ***Inflation heterogeneity for food products***

Product	Inflation rate	Product	Inflation rate	Product	Inflation rate
'Bula'	4.12%	Cooking Oil (Local)	1.93%	Peas Green(dry)	3.21%
'Dabo' (Traditional Ambasha)	2.59%	Cow Milk (Unpasteurized)	2.33%	Peas Milled	2.57%
'Dube' Flour	5.47%	Cow Milk (pasteurized)	1.85%	Peas Mixed	3.33%
'Enjera' ('Teff' Mixed)	1.79%	Dry Yeast(Baking powder)	1.79%	Peas Split	2.43%
'Fafa' Flour	2.39%	Durrah	2.99%	Peas Split(Roasted)	2.31%
'Furno Duket' Locally Processed	1.58%	Ethiopian Kale	9.95%	Peas White	3.30%
'Kocho' (Unprocessed)	5.45%	Fenugreek Milled	1.98%	Pepper Green	7.22%
'Teff' Black (Red)	2.60%	Fenugreek(Green)	2.12%	Pepper Whole	5.30%
'Teff' Mixed	2.13%	Fish Fresh	3.70%	Potato	3.83%
'Teff' White	2.09%	Garlics	5.19%	Powdered Milk (Me&My)	2.57%
African Millet	2.54%	Ginger Dry(Local)	3.58%	Pumpkin	9.62%
Avocado	5.39%	Ginger Wet(Local)	6.09%	Rape Seed	4.32%
Banana	4.02%	Grapes	2.85%	Rice (Imported)	0.78%
Barley Black	2.49%	Green Peas	4.76%	Salt	1.54%
Barley Mixed	2.54%	Ground Nut Shelled	2.60%	Sardines (Imported)	2.02%
Barley White	2.18%	Haricot Beans	3.27%	Sesame Seed Red	2.76%
Barley for Beer	1.72%	Honey	2.30%	Sorghum	3.65%
Basil Dry	4.09%	Horse Beans	3.52%	Sorghum Red	3.49%
Beet Root	5.17%	Horse Beans Milled	2.75%	Sorghum White	3.02%
Biscuits	1.50%	Horse Beans Split(Roasted)	2.67%	Sorghum Yellow	1.98%
Black Cumin(Local)	2.32%	Hulled Barley	3.11%	Soya Beans	2.69%
Black Pepper(Local)	6.20%	Leaks	5.08%	Spaghetti (Local) Without Egg	1.30%
Bread Wheat (Bakery)	1.42%	Lemon	8.59%	Spinach	6.58%
Butter Unrefined	2.81%	Lentils	2.67%	Sugar	0.59%
Cabbage	6.59%	Lentils Split	2.34%	Sunflower	3.33%
Cactus	8.96%	Lettuce	9.53%	Sweet Potato	7.60%
Camel Milk	3.75%	Lima Beans	4.09%	Tangerine	3.24%
Canned Tomato (Local)	2.77%	Linseed Red	4.05%	Tomatoes	6.91%
Cardamon(Local)	2.10%	Linseed White	3.23%	Tumeric Flour(Local)	2.34%
Carrot	6.17%	Long Pepper(Local)	4.25%	Vegetable Butter(Sheno & Shady)	1.70%
Castor Beans	5.07%	Macaroni (Local) Without Egg	1.17%	Vetch	3.78%
Cauliflower	2.31%	Maize (White)	2.60%	Vetch Milled	3.13%
Cheese Cottage	4.11%	Mango	4.07%	Vetch Split(Roasted)	3.03%
Chick Peas	2.34%	Mixed Pulses Milled	2.81%	Wheat Black (Red)	2.62%
Chick Peas Milled	2.34%	Niger Seed	4.02%	Wheat Mixed	2.72%
Chick Peas Split(Roasted)	2.16%	Oats	2.29%	Wheat White	2.15%
Chillies Whole	6.63%	Onions	4.42%	White Cumin(Bishop's Weed)Local	4.59%
Cinnamon(Imported)	2.36%	Orange	4.50%	Yoghurt (Traditional)	2.13%
Cloves(Imported)	6.88%	Papaya	6.09%		
Cooking Oil (Imported)	1.34%	Pastini	1.36%		

Notes: this table reports the average inflation by product for food products.

1.6.2 Additional figure

Figure 1.5: ***Estimated Pass-through rates and distance***



Notes: This figure presents a local polynomial regression of pass-through rates on distance (Log) using non-food products dataset. I used an Epanechnikov kernel and a bandwidth of 0.5

CHAPTER II

TRADE COSTS, PRICES, AND CONNECTIVITY IN RWANDA

Abstract

We combine high-frequency market price data and the road network to estimate the size of internal trade costs in Rwanda and evaluate whether better connectivity reduces these costs. We find that internal trade costs in Rwanda, defined as a function of distance are ten times higher than in the U.S., three times higher than in Ethiopia, and two times higher than in Nigeria. Holding Rwanda pass-through rate at these countries levels would have reduced the size of internal trade costs by half in Rwanda. We also find that these costs are particularly high for manufactured products. We test two mechanisms related to the intensity of competition, and the accessibility to road infrastructure. We show that intermediaries are charging higher markups in locations where there is less competition. We also show that better road connectivity through access to feeder and national roads matter as connecting a location to a feeder road reduces trade costs by approximately two cents per mile.

Keywords: Trade costs; Price dispersion; Feeder roads; Rwanda.

JEL classification: O12; R12; R23.

2.1 Introduction

Despite a significant reduction in barriers to international trade over the past decade all over the world, developing countries are still characterized by many non-tariff barriers that affect trade flows and reduce potential regional integration benefits. In many developing countries, the movement of goods and people are affected by many additional challenges: old truck fleets, terrible road conditions that necessitate frequent truck repairs, poor logistics, high fuel costs, long waiting times for loading and unloading frequent checkpoints often accompanied by bribe demands. These infrastructure and non-tariff barriers add to an already high transportation costs toll leading to inflated prices of domestically manufactured products, as most raw materials used for manufacturing are imported. The situation is exacerbated in rural areas where the population suffers from a lack of adequate transport infrastructure due to the high costs of this type of investment and political issues (Berg et al., 2017). Rural areas are not all integrated to the rest of the country, and this geographic isolation might have disastrous economic consequences. High trade costs isolate lagging regions from the rest of the country. As a consequence, consumers in these areas cannot take advantage of competitively priced goods from big cities and abroad, and their firms cannot access high-quality inputs or sell to unaccessible urban markets. One way to reduce the cost of isolation is to reduce the infrastructure gap in rural areas. Over the past decades, multilateral development agencies (MDB) are increasingly financing transport infrastructure projects in developing countries. In 2018 for example, the World Bank Group approved 3.44 billion and the African Development Bank disbursed 1.9 billion dollars for transport infrastructure projects in Africa, including the rehabilitation or the construction of rural feeder roads (AFDB, 2018).

Rwanda's landlocked country status, coupled with its rugged terrain geography, and the absence of a significant industrial rail system, represent a major constraint to national, regional, and international trade. The country lacks a link to regional railway networks, which means all trade is conducted by road or by air. Most of the country's imports and exports are shipped by road from the ports of Mombasa (Kenya) and

Dar-es-Salaam (Tanzania) which are up to 1,500 km from the Kigali, the country main city. There are large internal trade costs associated with remoteness, and these costs are magnified by the poor state of road infrastructure. However, very little is known about the size of internal trade costs, especially in the context of a landlocked country like Rwanda. How large are internal trade costs in Rwanda? What are the effects of better road connectivity on these costs? The government of Rwanda (GoR) with the support of four donor agencies (World Bank, USAID, EU, and the Netherlands), has designed an ambitious feeder roads program which aims to enhance connectivity to rural markets and reduce transport costs for people and goods. Given the importance of the funds invested, policymakers, donors, and researchers are naturally questioning the expected wider economic benefits of such investments. In particular, they are interested in understanding the effects of these investments in reducing the size of trade costs in Rwanda.

To address these questions, this paper uses micro-geographic market price data and the Rwandan existing road network. Specifically, we estimate the size of internal trade costs and evaluate whether better road connectivity can reduce the size of these costs. Previewing our key results, we find that internal trade costs are particularly high in Rwanda compared to other landlocked African countries like Ethiopia. Exploiting the information on a feeder roads program and localities connection to the existing road network, we find that markets connected to feeder roads experience lower trade costs, especially when feeder roads allow access to national roads. Therefore, a simultaneous connection of markets to both feeder roads and national roads has the potential to substantially reduce the size of internal trade costs.

In the first part of our analysis, we rely on the methodology proposed by Atkin and Donaldson (2015) to estimate the size of internal trade costs. The basic idea is to use spatial dispersion of prices purged from spatial variation of markups to estimate internal trade costs. As mentioned by Atkin and Donaldson (2015), this exercise involves three major difficulties. First, one needs to collect prices of identical products to ensure that spatial price differences do not reflect differences in product quality. To fulfill this requirement, we use an ongoing high-frequency market price survey to select narrowly defined products for which type, quantity, and units are explicitly specified. Second, one needs to identify the origin locations of products to ensure that spatial price differences correctly identify trade costs. Combining the information on the source of products included in the price and the market listing surveys allowed us to retrieve the origin locations of products. Third, one needs to correct spatial differences for the markups charged by intermediaries. We fulfill this requirement by first estimating the effect of a change in origin prices on destination prices, and then using this effect to correct price differences. After overcoming these three challenges, we find that internal trade costs increase with distance and are high in Rwanda. They are approximately ten times higher than in the U.S., three times higher than in Ethiopia and two times higher than in Nigeria. More specifically, an intermediary who trades goods 10 miles away from the farm gate, pays an additional trade cost of 25 cents in Rwanda, compare at 9 cents in Ethiopia, 13 cents in Nigeria, and only 2.5 cents in the U.S. [1] We also show that intermediaries are charging higher markups in locations where there is less competition (measured by the number of traders in a market). We also find that internal trade costs increase more with distance when we consider manufactured products or Food. All these results are robust to alternative measures of distance.

In the second part of our analysis, we investigate the effects of improved road connectivity on the size of internal trade costs. We start by capturing the evolution of some key economic variables related to internal trade costs for connected and non-connected markets. We find that better road connectivity through feeder and national roads is positively correlated to competition intensity and trade concentration in local markets. This finding might have some implications in terms of internal trade costs. The idea here is that with the advent of new rural roads, there will be increased competition. This intensification of competition could affect the size of trade costs by reducing markups. We then try to explore this correlation by assessing whether markets that are connected to feeder roads are experiencing reduces trade costs. We find that feeder roads reduce the size of internal trade costs defined as a function of distance, especially when they allow

1. Results for Ethiopia, Nigeria and the U.S. are estimated by Atkin and Donaldson (2015).

accessibility to primary or national roads. On average, connecting a location to a feeder road reduces trade costs by approximately 2 cents per mile. This effect transits through the connection of this feeder road to a national road. Finally, to complement our analysis, we construct three variables to capture different levels of road connectivity with respect to the proximity of the markets with feeder and national roads. These levels of road connectivity are named as follows: low connectivity, intermediate connectivity, and high connectivity. What stands out here is that, the better the road connectivity to markets, the lower the internal trade costs. Overall, our results suggest that the combination of feeder roads and national roads leads to a substantial reduction of internal trade costs measured as a function of distance.

Our work contributes to the literature in several ways. First, we contribute to the emerging studies which try to estimate internal trade costs in developing countries (e.g., Porteous, 2019; Sotelo, 2020). Atkin and Donaldson (2015) use price differences of manufactured goods across cities to measure the size of trade costs in Ethiopia, Nigeria, and the U.S. In the same vein, de Rochambeau (2018a) proposes an estimation of trade costs in Liberia using spatial dispersion of 13 products across 22 urban markets. We complement these works by using prices of a larger set of products including manufactured and agricultural products. In addition, we focus on spatial price dispersion across rural markets. Second, we also contribute to the growing literature estimating the impact of large transport infrastructure investments. Redding and Turner (2015) proposes a qualitative review of this growing literature whereas Roberts et al. (2020) provide a quantitative review. This literature takes two main directions. The majority of studies estimate a reduced-form relationship between transport infrastructure investments and economic outcomes (e.g., Faber, 2014; Baum-Snow et al., 2017; Berger and Enflo, 2017). A recent strand of this literature estimates the impact of transport infrastructure investments using a general equilibrium framework (e.g., Donaldson and Hornbeck, 2016; Donaldson, 2018; Allen and Arkolakis, 2019). Overall, this literature finds a positive but variable effects of transportation infrastructure on economic outcomes. The studies often differ in the features of transport infrastructure. These features includes, the mode of transportation (rails, roads, waterways), the type of connection (intra-city, inter-city, urban-rural), and the type of construction (new system or rehabilitation). The majority of this literature pays attention to transport infrastructure which connects big urban markets. We complement their works by evaluating the effect of increased connectivity through access to both feeder and national roads on the size of internal trade costs.

The remainder of the paper is organized as follows. Section 2.2 presents our data and descriptive statistics. We describe the methodology used to estimate internal trade costs in section 2.3. Section 2.4 presents the baseline results on the estimation of the size of internal trade costs. Section 2.5 investigates the effects of road connectivity on internal trade costs. Section 2.6 concludes.

2.2 Data and Descriptives

This section presents our price data as well as the economic and geographic variables used throughout our empirical analysis.

2.2.1 **Price Data**

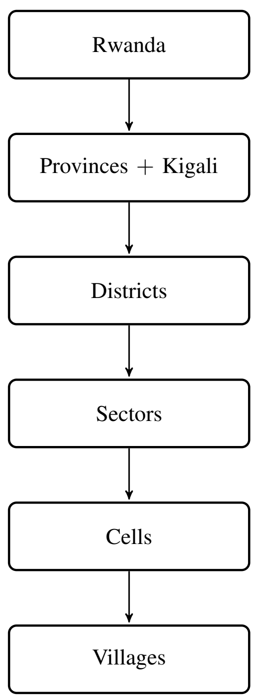
Source and data collection. The prices data used in this paper were collected by the Government of Rwanda in collaboration with the World Bank to evaluate the impact of a large feeder road rehabilitation program. Each month, enumerators visit each market, ask about different products, record product availability as well as prices.[2] Data collection started in November 2016 and ended in July 2018. It covers 139 markets located in all the parts of Rwanda excepted in the capital city Kigali. These markets are rural and located in cells that are the smallest administrative division in Rwanda after villages (see Figure 2.1).
Figure 2.2 maps the markets included in this study as well as the road network to which

2. The Enumerators visit a village twice and stay a few days for data collection. It is not necessarily the same enumerator who visits the village.

they are connected. A market here is a village located in a district. In this village, there are retail spaces. The businesses are of various types: stalls, permanent display cases, temporary kiosks, and often even just covers on the ground on which the products are displayed. The businesses are either open air or closed and products are generally available at the same time. The majority of the market are selected based on their proximity to a feeder roads or a national roads. The criterion of wide geographic coverage was also considered. So we end up with steps connected to the feeder road or not, steps connected to the national road or not, and markets connected neither to a feeder road, nor to a national road.

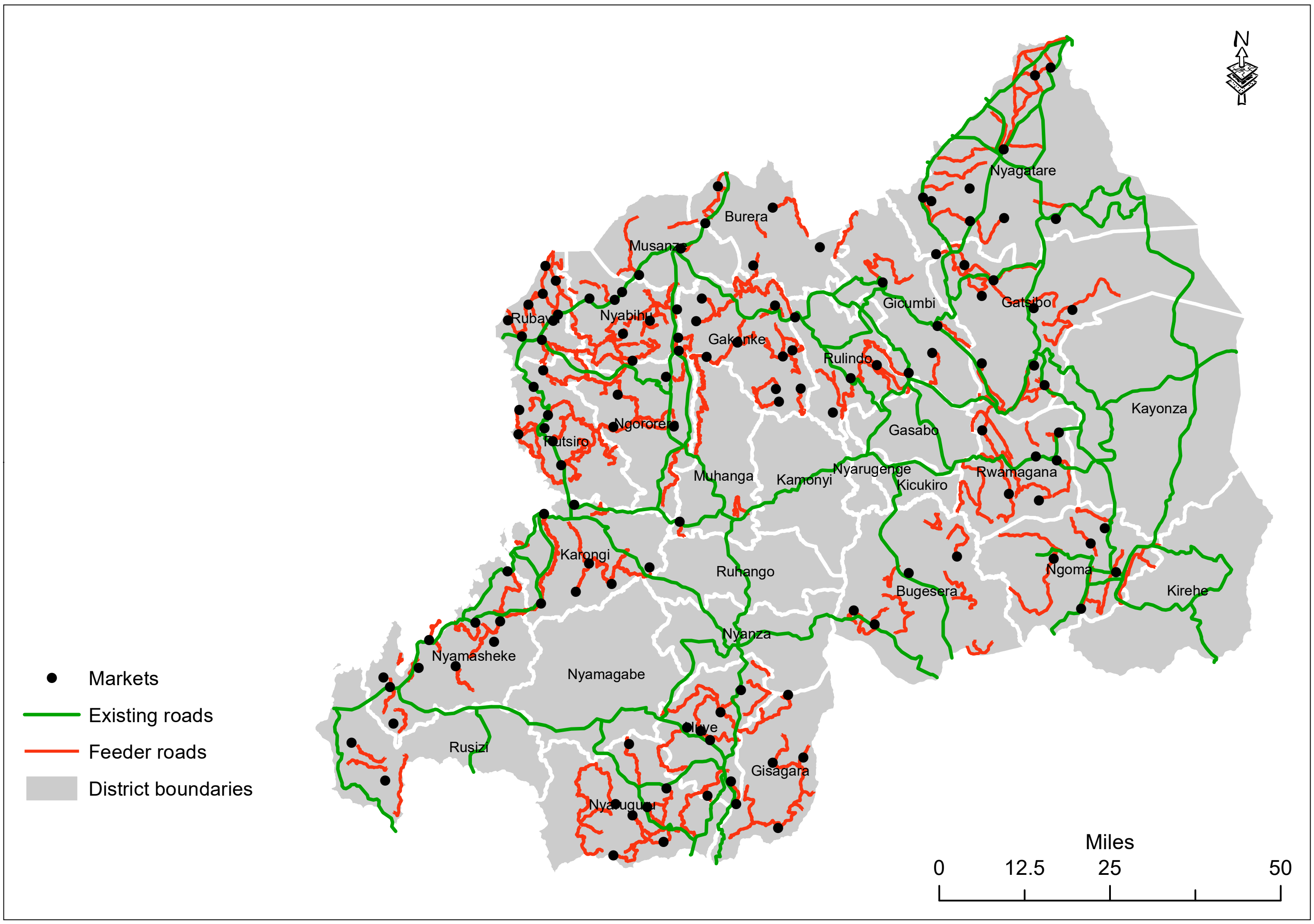
Table 2.1 shows the spatial coverage of our price data. These data cover 70% of districts and 30% of sectors in Rwanda.

Figure 2.1: ***Administrative structure in Rwanda***



Notes: This figure presents the administrative structure in Rwanda. Rwanda is divided into 4 provinces and the capital city Kigali, 30 districts, 416 Sectors, 2148 cells, and 14,837 villages.

Figure 2.2: ***Markets and roads in Rwanda***



Notes: This figure maps the 139 markets covered by our price data. The road network is equally depicted. It includes the existing national roads (in green) and the new feeder roads rehabilitated by the Government of Rwanda (in red).

Table 2.1: ***Commodity coverage and spatial coverage of price data***

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Province	# months	# districts	# sectors	# cells	# markets	# products	# categories
East	22	5	34	37	38	51	6
North	22	5	23	24	25	51	6
South	22	4	21	22	22	51	6
West	22	7	45	50	54	51	6
Total	22	21	123	133	139	51	6

Notes: This table shows the spatial and commodity coverage of our price data. Column (1) enumerates the provinces; Column (2) shows the number of months cover in our data. Columns (3) to (6) show respectively the numbers of districts, sectors, cells, and markets covers in each province. Column (7) presents the number of products; Column (8) presents the number of product categories. Except the capital Kigali, all the provinces of Rwanda are covered in our price data.

Description. The data collection expanded to a basket of 68 goods frequently consumed in Rwanda. In particular, this basket represents 45% of total household consumption. As we are interested in estimating trade costs using spatial price differences, we selected products in this basket that were narrowly defined (quantity and units to which the price refers). For instance, we excluded products like meat or fish because they were not narrowly defined. Selected products include for example "Maggi Bouillon Cube-12 Gm" or "Soft drink (Fanta)-30 Cl". We end up with 51 products ranged in 6 categories including fruits; vegetables; household items; staple food; personal care; processed Food and drinks. Our product sample which is very diversified includes 48% of manufactured products and 52% of non-manufactured products. Table 2.2 describes in detail the list of products and categories selected.

Table 2.2: ***Product and category List***

Fruit	**Household Item**	**Personal Care**
Banana-1 Kg	Airtime Cell Phone Card-1 Piece	Hair Dye Powder-4.5 Gm
Mandarine-1 Kg	Candle-1 Piece	Sanitary Napkins-10 Pieces
	Charcoal-1 Kg	Shampoo-1 L
	Detergent-600 Gm	Soap-80 Gm
	Mattress-1 Piece	Toilet Paper-1 Piece
	Nails-1 Kg	Toothpaste-70 Gm
	Suitcase-1 Piece	
Processed Food and Drinks	**Staple Food**	**Vegetable**
Brochette-1 Piece	Cassava Flour- 1 Kg	Amarante-1 Bunch
Commercial Beer (local)-72 Cl	Cassava leaves-1 Kg	Cabbage-1 Kg
Maggi Bouillon Cube-12 Gm	Cooking Banana-1 Kg	Carrots-1 Kg
Mineral Water-1.5 L	Cooking Oil-1 L	Cassava Root- 1 Kg
Softdrink (Fanta)-30 Cl	Dry Beans-1 Kg	Celery-1 Kg
Spaghetti-500 Gm	Egg-1 Piece	Garlic-1 Kg
Tea local-10 Gm	Green Pea-1 Kg	Irish Potato-1 Kg
	Groudnut Flour-1 Kg	Leek-1 Bunch
	Groundnut-1 Kg	Onion-1 Kg
	Imported Rice-1 Kg	Pepper-1 Kg
	Maize-1 Kg	Sweet Potato-1 Bunch
	Maize Flour-1 Kg	Tomato-1 Kg
	Milk Fresh-1 L	
	Sorghum-1 Kg	
	Sorghum Flour-1 Kg	
	String Bean-1 Kg	
	Sugar-1 Kg	

Notes: This table shows the list of products we considered in this paper as well as their categories. The description of each product includes the name, the unit, or the weight.

Outliers and availability. A graphical check suggests that there are some outliers in the dataset. For some products, prices might be abnormally high because they do not refer to the same units. Prices are cleaned by dropping monthly observations for which the log-price is five times higher or five times lower than the median. With this procedure, we drop only 0.41% of our observations. Turning to the information about availability of products, it is recorded by enumerators after each visit. In our final dataset, a product is considered unavailable for about 21% of the product-location pairs. The shortage for these 21% product-location pairs is permanent as the enumerator comes twice a month and spends some time in the village.

Correction for exchange rates. Initial price data were in Rwandan francs. For comparison purposes, we converted these prices in U.S. dollars using the U.S. dollars-Rwandan francs exchange rate prevailing in November 2016. [3]

Trading markets pairs

One important challenge to overcome when estimating trade costs using spatial price gaps is to identify markets that are effectively trading. The price difference between two markets is informative for trade costs only when there is trade between these markets (Atkin and Donaldson, 2015). There is trade between two markets when one of them is the origin location of the product and the other, the destination.

We define a trading market pair as a pair of markets for which one market is the location where the intermediary sells the product (destination) and the other market is a **different location** where the product come from (origin). A match between the price survey and the market listing survey allows us to identify origin locations. In the price survey, traders were asked where their products come from. We classify products into four possible origin locations: in the village; inside district; outside district, and outside Rwanda (for imported goods). [4]

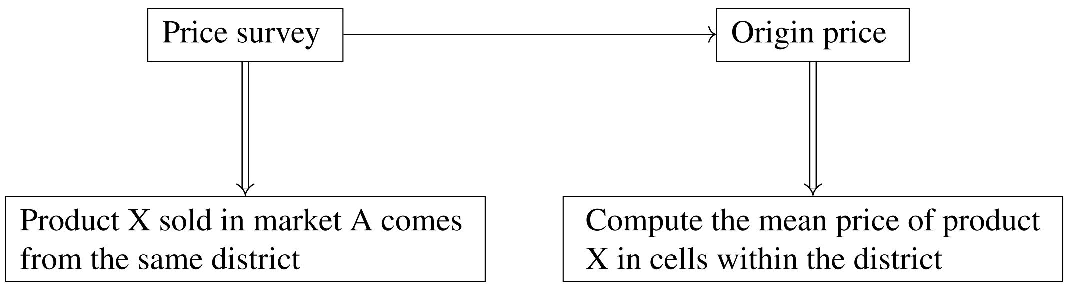
In the village. In a given market, if a product originating from the same village, origin price equal to destination price. we dropped all market pairs for which the origin location were the same village because they did not correspond to our definition of trading markets pairs.

3. November 2016 is the first month of price data collection. 1 U.S. dollar was equal to 806.25 Rwandan francs on November 1st, 2016.

4. In our data in 23% of the cases, the products comes from the same village, 21% from the same district, 19% from another district, 14% from outside Rwanda. For 23% of the cases, the information on origin is unknown.

Inside district. Figure 2.4 describes the process we used to identify the origin of a given product coming from the same district. In some cases, the only information we have about the origin of a product sold on a market is that, it comes from somewhere in the same district in which the market is located, but not in the same village. In these cases, we take as origin price, the average price in the component cells of the district. [5] We consider the centroid of the district as the origin location of the product.

Figure 2.3: ***Retrieving the origin of an "inside district" product***



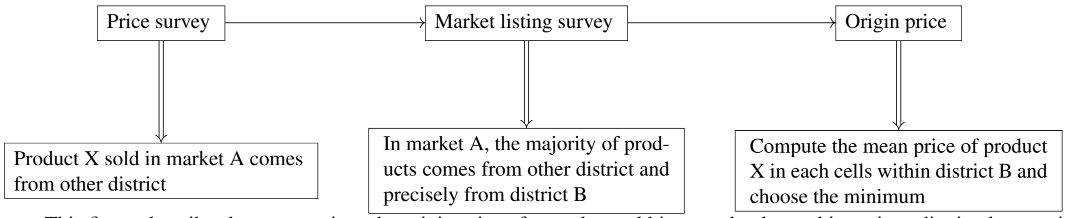
Notes: This figure describes how we retrieve the origin price of a product sold in a market located in a given district and originating from the same district.

Outside district. Figure 2.4 describes the process we used to identify the origin of a given product coming from another district. For products with an origin outside the district, the price survey is not sufficient to identify precisely the corresponding district. Therefore, we use information from the market listing survey to identify this district. In the market listing survey, the following question is addressed in each market: Where do the majority of products sold in this market come from? The possible answers are the same village; same cell; same district; same sector; Kigali; another district; another country; other sources. Once this general source is known, and provide the name of the district or the name of the country, another question is how to identify the exact source location. To retrieve it, after the district is identified, we compute the mean price in every cell within this district and choose the minimum price as the origin price. Figure 2.5 shows an example of the path followed by a product sold in a market and originating from outside the district where the market is located. More specifically, it pictures the

5. The results are not sensitive when we consider the median price instead of the average price.

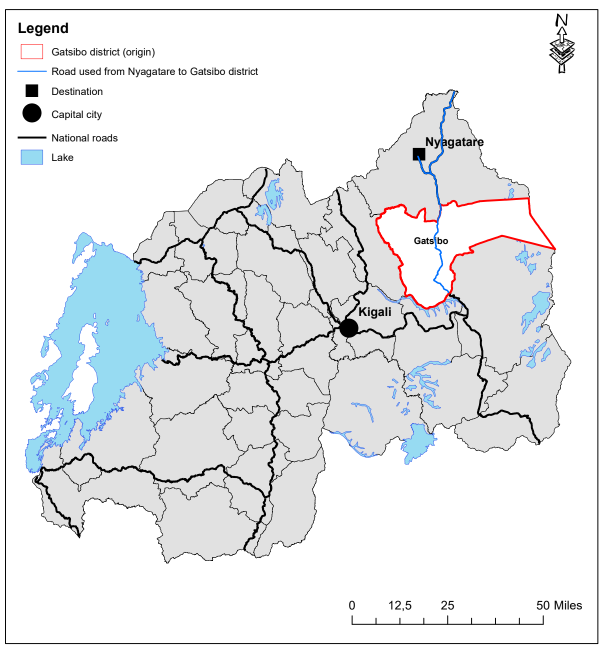
path followed by sugar sold in Nyagatare (the destination market) and coming from another district (Gatsibo).

Figure 2.4: ***Retrieving the origin of an "outside district" product***



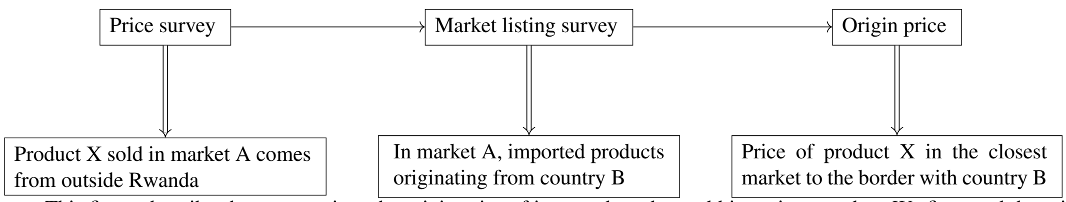
Notes: This figure describes how we retrieve the origin price of a product sold in a market located in a given district, but coming from another district. The price survey only mentions that the product comes from outside district without other indications. The market listing survey gives for each market, the information about the precise district source of products coming from outside district.

Figure 2.5: ***Distribution network of sugar sold in Nyagatare market***



Outside Rwanda Figure 2.6 presents the process we used to identify the origin of a given product coming from abroad. As can be seen, when a product is imported, the price survey only mentions that the product comes from another country without other indications. To complement this information, we use the market listing survey. In this survey, we know exactly in each market, the origin country of imported products. Once the country identified, we choose as origin price, the price in the closest market to the border with this country. Due to the similarity of prices in border markets or markets not far from the border, our strategy here does not affect the results.. [6]

Figure 2.6: ***Retrieving the origin of an imported product***



Notes: This figure describes how we retrieve the origin price of imported product sold in a given market. We first used the price survey, and then used the market listing survey.

Imported products in our dataset come from four countries including: Burundi, Democratic Republic of Congo (DRC), Ouganda, and Tanzania. For these products, we choose as origin the closest market to the border in our dataset.

Burundi border: The major border between Burundi and Rwanda is Fugi. However, since Fugi is not in our market sample, we had to choose the closest market to Fugi as the origin of imported products coming from Burundi. In our sample, that market was Viro.

Democratic Republic of Congo (DRC) border: There are two major border towns between Uganda and Rwanda: Gisenyi and Cyangugu. Gisenyi is present in our market sample market while Cyangugu is not. We replaced Cyangugu by Bambiro which is the closest market to Cyangugu in our dataset. We end up with two possible origins for products imported from DRC: Gisenyi and Bambiro. If an imported product is sold on

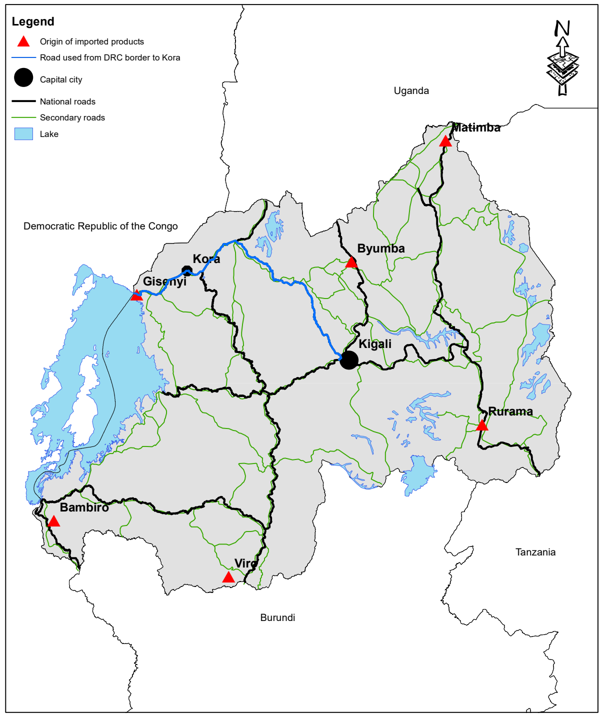
6. Foreign products in our database arrive from border countries by road. This is why we consider the origin price for a given product coming from abroad to be the price in the village closest to the border.

a market and the product comes from the DRC, we define Gisenyi or Bambiro as origin location depending on the closest origin to this market.

Uganda border: There are two major border towns between Uganda and Rwanda: Gatuna (the most important) and Kagitumba. However, Gatuna and Kagitumba are not present in our market sample. Then, we chose respectively Byumba and Matimba as origins for imported products coming from Uganda. We end up with two possible origins for products imported from Uganda: Byumba and Matimba. If an imported product sold on a market comes from Uganda, we choose as origin Byumba or Matimba, depending on the closest origin to this market.

Tanzania border: The major border between Tanzania and Rwanda is Rusumo. However Rusomo is not in our markets sample, hence we chose Rurama (the closest market to Rusumo) as the origin of imported products coming from Tanzania. The origins of the imported market are mapped in Figure 2.7. This figure presents also the path followed by an imported toothpaste (from DRC) sold in Kora.

Figure 2.7: ***Road network of imported toothpaste sold in Kora market***



Geographic Variables

We complement our dataset with geographic variables. To estimate the elasticity of trade costs relative to the distance we need distance measures. Moreover, we need information on road infrastructure in order to examine their impacts on internal trade costs.

Distance and travel time. To assess how far a market is from another, we use the geodesic distance in miles computed by the Stata package Geodist. For robustness checks, we consider three other measures of distance. We first compute the great circle distance using the Stata package Geodist. We then use the Stata package Georoute to compute travel times and travel distances (Weber et al., 2016). Whereas travel distance measures the distance by car between two locations, travel time is defined as the time in minutes one should drive by car from one market to another during normal traffic conditions.

Road infrastructure. Figure 2.2 maps all the road infrastructure we consider in this study. We use two categories of road infrastructure in our analysis: the new feeder roads rehabilitated and improved by the government of Rwanda with the World Bank and other donors support; and the existing national roads. A feeder road is a rural road whereas a national road is a road that connects to urban centers. The market listing survey contains information about all feeder roads segments. A map of all feeder road segments was overlaid with a map of all market locations, and all markets within a 1-kilometer buffer of the sampled feeder roads. All markets within a 1-kilometer buffer around the national roads were also sampled.

Economic Data

To take into account the market structure through our empirical analysis, we computed two measures. First, a measure of competition intensity and second, a measure of trade concentration.

Competition intensity in a market. In the price survey, there is information about the numbers of traders per market. We measure the intensity of competition in a market by the log of the number of traders. We also use the log of the number of traders in the district where the market is located.

Trade concentration index. Using the data on the number of traders by market, we calculated for each market, the share of traders in the total traders of the sector where the market is located. Our measure of trade concentration is the inverse of the Herfindahl index computed with these shares.[7] Equation 2.1 presents our trade concentration index.

$$H_i = \frac{1}{\sum_{k=1}^{n} \left(share_k\right)^2} \tag{2.1}$$

In equation 2.1, H_i is the trade concentration index of sector i. $share_k$ is the share of traders in market k, in the total of traders in sector i where market k is located.

2.2.2 Descriptive Statistics

Our final datasets collect information on the prices of 51 products in 139 markets as well as the price differences between these markets. In addition, they provide geographic and economic characteristics of these markets (distances, travel times, road infrastructure, competition intensity, trade concentration). More specifically, we end up with two different datasets. The first only consider trading pair markets, that is market pairs for which one market is the origin location of the product while the other market is the destination. The second dataset considers all the possible pairs of markets.

Table 2.3 displays some descriptive statistics on our final datasets. Several features are worth mentioning here. First, when considering only the trading market pairs, the average price gap is lower. As can be seen, the average price gap between all market pairs is 0.288 dollar while it is only 0.08 dollar between trading market pairs. Second, Rwanda is a small country relative to the other African countries studied in the literature (Ethiopia, Liberia, Nigeria) with an average distance of 18 miles between trading markets compared at 350 miles in Nigeria, 90 miles in Liberia, and 221 miles in Ethiopia (see Atkin and Donaldson, 2015; de Rochambeau, 2018a). However, in terms of prices, the average price gap in Rwanda is almost in the same range as in these countries. The average price gap in Rwanda is 0.08 dollars compared to 0.07 in Liberia,

7. While the competition intensity index expresses more of a measure extensive, The trade concentration index is interpreted as an intensive measure.

0.062 in Nigeria, and 0.025 in Ethiopia.

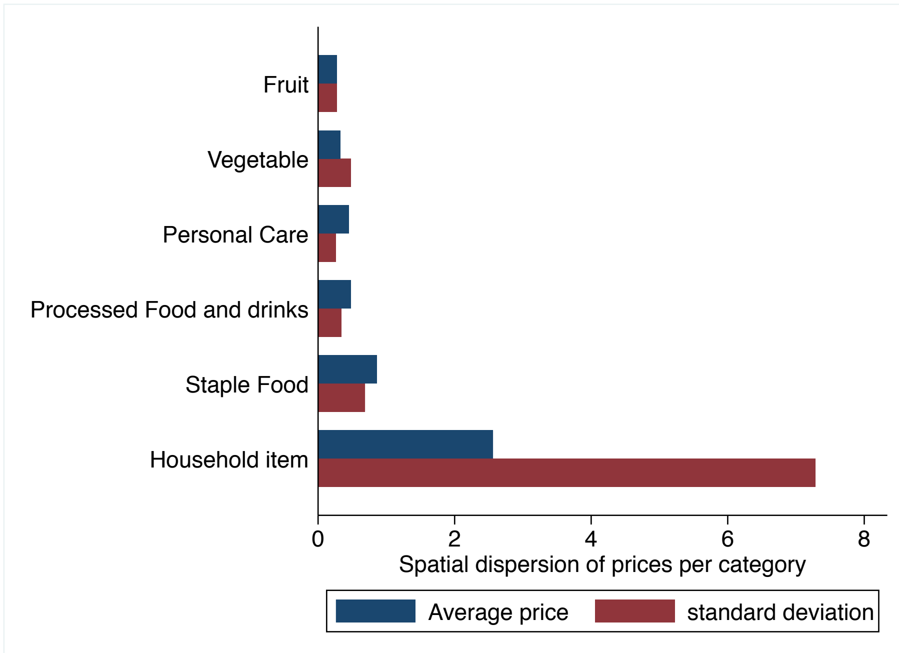
Figure 2.8 presents the distribution of prices by categories. It shows that products in the Household item category are more expensive and the most dispersed. However, the fruits are cheaper and the least dispersed.

Table 2.3: ***Summary Statistics***

	Rwanda
Origin Price	0.914
	(3.077)
Price Gap (Trading market pairs)	0.088
	(1.570)
Absolute Price Gap (All market pairs)	0.288
	(1.458)
Distance to origin location (Trading market pairs)	18.377
	(18.827)
Distance between all locations (All market pairs)	52.346
	(26.361)
Log Distance to origin location (Trading market pairs)	2.491
	(0.904)
Time to drive from origin location to destination (Trading market pairs)	56.452
	(50.422)
Number of traders in markets	459.085
	(1627.072)
Observations (Trading market pairs)	48,678
Observations (All market pairs)	9,598,870

Notes: This table presents the means and standard deviations (in parentheses) of the key variables used in this study. The first row reports the price at the origin location of the products. The second row reports the price gaps only using data from trading market pairs. The third row reports absolute price gaps using data from all locations pairs. The fourth row reports the distances in miles between trading market pairs. The fifth row reports distances in miles between all location pairs. The sixth row reports log-distance in log-miles between trading market pairs. The seventh row reports the time in minutes to drive between trading market pairs. The eighth row reports the number or traders. Prices are converted in u.s. dollars using the prevailing exchange rate in November 2016.

Figure 2.8: ***Raw distribution of prices***



Notes. This figure presents the spatial distribution of prices per category of products.

2.3 Methodology

This section describes the methodology used in this paper. We first develop the model, then we present the empirical strategy.

2.3.1 Model of intermediated internal trade

The estimation of internal trade costs in this paper relies on the model of intermediated internal trade developed by Atkin and Donaldson (2015). We only present here the intuition and the main assumptions/equations of the model. The full framework, as well as the proofs, can be found in the initial paper.

The model idea The environment of the model considers multiple products j potentially sold in multiple markets d at multiple periods t. Each good j can be domestically produced at a factory location o , or imported into the country through a border o. Firms or intermediaries purchase goods from the origin o, transport these goods to the destination d and sell them to consumers in this location. The price at the origin is denoted by P^j_{ot} and the price in location d is P^j_{dt}. Equation 2.1 is the basic equation to infer internal trade costs from spatial price differences.

$$P^j_{dt} - P^j_{ot} = \tau(X^j_{odt}) + \nu^j_{dt} \tag{2.1}$$

In this equation, ν^j_{dt} is the markup charged by intermediaries in location d and time t when selling product j. $\tau(X^j_{odt})$ is the trade costs incurred by the intermediary when buying product j in location o and selling it in location d. X^j_{odt} represents the set of cost-shifters. The main cost-shifter is the distance between the origin o and the destination d. Equation 2.1 implies that to infer trade costs from price differentials, it is important to identify precisely the pairs of markets which actually trade by recovering the locations where intermediaries bought their products. Moreover, it is crucial to purge price gaps from markups such that these price differences only identify trade costs.

Main assumptions To control for the markups charged by firms or intermediaries, Atkin and Donaldson (2015) make four assumptions about the structure of the market.

Assumption 1 : An intermediary selling q^j_d unit of product j in location d faces fixed costs F^j_d and is subject to the following total costs:

$$C(q^j_{od}) = \left(P^j_o + \tau(X^j_{od})\right) q^j_d + F^j_d \tag{2.2}$$

Assumption 2 : In market d, there are m_d identical intermediaries selling the product j and choosing the quantity q_d which maximizes their profits given the competitiveness in the market d.

Assumption 3 : Consumer preferences take the form of constant pass-through inverse demand as developed by Bulow and Pfleiderer (1983).

Assumption 4 : The pass-through rate ρ^{j}_{odt} measures the effect on the destination price of a shock on the origin price. ρ^{j}_{odt} is fixed across time within a product-origin-destination triplet.

That is $\rho^{j}_{odt} = \rho^{j}_{od}$

Fundamental equation to infer trade costs. Combining all the previous assumptions leads to the following equation:

$$P^{j}_{dt} - P^{j}_{ot} = \rho^{j}_{od}\tau(X^{j}_{od}) + (1-\rho^{j}_{od})(b^{j}_{d} - P^{j}_{ot}) \tag{2.3}$$

Equation 2.3 shows that the pass-through ρ^{j}_{od} and the level of the inverse demand b^{j}_{d} are sufficient to control for markups, competitiveness and preferences. This theoretical equation is fundamental to infer trade costs from spatial price differences.

2.3.2 Empirical strategy to estimate internal trade costs

The empirical strategy to estimate trade costs consists of two steps. In the first step, we estimate pass-through rates ρ^{j}_{od}; and then we use these estimates to infer the effect of a change in distance on trade costs.

Step 1: recover estimates $\widehat{\rho^{j}_{od}}$ of pass-through rates ρ^{j}_{od}. Combining equation 2.3 with the assumptions of the model leads to:

$$P^{j}_{dt} = \rho^{j}_{od}P^{j}_{ot} + \rho^{j}_{od}\tau(X^{j}_{od}) + (1-\rho^{j}_{od})(b^{j}_{d}) \tag{2.4}$$

Applying a fixed effects approach to decompose b^{j}_{d} (the level of inverse demand) and

$\tau(X_{od}^{j})$ leads to the following equation:

$$P_{dt}^{j} = P_{ot}^{j}\rho_{od}^{j} + \mu_{odt}^{j} + \mu_{odt}^{j}t + \chi_{dt}^{j} \tag{2.5}$$

Where μ_{odt}^{j} is a destination-product fixed effect, $\mu_{odt}^{j}t$ is a destination-product linear time trend and χ_{dt}^{j} is an error term.

Therefore, pass-through rates can be estimated by a regression of the destination prices P_{d}^{j} on the origin prices P_{o}^{j} and adequate fixed effects.

Step 2: recover estimates of the effect of a change in distance on trade costs After recovering the pass-through rates estimates $\widehat{\rho_{od}^{j}}$, we can estimate $\frac{\partial\left[\tau(X_{odt}^{j})\right]}{\partial Log(x_{od})}$, the effect of a change in distance on trade costs.

With the inclusion of $\widehat{\rho_{od}^{j}}$, equation 2.4 becomes:

$$\frac{P_{dt}^{j} - P_{ot}^{j}\widehat{\rho_{od}^{j}}}{\widehat{\rho_{od}^{j}}} = \tau(X_{odt}^{j}) + \frac{(1-\widehat{\rho_{od}^{j}})}{\widehat{\rho_{od}^{j}}}b_{dt}^{j} \tag{2.6}$$

Equation 2.6 shows that internal trade costs can be estimated by using price gaps adjusted by the estimates of the pass-through. However, the demand shifter b_{dt}^{j} is not observable. Therefore, a fixed effect approach can also be used thereby decomposing b_{dt}^{j} as a sum of fixed effects.

We decompose also $\tau(X_{odt}^{j})$ as a linear function of the log of distance and other cost-shifters as follows:

$$\tau(X_{odt}^{j}) = \beta Log\,(x_{od}) + \theta_{dt} \tag{2.7}$$

In this equation, x_{od} is the distance from the origin location to the destination and θ_{dt} represents all the others determinants of trade costs. β is the effect of distance on internal trade costs, the parameter of interest. In the empirical analysis, we also allow a non-parametric relation between trade costs and the log of distance.

Using all these assumptions leads to the equation 2.8 which is the key equation of our empirical estimation of the internal trade costs in Rwanda:

$$\frac{P_{dt}^{j} - P_{ot}^{j}\widehat{\rho_{od}^{j}}}{\widehat{\rho_{od}^{j}}} = \beta Log\left(x_{od}\right) + \frac{(1-\widehat{\rho_{od}^{j}})}{\widehat{\rho_{od}^{j}}}\gamma_{t}^{j} + \frac{(1-\widehat{\rho_{od}^{j}})}{\widehat{\rho_{od}^{j}}}\gamma_{d} + \pi_{dt}^{j} \tag{2.8}$$

In equation 2.8, γ_t^j is a time-product fixed effects and γ_d is a destination fixed effects. Moreover, π_{dt}^j is an unobserved error term that captures any others determinants of trade costs and measurement error in $\frac{P_{dt}^{j} - P_{ot}^{j}\widehat{\rho_{od}^{j}}}{\widehat{\rho_{od}^{j}}}$.

2.4 Internal trade costs as a function of distance: the size in Rwanda

We turn now to our empirical analysis. This section presents estimates of the extent to which internal trade costs vary with distance. First, we present our baseline results. Second, we assess the robustness of these results. Then, we focus on heterogeneous effects. Finally, we compare our results with those estimated in the literature.

2.4.1 Effect of distance on internal trade costs: baseline results

From all markets pairs to trading market pairs

We first estimate equation 2.8 in the case where the price differences are not purged from spatial variation of markups. That is the case when $\widehat{\rho_{od}^{j}} = 1$. Our estimation uses both trading market pairs and all market pairs. Table 2.4 presents the results of these estimations. In this table, columns (1) and (2) show respectively the effect of distance (in logs) on price gaps using all market pairs and trading market pairs. The main message that emerges is that price gaps increase with distance but this relation is underestimated when we use all market pairs of markets instead of trading markets pairs. More specifically, the estimates in columns (1)-(2) imply that an intermediary who trades goods 10 miles away from a given location, pays an additional trade cost of 10 cents if we consider trading market pairs, but only 7 cents if we consider all market pairs. Thus, we underestimate by a factor of 1.4 the effect of distance on trade costs by using price difference between all market pairs of markets instead of using price difference between trading markets pairs. The concavity of the distance effect plays an important role here to explain this result. [8]

Table 2.4: ***Estimating the effect of distance on internal trade costs***

	From all market pairs to Trading market pairs		From price gap to adjusted price gap	
	(1)	(2)	(3)	(4)
Dependent variable:	Absolute Price Gap (All market pairs)	Price Gap (Trading market pairs)	Price Gap (Trading market pairs)	Adjusted Price Gap (Trading market pairs)
Log distance to origin	0.030[a]	0.042[a]	0.042[a]	0.106[b]
	(0,002)	(0.012)	(0.012)	(0.048)
Observations	9,598,870	48,661	48,661	41,593
R-squared	0.480	0.220	0.220	0.809
Time-Product FE	Yes	Yes	Yes	Yes
Time-product $\times \frac{1-\hat{\rho}}{\hat{\rho}}$ FE	No	No	No	Yes
Destination $\times \frac{1-\hat{\rho}}{\hat{\rho}}$ FE	No	No	No	Yes

Notes This table investigates the effect of distance on trade costs. Column (1) uses data on the absolute price gap between All market pairs of markets. Columns (2) and (3) present the same regression. It uses data on the price gaps between Trading market pairs. Column (4) uses the price gaps adjusted with pass-through rates. Standard errors are in parentheses and clustered at the time-product level. Prices are converted in US dollars using the prevailing exchange rate in November 2016. Distances are in miles. [c] significant at 10 percent level, [b] at 5 percent level, [a] at 1 percent level.

Estimating pass-through rates

As described in the previous section, the first step to correctly estimate the effect of distance on trade, costs is to estimate the pass-through rate for each origin-destination-product triplet.

The pass-through is a sufficient statistic to assess the distribution of margins in the markets. It is defined as the effect on destination prices of a change in the marginal cost of intermediaries. (In this chapter, this variation in marginal cost is caused by a variation in the origin-destination distance traveled by intermediaries). It can be interpret as the effect of a variation in the distance traveled by intermediaries on the markups they charge (this effect of the distance passes through an effect on the marginal cost of intermediaries). A greater estimated pass-through suggests that a change in distance

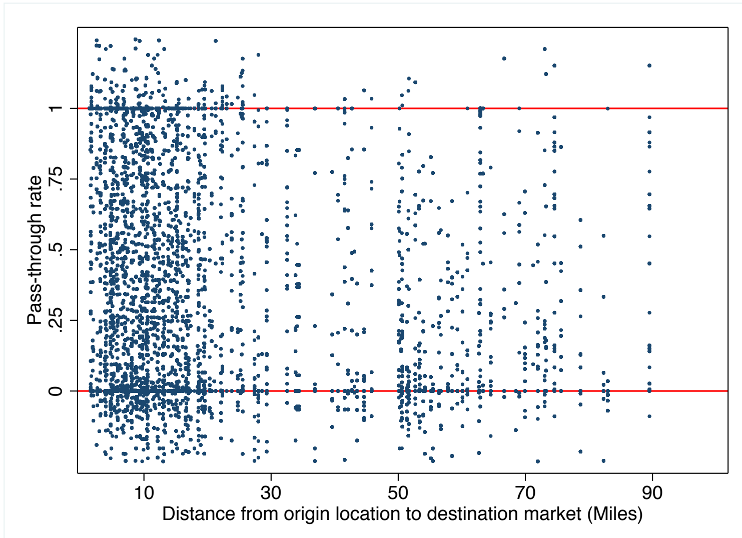
8. See figure 2.17 in the appendix

traveled (and therefore in marginal cost) has more impact on markups and therefore prices.

Figure 2.9 plots the pass-through rates for all products and all origin-destination pairs. The majority of pass-through rate estimates are between 0 and 1. This result shows that the intermediaries market in Rwanda is characterized by imperfect competition. Indeed, as shown in the column (1) of Table 2.5, a 1 dollar increase in origin price leads to an increase of 34 cents in destination price on average. In addition, Table 2.5 reveals an interesting feature on the distribution pass-through rates. This distribution is positively skewed since the average pass-through rate is approximately 1.4 times the median pass-through rate. Our estimates of pass-through rates are in range with those estimated in the literature. Atkin and Donaldson (2015) find an average pass-through rate of 0.58 for Ethiopia and 0.39 for Nigeria. As in the literature, some pass-through rates are negative. To deal with this problem we follow Atkin and Donaldson (2015) by winsorizing the pass-through rates at 0.2.

Turning now to the relation between pass-through rate and distance, Figure 2.10 presents a non-parametric regression of pass-through rate estimates on distance. This figure shows that pass-through decreases with distance. It means that the intermediaries tend to charge lower markups in remote locations.

Figure 2.9: ***Estimated pass-through rates for all goods***



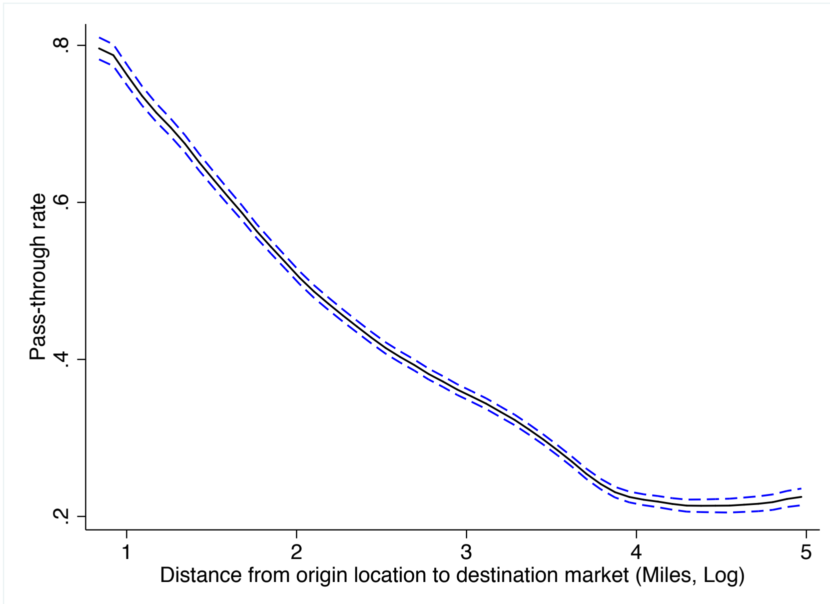
Notes: This figure plots the estimated pass-through rates against distance from origin location to destination markets.

Table 2.5: ***Estimated pass-through rates***

	(1)	(2)	(3)	(4)	(5)
			Percentile		
	Mean	SD	75	50	25
Pass-through rate	0.339	0.996	0.824	0.241	0.102

Notes: This table presents the basic characteristics of our estimates of pass-through rates.

Figure 2.10: ***Estimated pass-through rates and distance***



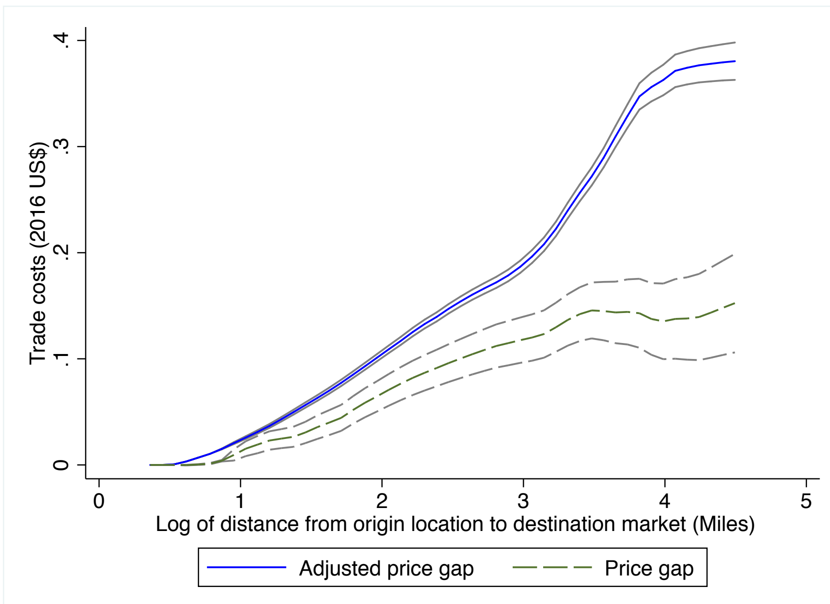
Notes: This figure presents a non-parametric regression of pass-through rates on distance. We use a locally weighted polynomial (Epanechnikov kernel, bandwidth=0.5). 95% confidence intervals is shown.

The effect of distance on internal trade costs: from price gap to adjusted price gap

The estimation of pass-through rates allows us to correct price differences from the spatial variation of markups. We refer to the "adjusted price gap", the price differences corrected by the pass-through rate estimates. We first start by estimating a non-parametric relationship between price gaps and distance as well as between adjusted price gap and distance. Figure 2.11 shows these non-parametric relationships. The main result that emerges is that the effect of distance on trade costs is positive but underestimates when we do not correct price differences by the pass-through estimates. Second, we estimate a linear relationship between price gaps (adjusted and non-adjusted) and distance as described in equation 2.8. Columns (3)-(4) of Table 2.4 present the results of these estimations.

These results imply that an intermediary who trades goods 10 miles away from a given location, pays an additional trade cost of 25 cents if we purge price gaps from spatial variation of markups. However, he pays an additional trade cost of only 10 cents if we do not take into account the spatial variation of markups. As a result, we tend to underestimate by a factor of 2.5 the effect of distance on trade costs if we do not correct price gaps by pass-through estimates. This result is consistent with the results estimated in the literature for Ethiopia, Nigeria, and Liberia. Our preferred estimates of the effect of distance on trade costs are presented in column (4) of Table 2.4.

Figure 2.11: ***The effect of distance on internal trade costs***



Notes: This figure presents a non-parametric regression of prices gaps and adjusted price gaps (trade costs) on distance. We use a locally weighted polynomial (Epanechnikov kernel, bandwidth=0.5). 95% confidence intervals is shown.

2.4.2 Robustness checks

In this subsection, we check the robustness of our baseline empirical regularities. We conduct two types of checks. First, we check if our baseline results are sensitive to other measures of distance. Thus, we consider the following alternative measures of distance: the great circle distance; the travel time; and the travel distance. As shown in table 2.6, using other measures of distance does not change that much the magnitude of the effect of distance on trade costs. Indeed, trade costs increase with travel distance and travel time. Column (2) suggests that an intermediary who travel 10 more miles to trade goods, pays an additional trade cost of 19 cents.

Table 2.6: ***Internal trade costs and distance: other metrics***

Dependent variable:	Adjusted Price Gap		
	(1)	(2)	(3)
Log of great circle distance to origin	0.105^{b}		
	(0.048)		
Log of travel distance to origin		0.081^{c}	
		(0.044)	
Log of travel time to origin			0.094^{c}
			(0.052)
Observations	41,593	41,593	41,593
R-squared	0.809	0.809	0.809

Notes. This table presents effect of distance on internal trade costs using alternative metrics for distance. These alternative metrics are respectively, the great circle in column (1), the travel distance in column (2) and the travel time in column (3). Distances are in miles. Standard errors are in parentheses and clustered at time-product level. Prices are converted in US dollars using the prevailing exchange rate in November 2016. All these regressions include Time-Product fixed effects, $Time-product \times \frac{1-\hat{\rho}}{\hat{\rho}}$ fixed effects and $Destination \times \frac{1-\hat{\rho}}{\hat{\rho}}$ fixed effects. ^{c} significant at 10 percent level, ^{b} at 5 percent level, ^{a} at 1 percent level.

In a second robustness check, we investigate whether the relationship between trade costs and distance is stable across different distance intervals. We consider four distance intervals: less than 20 miles; between 20 and 50 miles; between 50 and 80 miles; more than 80 miles. We regress the adjusted price gaps on distance interval dummies. Table 2.7 reports the results and suggests that the further the destination from the origin, the greater the effect of distance on trade costs. Indeed, a coefficient equality test shows that the coefficients are significantly different from each other. An interesting point here is that this increase in trade costs is not always significant toward markets lying at a distance between 20 and 50 miles of the origin location.

Table 2.7: ***Trade costs and distance intervals***

Dependent variable:	Adjusted Price Gap
20 < distance to origin (miles) < 50	0.0683
	(0.073)
50 < distance to origin (miles) < 80	0.2148^c
	(0.114)
80 < distance to origin (miles)	0.3706^a
	(0.127)
Observations	41,593
R-squared	0.809

Notes. This table examines the impact of distance on trade costs using distance intervals. Dummy is equal to 1 if the distance between origin and destination belongs to the interval and 0 otherwise. Standard errors are in parentheses and clustered at time-product level. Prices are converted in US dollars using the prevailing exchange rate in November 2016. All these regressions include Time-Product fixed effects, $Time-product \times \frac{1-\hat{\rho}}{\hat{\rho}}$ fixed effects and $Destination \times \frac{1-\hat{\rho}}{\hat{\rho}}$ fixed effects. ^{c} significant at 10 percent level, ^{b} at 5 percent level, ^{a} at 1 percent level.

2.4.3 Heterogeneous effects of distance on internal trade costs

We now investigate potential heterogeneity in our estimation of the effect of distance on internal trade costs in Rwanda. We focus on three types of products: Non-manufactured products, imported products, and Food. The idea here is to go beyond our basic results and ask if there is some heterogeneity in the calculation of trade costs elasticity with respect to distance if we consider different categories of products. The distinction between manufactured and non-manufactured products makes it possible to see whether in a rural context there is a difference in terms of trade costs. The domestic / imported distinction makes it possible to emphasize the specificity of foreign products. The food / non-food distinction makes it possible to separate the manufactured goods that are food from those that are not food. In the context of rural area in a poor country, food products are a stake for households. Whether or not it costs more to get it to them seems important. We identify the observations corresponding to these types of products by using adequate dummies that we interacted with the log of distance. Table 2.8 presents our results. Starting with non-manufactured products, column (2) of Table 2.8

shows that there is significant heterogeneity across product type. [9] Trade costs increase less with distance for non-manufactured products than for manufactured products. The difference in the mode of transportation used to ship non-manufactured products may explain this result. Likewise, there is a significant heterogeneity for food products (see columns (4) and (5)). The effect of distance on trade costs is lower for food products compared to non-food products. We work in a rural area configuration. In this configuration, the majority of food / non-manufactured products are produced locally to such an extent that one would expect lower trade costs for food / non-manufactured products compared to non-food / manufactured products. When looking at imported products, columns (3) and (5) show no significant difference between imported versus non-imported goods. This result might be drive by the fact that very limited imported goods are included in the sample of products which was designed from a basket of common goods in a rural setting.

Table 2.8: ***Internal trade costs and distance: heterogeneity***

Dependent variable:	Adjusted Price Gap				
	(1)	(2)	(3)	(4)	(5)
Log distance to origin	0.106^b	0.172^a	0.127^b	0.253^a	0.319^a
	(0.048)	(0.059)	(0.053)	(0.095)	(0.096)
Log distance to origin × Non-manufactured		-0.180^a			-0.075
		(0.051)			(0.046)
Log distance to origin × Imported			-0.015		-0.039
			(0.034)		(0.034)
Log distance to origin × Food				-0.240^a	-0.211^b
				(0.090)	(0.102)
Observations	41,593	41,593	41,593	41,593	41,593
R-squared	0.809	0.809	0.809	0.809	0.809

Notes. This table presents the heterogenous effect of distance on internal trade costs. Column (1) represents our baseline regression. In column (2) we add an interaction term with a dummy for non-manufactured products. Columns (3)-(4) include respectively an interaction terms with dummies for imported and food products. In column (5) we take into account all the previous interaction terms. Distances are in miles. Standard errors are in parentheses and clustered at time-product level. Prices are converted in US dollars using the prevailing exchange rate in November 2016. All these regressions include Time-Product fixed effects, $Time-product \times \frac{1-\hat{\rho}}{\hat{\rho}}$ fixed effects and $Destination \times \frac{1-\hat{\rho}}{\hat{\rho}}$ fixed effects. [c] significant at 10 percent level, [b] at 5 percent level, [a] at 1 percent level.

9. We do the same exercise in Table 2.18 in the appendix using price gaps instead of adjusted price gaps. Table 2.19 in the appendix presents some indicators within industries, in particular non-manufacturing and agricuture.

2.4.4 How large are internal trade costs in Rwanda compared to other countries?

In this subsection, we compare our estimates to those estimated in the recent literature. Estimates are presented in Table 2.9 and includes results from Ethiopia, Nigeria, U.S. (Atkin and Donaldson, 2015), and Liberia (de Rochambeau, 2018a). Starting by comparison within East Africa, Rwanda and Ethiopia are two landlocked countries and are similar in terms of GDP per capita. However, they are characterized by a significant difference in their internal trade costs. Columns (1)-(2) of Table 2.9 show that internal trade costs in Rwanda are approximately three times higher than in Ethiopia. To go further with the comparison of trade costs between Ethiopia and Rwanda, we restrict our comparison to food products. Table 2.10 presents the total elasticity of trade costs relative to distance in Rwanda and Ethiopia for foods products. The elasticity in column (1) is obtained by running the regression 2.8 and divided the coefficient on distance (0.0782) by the mean of price gap (0.088). [10] The elasticity in column (2) comes from chapter 1, where we computed the actual trade costs elasticity of foods in Ethiopia using World Food Program data. The results show that internal trade costs for foods in Rwanda are approximately 1.2 times higher than in Ethiopia.

We hypothesized three reasons that explain these differences. First, the market power enjoys by firms in the transport sector, especially in the trucking industry. In a policy note, Vincent Safari (2016) [11], Technical Adviser with TradeMark East Africa (TMEA) in Rwanda stress that almost 30% of the price of goods in Rwanda is for transport. He argues that if the costs of transportation go down, it will have a huge impact on the man in the street. Our pass-through rates estimates provide a good idea on the market power of transportation firms since a high pass-through means a high competition and a less market power. The estimations show that the average pass-through rate in Ethiopia is about 1,7 times higher than in Rwanda. If the trucking industry is not competitive, it

10. Regression 2.8 gives the semi-elasticity. To obtain the total elasticity, we divide the semi-elasticity by the mean of the price gap

11. Safari, V, UD: Reducing non-tariff barriers in Rwanda equals reduced prices for all. TradeMark East Africa (TMEA), https://www.trademarkea.com/impactstories/reducing-non-tariff-barriers-in-rwanda-equalsreduced-prices-for-all/ 08/07/2016

probably means, firms have a high market power, which can lead to high internal trade costs.

Second, the low quality of transport vehicles (Hartmann and Asebe, 2013) . The quality of transport vehicles affects internal trade costs. This is reflected in the high cost of poor mechanical condition of vehicles, leading to high operating costs that are in turn passed on to the consumers. This is exacerbated by the fact that transport vehicles operation is dominated by small fleets that do not benefit from economies of scale in logistics and pricing. Third, the poor road quality and transport infrastructure, in general, is a major factor (Hartmann and Asebe, 2013). In our sample, the quality of road connecting markets in rural areas is poor. Finally, being landlocked, Rwanda is far from the nearest port, and transportation costs to export markets are high. This remoteness constitutes a challenge for internal trade costs.[12] The country is accelerating infrastructure investments that can help mitigate the negative impacts. The governments of Rwanda and Tanzania have recently agreed to expedite the process to implement the Isaka-Kigali Standard Gauge Railway (SGR) project, which is a cross-border railway link connecting the two countries.

Turning to a comparison with west African countries, columns (1), (3) and (4) of Table 2.9 show that internal trade costs in Rwanda is approximately 1.3 times lower than in Liberia but two times higher than in Nigeria. More specifically, an intermediary who trades goods 10 miles away from the farm gate, pays an additional trade cost of 25 cents in Rwanda, 33 cents in Liberia, and 13 cents in Nigeria. The explanations previously mentioned might explain the difference between Nigeria and Rwanda. In addition to the arguments mentioned above, the difference in terms of internal trade costs between Liberia and Rwanda might be explained by corruption issues. Indeed, bribes constitute an important trade costs-shifter in the African context. As shown in Table 2.9 corruption seems to be more problematic in Liberia than in Rwanda.

---

12. Rwanda lacks a link to regional railway networks, which means that all trade is conducted by road or air. Most of the country's imports are shipped by road from the ports of Mombasa (Kenya) and Dar-es-Salam (Tanzania).

Table 2.9: ***Comparison with the literature***

	East Africa		West Africa	
	(1) Rwanda	(2) Ethiopia	(3) Liberia	(4) Nigeria
Country size ratio (U.S. as benchmark)	373.4	8.9	88.3	10.64
GDP per capita ratio (U.S. as benchmark, 2017)	27.46	26.85	44.16	10.39
Average distance between trading markets (miles)	18	221	90	350
Trade costs elasticity estimates ratio (U.S. as benchmark)	9.96	3.53	13.40	5.26
Corruption index Rank (2018)	48/180	114/180	120/180	144/180

Notes: This table compares the results estimated in this paper with those estimated in the literature. The corruption index ranks countries from the less corrupt (1) to the more corrupt (180) and is provided by transparency international. For all the ratios computed in this table, U.S.A is the benchmark.

Table 2.10: ***Trade costs for foods : Rwanda versus Ethiopia***

	(1) Rwanda	(2) Ethiopia
Trade costs elasticity for foods	0.888	0.757

Notes: We obtain the number in column 1 by doing 0.0782/0.088. 0.0782 is the semi-elasticity of trade costs relative to transport. 0.088 is the mean value of price gaps in Rwanda.The number in column 2 comes from chapter 1.

Turning next to the extent to which the size of internal trade costs in Rwanda would have changed if they were estimated by using a level of pass-through that is similar to other countries like Nigeria (0.39), Ethiopia (0.58), and the U.S. (0.77). We used average pass-through rates estimated in Atkin and Donaldson (2015). Table 2.11 shows our results, both for the Rwanda baseline and the estimation with other countries' pass-through rates. As shown, all coefficients on the adjusted price gap are positive and significant, except for the model with Nigeria pass-through rate which is only significant at 10% level. Holding the pass-through rate at the U.S. levels, the size of internal trade costs in Rwanda, would have been 56.6% less. The same holds for Ethiopia and Nigeria, with the size of internal trade cost reducing by 53.7% and 49.1% respectively.

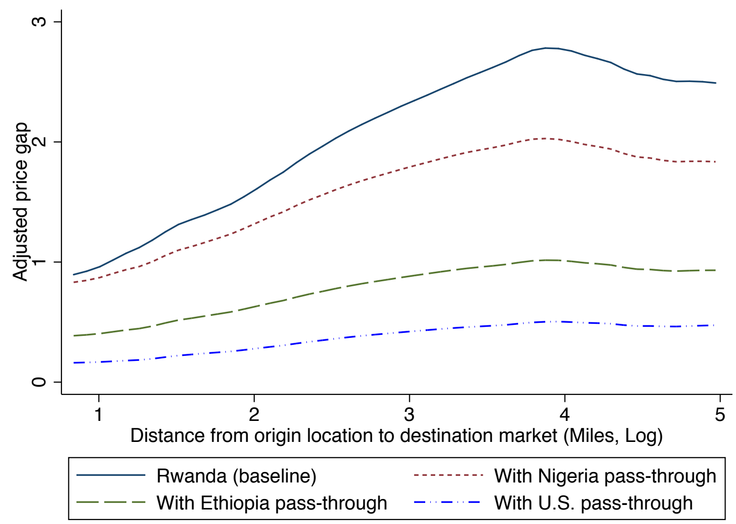
As can also be seen from Figure 2.12, the gap between the Rwanda baseline estimate and those using other countries' pass-through rates is increasing with the distance from the origin to destination locations. This implies that had the Rwanda pass-through rates fixed at those countries levels, the size of internal trade cost fin Rwanda would have been reduced by a factor of almost 2.5.

Table 2.11: ***The effect of a change in the pass-through level on the size of trade costs in Rwanda***

	Coefficient from regression of adjusted price gap on the logarithm of distance
Baseline	0.106^{b}
	(0.048)
Using Nigeria Pass-through	0.054^{c}
	(0.031)
Using Ethiopia Pass-through	0.049^{b}
	(0.022)
Using U.S. Pass-through	0.046^{b}
	(0.018)

Notes: We calibrate the pass-through in Rwanda at the level of the average pass-through in Nigeria (0.39), Ethiopia (0.58), and the U.S. (0.77). All regressions include Time-Product fixed effects, $Time-product \times \frac{1-\hat{\rho}}{\hat{\rho}}$ fixed effects and $Destination \times \frac{1-\hat{\rho}}{\hat{\rho}}$ fixed effects. ^{a}, ^{b}, ^{c} denote coefficients significant at the 1%, 5% and 10% levels, respectively.

Figure 2.12: ***Internal trade costs and pass-through : counterfactual analysis***

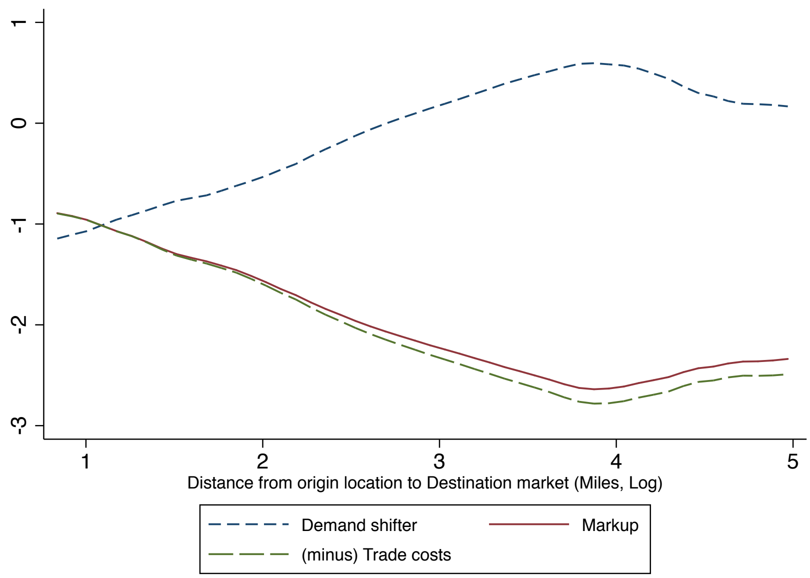


Notes: This figure presents non-parametric regressions of adjusted price gaps on distance using different level of pass-trough rates. We calibrate the pass-through level in Rwanda using respectively the average level in Nigeria (0.39), Ethiopia(0.58) and the U.S.(0.77). We use a locally weighted polynomial (Epanechnikov kernel, bandwidth=0.5).

2.4.5 Unpacking the Intermediary Markup

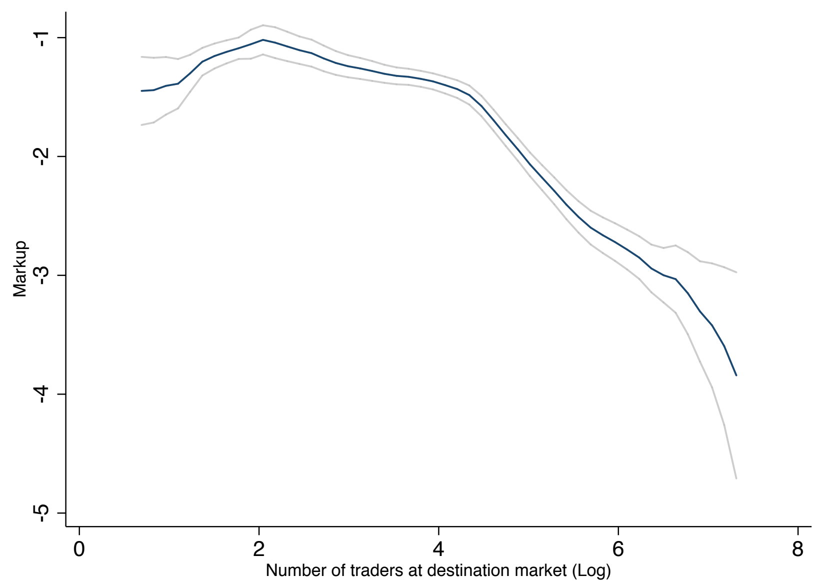
We turn to an investigation of how and why markups vary over space. Figure 2.13 shows that markups decline with distance implying that intermediaries charge lower markups in remote areas. This pattern has two possible explanations. First, consumers in remote locations are in majority poor and traders in these areas faced higher trade costs. Consequently, the demand in remote areas is strongly elastic and the traders charge low markups. Figures 2.14 and 2.15 propose another explanation of the decrease of markups with distance. We are in a rural setting where there are few traders in markets close to production locations. Traders are therefore charging high markups in these markets location, but tend to charge lower markups in markets far away from the farm gates. Therefore, a rise in prices would lead to a substantially higher fall in the demand, and traders would not be able to increase their total revenue by increasing prices to a level that consumers cannot afford. Moreover, the costs for firms (land and labor in particular) are lower in rural areas, meaning they can easily afford and survive by charging lower prices. On the contrary, because incomes are lower in rural areas, consumers cannot afford to pay or support higher prices.

Figure 2.13: ***Markups and distance***



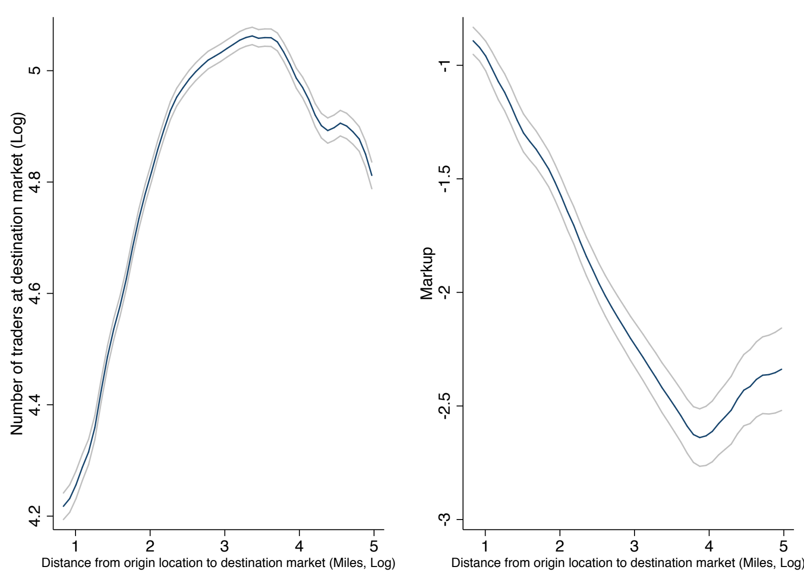
Notes: This figure presents respectively non-parametric regressions of trade costs (minus), markups and demand shifter on distance. We use a locally weighted polynomial (Epanechnikov kernel, bandwidth=0.5).

Figure 2.14: ***Correlation between markups and competition intensity2***



Notes: This figure presents a non-parametric regression of markups on competition intensity. We use a locally weighted polynomial (Epanechnikov kernel, bandwidth=0.5).95% confidence intervals is shown.

Figure 2.15: ***Competition intensity, markups and distance***



Notes: This figure presents respectively non-parametric regressions of the number of intermediaries and markups on distance. We use a locally weighted polynomial (Epanechnikov kernel, bandwidth=0.5). 95% confidence intervals is shown.

2.5 Internal Trade Costs and Connectivity

This section aims to investigate the potential effects of improved road connectivity on internal trade costs defined as a function of distance. More specifically, we focus on the effect of connecting markets to feeder roads and national roads on internal trade costs. To begin with, we will start by examining the evolution of key economic variables related to internal trade costs. We will then evaluate the effects of improved connectivity (through better access to feeder and national roads) on the size of internal trade costs.

2.5.1 Effects of Improved Connectivity on Internal Trade Costs

The previous subsection has shown that better connectivity through access to feeder and national roads is correlated with the change in key economic variables related to trade costs. In this subsection, we will investigate the potential effects of improved connectivity through better access to feeder roads and national roads on internal trade costs. We begin by fa descriptive analysis, which focus on the comparison and evolution of key economic variables in market that are connected versus those that not connected to a feeder or national roads. The economic variables we are interested in can be classified into two categories: price-related variables and market characteristics variables. We then run our baseline regression of the adjusted price gap on the logarithm of distance with an interaction term to account for better connectivity. Results are presented in Table 2.12, Table 2.13, and Table 2.14 below.

Table 2.12: ***Summary statistics of economic variables in connected versus non-connected markets***

	Mean	SD
Panel A: Three first months of dataset		
Adjusted Price Gap to origin price ratio	0.519	2.710
Competition intensity index	4.663	0.964
Price Gap	0.101	1.421
Price Gap to origin price ratio	0.087	0.819
Inverse of Herfindhal index	614.602	273.827
Panel B: Three last months of dataset		
Adjusted Price Gap to origin price ratio	0.323	0.712
Competition intensity index	4.992	1.431
Price Gap	0.063	1.266
Price Gap to origin price ratio	0.027	0.218
Inverse of Herfindhal index	577.884	288.709

Notes: This table presents the basic statistics of key economic variables related to prices and market structure. These statistics are computed for connected and non-connected markets. We focus on access to feeder and national roads. Our competition intensity index is the logarithm of the number of traders in a market. We use the inverse of the Herfindahl index, as our measure of trade concentration. A decrease in this index means a higher concentration. The Adjusted Price Gap to origin price ratio is the Adjusted Price Gap divided by the origin price. The Price Gap to origin price ratio is the Price Gap divided by the origin price.

Looking at prices-related variables, Table 2.12 shows that on average, price gaps decrease for better-connected markets. Both the adjusted price gap to origin price and price gap to origin price ratios also fall. As price gaps indirectly give information about trade costs, these results may reflect lower trade costs. For the market indicators, Table 2.12 suggests that on average, there is a higher competition intensity and trade concentration in the intermediaries market in better-connected markets. Since our trade concentration measure is the inverse of the Herfindahl index, a decrease of the index means a higher concentration. These descriptives suggest that better connectivity, through access to feeder and national roads is positively correlated to the intensity of competition and the concentration in the markets. This could potentially have implications for the size of internal trade costs defined as a function of distance.

We now move to our regression analysis. Table 2.13 displays our results for better access to feeder roads and national roads. In column (2) of Table 2.13 , we regress the adjusted price gap on the logarithm of distance and an interaction term between this measure of distance and a dummy for access to feeder roads. This dummy for access to feeder roads equal to one if the location origin is within a 1-kilometer buffer around a feeder road. In Column (3), we add to the previous explanatory variables plus an interaction term between the measure of distance and a dummy for national roads. This dummy for the national roads equal to one if the location origin is within a 1-kilometer buffer around a national road. A close look at table 2.13 shows that access to feeder roads helps reduce the size of internal trade costs defined as a function of distance when they allow access to national roads. On average, connecting a location to a feeder road reduces trade costs by approximately 2 cents per ile. The effect of feeder roads on trade costs transits through the connection with national roads. In column (2), the coefficient of the interaction term between the logarithm of distance and a dummy for the feeder roads is negative but not significant. However, when we add the interaction term between the distance and a dummy for access to national roads, this coefficient becomes significantly negative. Likewise, the interaction term between the logarithm of distance and a dummy for the national road is strongly negatively significant and has a higher magnitude.

2.5.2 Connectivity Intensity and the Size of Internal Trade Costs

In complement to the above analyses, we considered three dummies to capture the intensity and the different levels of road connectivity: low connectivity, intermediate connectivity, and high connectivity. We define these levels of connectivity as follows. First, an origin location has low connectivity if it is located neither within a 1-kilometer buffer of a national road nor a 1-kilometer buffer of a feeder road. Second, an origin location has intermediate connectivity if it is within a 1-kilometer buffer around a national road but not within a 1-kilometer buffer around a feeder road. Third, an origin location is highly connected if it is within a 1-kilometer buffer around both a national and a feeder road. We regress the adjusted price gap on the logarithm of distance and where we gradually added interaction terms between the measure of distance and our road connectivity dummies. [13]

As can be seen in Table 2.14, better connection to markets is associated with lower internal trade costs. When considering intermediate connectivity in column (2), there is no significant effect on trade costs; however, when considering high connectivity (see column (3)), the coefficient on the interaction term is significantly negative. [14] Finally, the coefficients on the interaction terms for both intermediate and high connectivity are significantly negative, with a stronger impact on the high-connectivity. These results are consistent with those obtained in Table 2.13. Overall, our results indicate the importance of the last mile connectivity as the combination of access to feeder and national roads leads to a significant reduction of the size internal trade costs.

Table 2.13: ***Road connectivity and trade costs***

Dependent variable:	Adjusted Price Gap		
	(1)	(2)	(3)
Log distance to origin location	0.106^b	0.152^a	0.191^a
	(0.048)	(0.052)	(0.059)
Log distance to origin × Feeder road		-0.036	-0.060^b
		(0.023)	(0.023)
Log distance to origin × National road			-0.123^a
			(0.035)
Observations	41,593	41,593	41,593
R-squared	0.809	0.809	0.810

Notes. This table investigates the effect of better connectivity on the size of trade costs. Standard errors are in parentheses. Prices are converted in U.S. dollars using the prevailing exchange rate in November 2016. All these regressions include time-product fixed effects, $Time - product \times \frac{1-\hat{\rho}}{\hat{\rho}}$ fixed effects and $destination \times \frac{1-\hat{\rho}}{\hat{\rho}}$ fixed effects. ^{c} significant at 10 percent level, ^{b} at 5 percent level, ^{a} at 1 percent level.

13. Due to the lack of observations and collinearity issues we were not able to run regression including the interaction term between the logarithm of distance and the low connectivity dummy.

14. We do the same exercise in Table 2.20 in the appendix using price gaps instead of adjusted price gaps.

Table 2.14: ***Connectivity and trade costs***

Dependent variable:	Adjusted Price Gap			
	(1)	(2)	(3)	(4)
Log distance to origin location	0.106^{b}	0.117^{b}	0.140^{a}	0.130^{b}
	(0.048)	(0.046)	(0.052)	(0.055)
Log distance to origin × Intermediate connectivity		0.036		-0.063^{b}
		(0.030)		(0.031)
Log distance to origin × High connectivity			-0.096^{a}	-0.123^{a}
			(0.028)	(0.025)
Observations	41,593	41,593	41,593	41,593
R-squared	0.809	0.809	0.809	0.810

Notes. This table investigates the effect of better connectivity on trade costs. Intermediate connectivity is a dummy variable which equals one if the origin location is within a 1-kilometer buffer of a national road but not within a 1 kilometer buffer of a feeder road. High connectivity is a dummy variable equal to one if the origin location is within a 1 kilometer buffer of both a national and a feeder road. Standard errors are in parentheses. Prices are converted in U.S. dollars using the prevailing exchange rate in November 2016. All these regressions include time-product fixed effects, $time - product \times \frac{1-\hat{\rho}}{\hat{\rho}}$ fixed effects and $destination \times \frac{1-\hat{\rho}}{\hat{\rho}}$ fixed effects. ^{c} significant at 10 percent level, ^{b} at 5 percent level, ^{a} at 1 percent level.

2.6 Conclusion

This paper builds on the emerging literature looking at the size of internal trade costs in the context of developing countries. In contrast to previous works using the spatial dispersion of prices in urban areas, our focus is on rural markets. We aimed to make two contributions to our understanding of the size of internal trade costs in the context of a developing and landlocked country. Using a new high-frequency market price data, our first contribution is to estimate the size of internal trade costs in Rwanda. We find that internal trade costs, defined as a function of distance, are high in Rwanda. They are approximately ten times higher than in the U.S., three times higher than in Ethiopia, and two times higher than in Nigeria. Looking at the heterogeneity of the size of internal trade costs, our results showed that internal trade costs are lower for non-manufactured products than manufactured products, and they are lower for food products compared to non-food products.

Our second contribution is to use the existing road network to test whether better road connectivity can reduce the size of internal trade costs in Rwanda. We find that better accessibility to markets, through feeder roads and national roads contributes to lower the size of internal trade costs. We also provide evidence of the importance of the last mile connectivity, as internal trade costs in markets that connect to both feeder roads and national roads are substantially lower. Finally, our third contribution is to empirically shed light on two potential mechanisms that are often linked to the size of internal trade costs in the context of a landlocked and developing country. We test the association between the size of internal trade costs, market power, and the quality of the transportation infrastructure. We show that intermediaries are charging higher markups in locations where there is less competition, measured by the number of traders in a market. We also show that access to feeders and national roads matter as connecting a location to a feeder or national road reduces trade costs by approximately two cents per mile.

This paper shows that trade costs within countries matters. Firms and people that face high costs of moving their goods from the factory gate to local markets or international gateway have an extra hurdle to clear. We show that investments in transport infrastructure and policy encouraging competition in the intermediary and transportation sectors can help mitigate this extra hurdle. This is relevant as understanding and tackling non-tariff barriers could increase the wider economic benefits of the recent framework agreement establishing the African Continental Free Trade Area (AfCFTA). One potentially important question for future research concerns the extent to which a reduction in trade costs impacts the cost of living, trade, productivity, and the specialization of places.

2.7 Appendix to Chapter 2

2.7.1 Additional tables

Table 2.15: ***Descriptive statistics of prices***

	Destination price					Origin price				
	Mean	SD	Median	Minimum	Maximum	Mean	SD	Median	Minimum	Maximum
Airtime	0.133	0.063	0.124	0.124	0.620	0.147	0.104	0.124	0.124	0.620
Amarante	0.109	0.047	0.124	0.025	0.620	0.104	0.045	0.124	0.025	0.620
Banana	0.235	0.248	0.124	0.006	1.860	0.221	0.243	0.124	0.006	1.860
Brochette	0.374	0.097	0.372	0.124	0.620	0.373	0.096	0.372	0.124	0.620
Cabbage	0.161	0.062	0.124	0.025	0.496	0.156	0.062	0.124	0.025	0.372
Candle	0.062	0.005	0.062	0.056	0.248	0.075	0.073	0.062	0.062	0.496
Carrots	0.299	0.215	0.248	0.062	1.860	0.284	0.201	0.248	0.062	1.860
Cassava flour	0.564	0.101	0.558	0.248	0.868	0.552	0.109	0.558	0.248	1.488
Cassava leaves	0.401	0.168	0.372	0.037	1.240	0.389	0.164	0.372	0.037	1.240
Cassava root	0.491	0.383	0.372	0.124	3.101	0.480	0.365	0.372	0.124	2.481
Celery	0.099	0.064	0.062	0.025	0.558	0.103	0.121	0.062	0.025	1.860
Charcoal	1.191	2.587	0.248	0.019	19.845	1.058	2.300	0.248	0.019	12.403
Commercial beer	0.941	0.160	0.992	0.496	1.240	0.875	0.174	0.930	0.099	1.240
Cook banana	0.927	1.544	0.310	0.124	11.163	0.863	1.442	0.310	0.124	11.163
Cook oil	1.986	0.655	1.984	0.434	8.434	1.816	0.477	1.860	0.124	8.310
Detergent	0.490	0.224	0.558	0.062	5.581	0.492	0.190	0.558	0.062	1.240
Dry beans	0.590	0.118	0.595	0.248	1.116	0.563	0.107	0.558	0.025	0.992
Egg	0.125	0.016	0.124	0.062	0.186	0.124	0.017	0.124	0.062	0.186
Garlic	0.586	0.876	0.124	0.062	9.922	0.600	0.855	0.124	0.012	9.922
Green pea	1.086	0.444	0.992	0.310	3.721	1.018	0.363	0.992	0.124	3.721
Goundnut flour	1.279	0.503	1.488	0.062	2.233	1.171	0.578	1.488	0.062	2.233
Ground nut	1.336	0.170	1.364	0.620	2.295	1.298	0.214	1.240	0.124	2.295
Hair dye	0.382	0.042	0.372	0.248	0.620	0.419	0.131	0.372	0.037	0.992
Imported rice	1.068	0.081	1.054	0.496	1.612	1.045	0.112	1.054	0.496	1.488
Irish potato	0.273	0.066	0.273	0.099	0.868	0.268	0.181	0.248	0.087	3.101
Leek	0.133	0.088	0.124	0.062	1.116	0.130	0.083	0.124	0.062	0.744
Maggi	0.084	0.030	0.062	0.015	0.186	0.079	0.036	0.062	0.031	0.868
Maize	0.412	0.107	0.434	0.124	0.992	0.394	0.103	0.409	0.124	0.843
Maize flour	0.631	0.147	0.620	0.124	1.488	0.642	1.300	0.558	0.124	20.465
Mandarine	0.541	0.391	0.496	0.012	1.488	0.530	0.384	0.558	0.012	1.488
Mattress	29.200	11.651	27.287	2.109	86.822	25.624	10.172	27.287	0.052	62.016
Milk	0.639	0.271	0.620	0.186	1.488	0.440	0.169	0.372	0.124	1.364
Mineral water	0.514	0.176	0.372	0.248	0.868	0.464	0.151	0.372	0.310	0.868
Nails	1.584	0.247	1.488	0.868	2.605	1.553	0.261	1.488	0.012	2.605
Onion	0.448	0.383	0.248	0.050	3.101	0.426	0.352	0.310	0.050	2.481
Pepper	0.277	0.522	0.062	0.012	4.341	0.298	0.522	0.062	0.012	4.341
Sanitary napkin	0.867	0.096	0.868	0.496	1.240	0.836	0.092	0.868	0.124	1.240
Shampoo	0.965	0.159	0.992	0.620	1.860	1.015	0.250	0.868	0.620	1.736
Soap	0.259	0.025	0.248	0.211	0.496	0.262	0.048	0.248	0.186	0.868
Softdrink	0.499	0.083	0.496	0.372	0.992	0.478	0.071	0.434	0.372	0.744
Sorghum	0.514	0.079	0.496	0.248	1.116	0.490	0.072	0.496	0.248	0.868
Sorghum flour	0.621	0.111	0.620	0.248	1.488	0.609	0.115	0.620	0.124	1.488
Spaghetti	0.723	0.193	0.868	0.248	0.992	0.730	0.187	0.868	0.310	0.992
String bean	0.354	0.237	0.372	0.062	4.341	0.337	0.252	0.372	0.062	4.341
Sugar	1.207	0.103	1.240	0.620	1.612	1.198	0.103	1.240	0.372	1.488
Suitcase	16.866	5.588	16.124	3.101	55.814	13.793	3.801	14.884	2.233	24.806
Sweet potato	0.300	0.348	0.248	0.074	3.101	0.306	0.379	0.248	0.074	3.101
Tea local	0.129	0.289	0.062	0.006	1.984	0.104	0.219	0.062	0.025	1.860
Toilet paper	0.248	0.006	0.248	0.223	0.310	0.246	0.015	0.248	0.062	0.372
Tomato	0.359	0.672	0.186	0.012	8.062	0.345	0.597	0.186	0.012	8.062
Toothpaste	0.448	0.236	0.372	0.124	1.488	0.459	0.277	0.372	0.037	1.488
Total	1.181	4.195	0.434	0.006	86.822	0.914	3.077	0.372	0.006	62.016

Notes. This table reports descriptive statistics on prices at origin location as well as at market destinations. Prices are converted in US dollars using the prevailing exchange rate in November 2016.

Table 2.17: ***Product and Industry match***

Agriculture, forestry and fishing	**Manufacturing**
Amarante	Airtime cell phone card
Banana	Candle
Brochette	Charcoal
Cabbage	Commercial Beer
Carrots	Cooking oil
Cassava flour	Detergent
Cassava leaves	Hair Dye Powder
Cassava root	Maggi bouillon cube
Celery	Mattress
Cooking banana	Mineral water
Dry beans	Nails
Egg	Sanitary napkin
Garlic	Shampoo
Green pea	Soap
Groundnut flour	Softdrink
Groundnut	Spaghetti
Imported rice	Sugar
Irish potato	Suitcase
Leek	Tea local
Maize	Toilet paper
Maize Flour	Toothpaste
Mandarine	
Milk	
Onion	
Pepper	
Sorghum	
Sorghum flour	
String bean	
Sweet Potato	
Tomato	

Notes. This table shows a correspondence between the products used for this study and the industries identified in the 2017 Census of Establishments published by the National Statistics Institute of Rwanda.

Table 2.18: ***Price gaps and distance: heterogeneity***

Dependent variable:	Price Gap				
	(1)	(2)	(3)	(4)	(5)
Log distance to origin	0.042[a]	0.078[a]	0.041[a]	0.123[a]	0.144[a]
	(0.012)	(0.023)	(0.010)	(0.040)	(0.037)
Log distance to origin × Non-manufactured products		-0.074[a]			-0.030[c]
		(0.024)			(0.016)
Log distance to origin × Imported			0.001		-0.016
			(0.013)		(0.013)
Log distance to origin × Food				-0.114[a]	-0.105[a]
				(0.041)	(0.040)
Observations	48,661	48,661	48,661	48,661	48,661
R-squared	0.22	0.22	0.22	0.22	0.22

Notes. This table presents the heterogenous effect of distance on price gaps. Column (1) represents the regression of price gap on the log of distance to origin. In column (2) we add an interaction term with a dummy for non-manufactured products. Columns (3)-(4) include respectively an interaction terms with dummies for imported and food products. In column (5) we take into account all the previous interaction terms. Standard errors are in parentheses and clustered at time-product level. Prices are converted in US dollars using the prevailing exchange rate in November 2016. All these regressions include a Time-Product fixed effects. [c] significant at 10 percent level, [b] at 5 percent level, [a] at 1 percent level.

Table 2.19: ***Indicators within industries***

Type of industry	**Agriculture, forestry and fishing**	**Manufacturing**
Pass-through rate	0.651	0.179
	(0.616)	(0.997)
Price Gap (Trading market pairs)	0.009	0.180
	(0.381)	(2.266)
Price Gap to origin percentage	0.030	0.045
	(0.302)	(1.237)
Adjusted price Gap	0.414	2.013
	(0.77)	(8.209)

Notes. This table shows the mean of some variable within industries. Standard errors are in parentheses. On average, prices gaps and pass-through rates are higher in manufacturing industry.

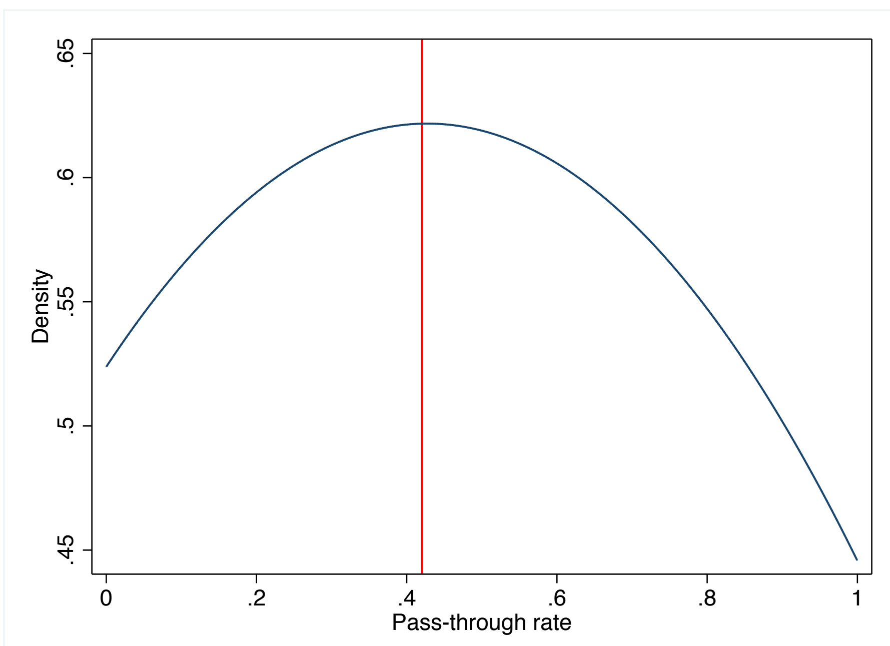
Table 2.20: ***Price gaps and connectivity***

Dependent variable:	Price Gap			
	(1)	(2)	(3)	(4)
Log Distance to origin location	0.042[a]	0.042[a]	0.047[a]	0.054[a]
	(0.012)	(0.012)	(0.013)	(0.016)
Log distance to origin × Intermediate connectivity		-0.003		-0.027[c]
		(0.008)		(0.015)
Log distance to origin × High connectivity			-0.013[b]	-0.030[b]
			(0.006)	(0.013)
Observations	48,661	48,661	48,661	48,661
R-squared	0.220	0.220	0.220	0.220

Notes. This table investigates the effect of connectivity on price gaps. Standard errors are in parentheses and clustered at time-product level. Prices are converted in US dollars using the prevailing exchange rate in November 2016. All these regressions include Time-Product fixed effects. [c] significant at 10 percent level, [b] at 5 percent level, [a] at 1 percent level.

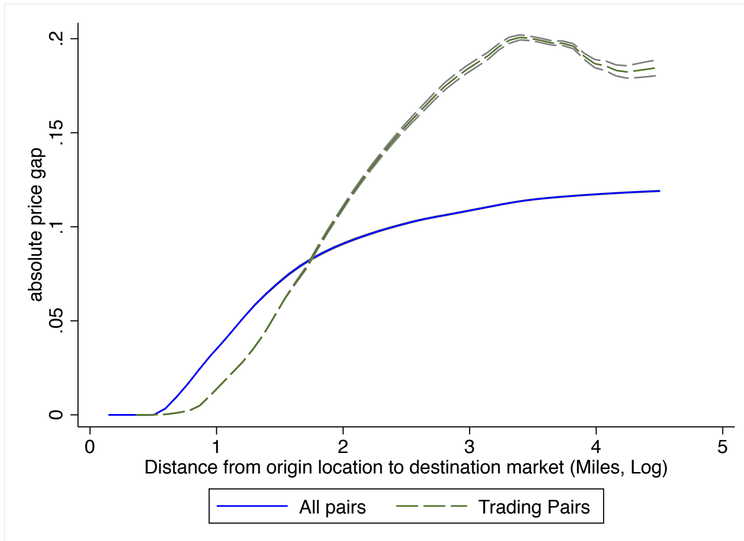
2.7.2 Additional figures

Figure 2.16: ***Kernel density of pass-through rates***



Notes. This figure presents the kernel density of estimated pass-through. Only pass-through between 0 and 1 are represented here. The red line represents the estimated pass-through with the highest density (0.42).

Figure 2.17: ***Price gaps and distance***



Notes. This figure displays the results of non-parametric of the absolute price gaps on distance using all pairs markets and trading pairs markets. The graph shows that relationship between absolute price gaps and distance is underestimated when we use all pairs markets instead of trading market pairs. We use a locally weighted polynomial (Epanechnikov kernel, bandwidth=0.5). 95% confidence intervals shown.

CHAPTER III

THE PRICE OF REMOTENESS: PRODUCT AVAILABILITY AND LOCAL COST OF LIVING IN ETHIOPIA

Abstract

We use the microdata underlying the Ethiopian CPI to examine the spatial dispersion in local prices and availability of 400 items across 110 cities. Remote cities face higher prices and have access to fewer products. Large cities also face higher individual prices but enjoy access to a wider set of products. To assess the welfare implications of these patterns, we build aggregate cost-of-living indexes. The cost of living is higher in remote areas, but not systematically related to population size. We then show spatial differences in the cost of living are a significant determinant of migration flows across Ethiopian regions.

Keywords: Cost of living; Product availability ; Remoteness; Ethiopia.
JEL classification: O12; R12; R23.

3.1 Introduction

Spatial inequality is pervasive within developing countries, and accounting for cost-of-living differences across places is key to a comprehensive view of its patterns (Ravallion and Van De Walle, 1991).[1] The literature so far has mainly focused on cost-of-living differences between urban and rural areas driven by price differentials of available products and services (see e.g. Ravallion and Van De Walle, 1991; Deaton and Tarozzi, 2000; Muller, 2002).

In this paper, we propose a fresh view on the geography of cost of living within a developing country - Ethiopia. First, we compute a measure of the cost of living that accounts for products and services availability on top of their price, which seems crucial because many products and services are locally missing in developing countries (see, e.g., FAO et al., 2018; WFP and CSA, 2018; FAO et al., 2019).[2] Second, instead of focusing on the urban-rural divide, we provide insights on spatial cost-of-living differences across cities, which allows us to relate the cost of living to two fundamental aspects of economic geography, namely, the population size and remoteness of cities. We find that cost of living is higher in remote places but that population size has no effect on it. Eventually, we show that properly accounting for spatial cost-of-living differences is important to understand the geography of internal migration flows.

The effect of economic geography on the local cost of living is not trivial, because several opposing forces operate. On the one hand, larger cities host higher-income households, which is likely to push prices up. On the other hand, consumers in small and remote cities may face higher trade costs, pushing the price of tradable goods up in those small and remote places.[3] Moreover, an under-studied component of the cost

1. See Ferré et al. (2012) and Young (2013) for nominal consumption and nominal income differences across urban and rural areas, and Gollin et al. (2017) for spatial differences in terms of non-monetary amenities.

2. In our data, about 35% of product-city pairs suffer from unavailability.

3. The role of transportation costs is likely to be prevalent in a developing country such as Ethiopia where internal transport costs are high (e.g., Atkin and Donaldson, 2015; Rancourt et al., 2014).

of living is product availability (Feenstra, 1994). Consumers in large and central cities may enjoy direct access to a greater number of products and services, which in the case of consumers' love for variety will push the relative cost of living down in those locations. What forces dominate is a quantitative question, but quite surprisingly, the literature has thus far provided little evidence on the spatial differences in terms of cost of living within developing countries.

We try to bridge this gap by estimating the impact of remoteness and population size on individual prices, product availability, and cost-of-living indexes at the city-level in Ethiopia. We leverage the microdata underlying the Ethiopian CPI to examine spatial dispersion in local prices and product availability across more than 100 markets (more or less equivalent to cities in our data - we refer to them as cities in the rest of the paper). We find that individual prices increase with city size and remoteness. Product availability, that is the probability that a product is available in a given city, also increases with city size, but it decreases with its degree of remoteness. Whereas remoteness has an unambiguous effect on the local cost of living by increasing prices and reducing product availability, the net benefit of living in a large city is less clear.

To make progress on this front, we aggregate our data and compute local cost-of-living indexes; we then examine how these indexes vary with city size and remoteness. We use two alternative measures of cost of living. The first measure is the spatial version of the Feenstra (1994) CES price index. Although this index was originally built to account for product availability over time, it has been adapted to measure spatial differences in the cost of living (Handbury and Weinstein, 2015). This index rests on the assumption that the price of unavailable varieties tends to infinity. We develop an alternative index that relaxes this assumption by assuming instead that, if a product is not available in a location, a consumer can still purchase it by traveling (thereby incurring travel costs) to the cheapest alternative location where the variety is available. [4]

4. In an inter-temporal context, the assumption that a consumer at date t cannot consume goods that disappeared between $t-1$ and t, and that consumers in $t-1$ cannot enjoy varieties that appear at date t is reasonable. However, although consumers cannot travel over time, they can travel over space - and even more so within a country. Therefore, the assumption that consumers in a location cannot benefit from varieties available in other locations may be more debatable when computing a spatial price index.

Armed with these two alternative measures, we show the cost of living is significantly higher in remote locations. However, we do not find systematic evidence that population size affects the cost of living, which suggests the price premium paid for available products in large cities is entirely compensated by access to a wider array of products. These results are robust to computing the cost of living with different methods and different weighting schemes. They also are not sensitive to the introduction of various control variables.

We then use data on income per capita in Ethiopian regions to show that accounting for cost-of-living differences is important to have an accurate view of spatial income inequality. The ranking of regions based on their income per capita adjusted for the price index changes significantly if one includes product availability in the computation of the cost-of-living index.

We eventually investigate how spatial differences in terms of cost of living shape migration flows across Ethiopian regions. We run gravity-like equations and find that cost-of-living differences are a significant determinant of internal migration flows. All else equal, people prefer moving to low-cost-of-living regions. Moreover, accounting for differences in terms of cost of living substantially improves the predictive power of the gravity model (the adjusted R-square of the regression rises by more 20% when cost of living is controlled for). Also note that using a standard price index that neglects product availability does not improve the predictive power of the gravity model. Last, accounting for the cost of living in the destination region completely cancels out the negative and significant impact of remoteness on migration; hence, higher prices and reduced product availability can entirely explain the lower attractiveness of remote regions for migrants. Having an adequate measure of local cost of living that accounts for both prices and product availability is thus crucial to understand internal migration patterns in a developing country such as Ethiopia.

Related literature. We contribute to four strands of the literature. First, we contribute to the literature on spatial differences in terms of cost of living in developing countries (see, e.g., Ravallion and Van De Walle, 1991; Deaton and Tarozzi, 2000;

Muller, 2002; Timmins, 2006; Ferré et al., 2012). We see three important contributions of our work with respect to existing studies: (1) Our measure of cost of living accounts for both the availability of products and services and for their price when they are available, whereas existing papers usually focus on the latter; (2) By leveraging the micro-data underlying the consumer price index in Ethiopia, our analysis has a greater spatial and industrial coverage than existing studies that usually focus on a narrower set of products and villages/regions; (3) Instead of emphasizing the urban/rural divide, we analyze cost-of-living differences across cities and relate them to economic geography, namely, population size and geographic remoteness of cities. [5]

Second, a couple of recent papers examine prices and product availability in the context of Ethiopia (Gunning et al., 2018; Krishnan and Zhang, 2018). Gunning et al. (2018) demonstrate that households in remote villages of Ethiopia have access to a lower variety of goods. Our paper differs from these other papers along three important dimensions. First, the focus is different. These papers examine individual prices and product availability in remote villages. Our analysis instead covers Ethiopian main cities. Second, our data allow us to compute local price indexes and thus to directly compare the welfare across cities - and relate welfare differences to city size and remoteness. [6] Third, we are the first to study the implications of prices and product availability differentials for migration decisions.

Third, our paper contributes to the literature on the measurement and determinants of price indexes in the presence of missing varieties. Seminal papers in this literature have

5. Ethiopia is an ideal laboratory to understand the impact of intra-national isolation on price variations across space in a developing economy context. With 100 million inhabitants, it is the second most populated country in Africa and the largest landlocked country in the world. The population is spread unevenly in the country. While 3% to 5% of the population live in the metropolitan area of Addis-Ababa, about 20 cities spread throughout the country have population with 5-digit figures. Like other countries in the region, Ethiopia has experienced strong GDP growth since the beginning of the 2000s and a sharp improvement in several development indicators.

6. The literature also examines the impact of remoteness on outcomes other than prices or product availability. For instance, Dercon and Hoddinott (2005) show that better access to market towns allows rural households to buy their inputs at a lower price and to sell their outputs at a higher price in Ethiopia; in the same vein, Aggarwal et al. (2018) find poor market access implies a poor harvest output in rural Tanzania.

developed methods to measure the costs and benefits of disappearing and appearing varieties over time (Feenstra, 1994; Hausman, 1996; Diewert and Feenstra, 2019). A few papers apply this method to a spatial context, and explore the link between the cost of living and city size (see, e.g., Handbury and Weinstein, 2015; Feenstra et al., 2019, using U.S. and Chinese data, respectively).[7] We complement this literature by showing that, in a country such as Ethiopia, remoteness has a stronger impact on spatial variations in terms of cost of living than city size. We also propose an alternative spatial price index that relaxes the standard assumption that products unavailable in a city cannot be purchased in surrounding cities by traveling.

Last, our paper contributes to the literature on the determinants of internal migration flows. Different factors have been shown to drive migration, including real wages (see, e.g., Krugman, 1991a), access to domestic and/or foreign markets (Crozet, 2004; Hering and Paillacar, 2016), access to public and private amenities (Bryan and Morten, 2019; Lagakos et al., 2020; Fafchamps and Shilpi, 2013), earnings differences (Young, 2013), conflicts (Kondylis and Mueller, 2014), or differences in housing prices (Monras, 2018; Hsieh and Moretti, 2019). We show that cost-of-living differences (particularly those driven by access to a wider set of products and services) are an important driver of migration flows across Ethiopian regions.[8]

The rest of the paper is organized as follows. Section 3.2 describes the data used in the empirical analysis. Section 3.3 presents the results on the impact of remoteness and population density on individual prices and product availability. Section 3.4 introduces our cost-of-living indexes, and presents our analysis of the determinants of the spatial dispersion in the cost–of-living across Ethiopian cities. Section 3.5 presents the results on the impact of cost-of-living differentials on migration across Ethiopian regions. Section 3.6 concludes.

---

7. See also Matsa (2011) on the link between competition and inventory shortfalls.

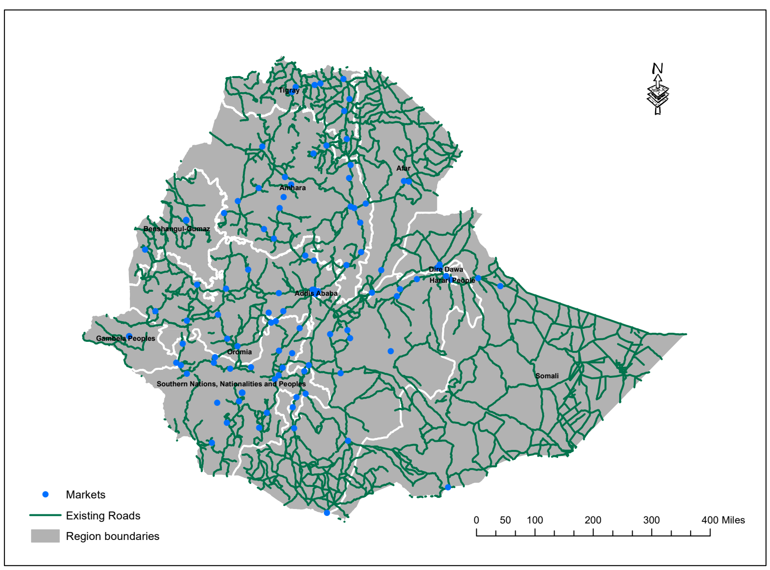
8. Jedwab and Storeygard (2017) show that better market access has spurred population growth in sub-Saharan cities. The effect is more pronounced for small and remote cities. Our results suggest that these differences in population dynamics might partly channel through increased availability and lower prices of products and services in these places.

3.2 Data

This section presents the different datasets used in the paper, and discusses their strengths and limitations. Crucial to our analysis are the consumer price data underlying the Ethiopian CPI. These data are described in the next section. We then describe the economic, demographic, and geographic variables used throughout our empirical analysis. The calculation of price indexes requires information on how consumers allocate their expenditure across the various products, and on the elasticity of substitution between these products. We describe these variables at the end of the section.

3.2.1 Consumer price index data

Figure 3.1: ***Markets covered by the Ethiopian CPI***



Source. Individual price data are collected by the Ethiopian Central Statistical Agency (CSA) to build the national Consumer Price Index.[9] We work with the data for the year

9. Figures 3.13, 3.14, **??** and 3.15 in the appendix compare price data from CSA and price data from

2015. The price quotes of more than 420 products are collected every month by enumerators in 117 markets. The markets are urban centers (cities or towns) in each woreda (a woreda being the smallest administrative division with local government in Ethiopia). Addis-Ababa, the Ethiopian capital city, is split into 12 markets. The different markets are mapped in Figure 3.1. In each market, enumerators survey a pre-determined sample of outlets (1 to 3 per location) every day during the first two weeks of each month. Outlets comprise a representative sample of open markets, kiosks, groceries, butcheries, pharmacies, super markets, and so on. Enumerators are asked to find precisely defined products and to report the product as missing if they cannot find it. When they find the product, they determine its typical price (after bargaining) by interviewing both sellers and consumers.

Based on the CPI calculation methodology by the CSA. The products selected are those most popular in major cities across the country. All regions are covered to represent tastes and preferences across the country The problem of the bias of selected products in favor of the capital arose sharply in the early 1960s when the agency only covered products consumed in the capital. Over time this bias subsided. And we are confident that for the year 2015 that we are using here, the selection of products takes into account national preferences and not only those of the inhabitants of the Capital. (Atkin and Donaldson, 2015).

Extraction. Every month, the CSA releases a 200-page document reporting the price of every product in each of the 117 cities. We extract this information from pdf files for all the months of year 2015. We then manually check the obtained dataset. We change the label of some products or product categories that appear to be obviously wrong. The resulting dataset is available on our personal websites.

Description. The survey covers 427 products and services that can be grouped into 12 major groups and 55 categories. [10] These products and services include food products

the World Bank Living standards measurement study (LSMS)

10. The list of products is presented in Table 3.12 in the Appendix.

such as bread and cereals, but also, among others, clothing and footwear products, household equipment products, or hair-cuts and restaurants. The product descriptions range from barcode-like data with brandname products ("Coca-Cola bottle 300c"), to very specific products without a brand (e.g., "bed sheet (Patterned Kombolcha) 1.90m x 2.50m"), to more generic product categories (e.g., "sorghum yellow, kg" or "rice imported, kg").

Outliers and yearly price. Prices are collected every month. In 2015, we thus have a maximum of 12 price quotes for a given product within a given location. We have detected a few outliers in the data. A visual check suggests two main types of outliers exist. First, the price might be abnormally high because, for instance, the marker for decimals was forgotten. Second, enumerators sometimes reported the monthly price of the product at the wrong line. For instance, in Akaki, the product "VCD-Player (Mayato Japan)" is missing every month of 2015 except the month of July. That month, the reported price is 5 Birrs. However, in the file, the product below "VCD-Player (Mayato Japan)" is "VCD Cassette rent". In Akaki, this product is available every month at the price of 5 Birrs, except for the month of July, where it is missing. Moreover, the median price of the "VCD-player (Mayato Japan)"' in markets where it is available is around 1,000 Birrs. Thus, an obvious mistake exists where the price of the "VCD Cassette rent" in Akaki was reported at the wrong line for the month of July.

To systematically detect and delete outliers, we drop monthly observations for which the price is 5 times higher or 5 times lower than the median price of the product across markets. We thereby exclude 469 observations out of more than 300,000. Note that with this procedure, we do not automatically create missing varieties, because, as explained below, our definition of "missing" is computed at the annual level.

Because we are interested in price differences across locations (and not over time), and to alleviate potential remaining issues related to misreporting, we take the median price per product and location computed over the months the product is not missing. In the original dataset, Addis-Ababa is split into 12 markets. We merge these markets into a single one by considering the median price across the 12 markets. This approach leaves us with information on 427 products across 106 cities.

Missing products. A missing price in our original dataset means the enumerators were unable to find the product in that city in that month. We follow Atkin and Donaldson (2015) and assume the product was not available if the price is missing. Our definition, however, is stricter than theirs because we require the product to be missing every month of 2015 to consider it unavailable. In our final dataset, a product is considered missing for about 35% of the product-location pairs.

Prices and vertical differentiation. Given our interest in building a local price index that best reflects the cost of living for consumers, we consider that using all the products and services in the dataset (including those with a less precise description) is an upside of our study. Not restricting the sample to barcode products matters because many products in the consumption basket of consumers have no barcode, especially in a developing country such as Ethiopia. This choice comes at the cost that for some products, price differences across locations might partly reflect differences in terms of quality. Actually, we identify in our dataset 31 products (out of 427) whose description clearly mentions a brand and a unit that make them as close as possible to barcode products. [11] As Figure 3.2 shows, across cities, the coefficient of variation of barcode products tends to be, as expected, lower on average than the one observed for the other products within the same product category. Still, Figure 3.2 shows the prices of barcode products also exhibit substantial spatial variation, suggesting the latter cannot be reduced to spatial variation in the quality of the available varieties.

Moreover, in most of our specifications, we use use a proxy of local prices purged from spatial differences in terms of quality; more precisely, we regress the price quote of product p in city c p_{pc} on the local per-capita income, and when the coefficient we obtain is statistically significant at the 5% level, we take as a proxy for the price of product p in city c the residual of this regression. The idea is that in the presence of non-homothetic preferences, wealthier consumers are more likely to buy higher-quality varieties; consequently, price quotes and per-capita income should be positively correlated at the local level. Note that in doing so, we also control for the possibility of wealthier consumers having a higher willingness to pay than poorer ones. The regres-

11. See the list in Table A.3 in the Appendix.

sion of price quotes on local per capita income is run separately for the 55 product categories and includes product-type fixed effects. The value of the coefficient exhibits a fair amount of variation across product categories (average equal to 0.07, standard-deviation equal to 0.14). The correlation is positive (resp. significantly positive) in 39 (resp. 21) of these categories, and negative (resp. significantly negative) in only 16 (resp. 6) product categories. Hence, the correlation is statistically significant for less than half of the product categories only.

Finally, we are interested in the spatial differences in terms of cost of living and their consequences on the internal migration patterns, which could well be driven by the type of varieties consumers have access to; one could thus argue the composition of the set of available varieties in terms of quality is part of the heterogeneity we want to capture. [12] For this reason, we also present the results with non-purged price quotes.

Overall, our results are similar regardless of the proxy we use for local prices, and thus we believe vertical differentiation is not a first-order issue for our study.

Product classification. We test for possible heterogeneity in the macro-determinants of local prices across specific broad categories of products.

More specifically, we manually identify services from the product description. Out of 427 products, 44 are services (10%), including, for instance internet services, haircuts or airplane tickets to Addis-Ababa.

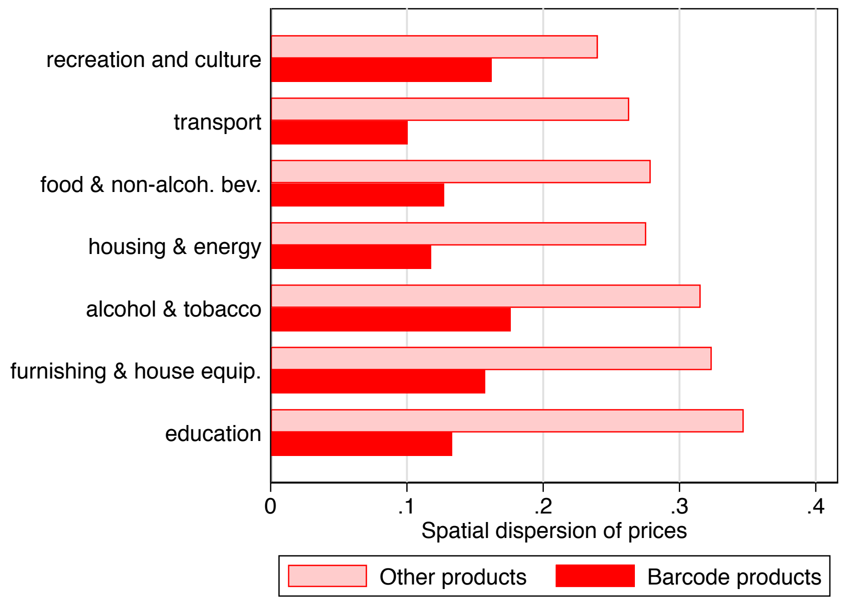
We also use the products' description to identify imported products. For some items, the description indicates the product is imported (e.g., "socks (cotton) imported"). For others, the country of origin is indicated (e.g., "under wear China"). We tag all the products whose description contains a word related to import as an imported product. [13] In our sample, 8% of the products are imported based on this definition (keeping in

---

12. The composition of varieties can be seen from two angles: from the angle of the type of variety but also the angle of the quality of the variety. Purge the price of the varieties from the revenue component allows controlling for the spatial difference in the quality of the varieties.

13. More specifically, a product is imported if its description includes at least one word among the following list: imp., import, imported, China, Japan, England, India.

Figure 3.2: ***Coefficient of variation of prices across space***

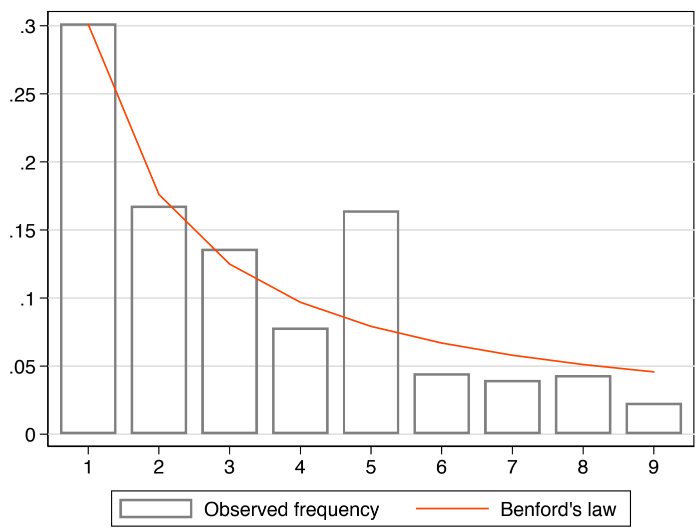


mind that it might not provide an exhaustive list of imported varieties). Finally, we define as "food products" those products in the major group "Food and non-alcoholic beverages".

Reliability of the price data. Because inflation is a major political concern in Ethiopia, one may worry the price-quote data are manipulated for political reasons. Note that 2015 was not a year of hyper inflation limiting the political motivations for price manipulation. To further assess the reliability of the price quotes, we test whether the Benford law holds in our data. The Benford law states that the frequency distribution of the first digit of numerical data is stable across samples (the frequency of number 1 is 30%, number 2 is 17%, and number 9 appears as the first digit in only 5% of the occurrences). Researchers have used deviations from this law to detect reporting issues in survey data (e.g., Judge and Schechter, 2009; Demir and Javorcik, 2018). The underlying idea of this test is that manipulating the data and still making them fit the Bendford distribution is difficult.

Figure 3.3 presents the frequency distribution of the first digit of price quotes in our data together with the Benford distribution. The observed frequencies fit with frequencies given by the Benford law. [14] Hence, the distribution of the first digit of price quotes in our data is consistent with the Benford law, suggesting price manipulation is not a major issue here.

Figure 3.3: ***Benford law for price quotes***



Each bar in the diagram represents the observed frequency of a number of the first digit of price quotes. The red curve represents the frequency of a price digit given by the Benford's law.

3.2.2 Demographic and economic data

We complement our price data with various variables extracted from different waves of the Ethiopian Census. For some locations, we did not find the required information in the Census, and we had to look for other data sources. We present the details below.

Population. The 2015 population of the urban centers considered in the analysis comes from the population projection figures for 2015 made by the CSA based on the 2007 population and housing census of Ethiopia (see CSA, 2013). For 25% of the urban

14. We also did the Kuiper test to assess the deviation of our data from the Benford law. The mean of the Kuiper statistic is 0.0264 which is over the 1% threshold to reject the null hypothesis that the observed distribution deviates from the Benford law.

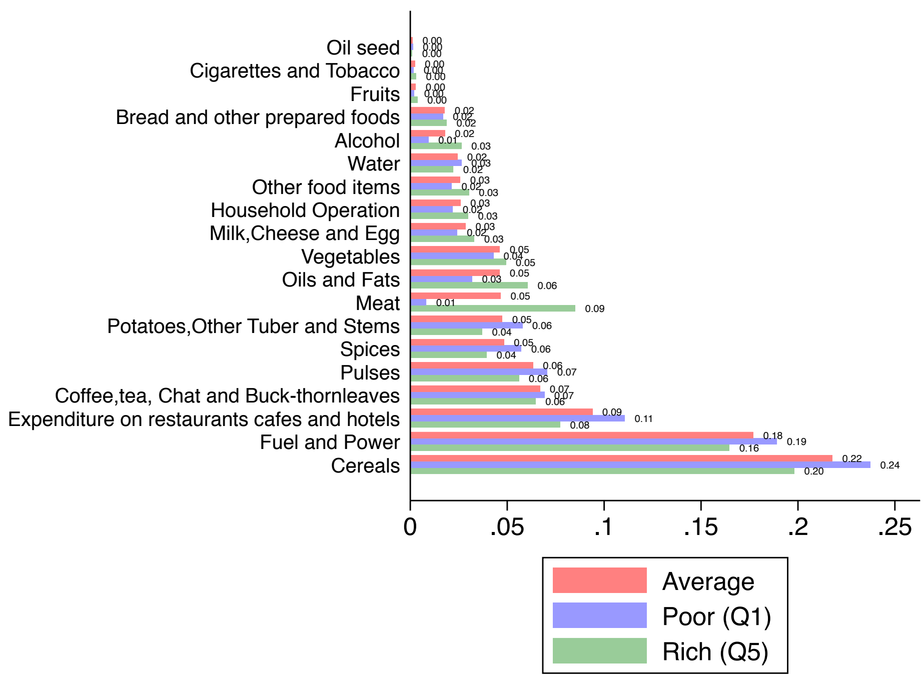
centers the CSA does not provide any projection. For these locations, we rely on projection figures for 2015 made by the Ethiopian Ministry of Water and Energy in 2011 as part of the urban water-supply universal access plan (see Ministry of Water and Energy, 2011). The size of the cities in our dataset ranges from 764 inhabitants (Deri) to 3,273,000 (Addis-Ababa). In our sample, 63% of cities have less than 30,000 inhabitants. These cities are defined as "rural".

Income. To measure per-capita income, we use data from the 2015/2016 Ethiopian Household Consumption Expenditure survey (HCE) conducted by the CSA (see CSA, 2018a). Per-capita income is measured by region (Ethiopia has 11 regions in total), with two separate measures for urban and rural areas in each region. It is thus available at a more aggregate geographic level than prices. It measures people's total expenditures in Birr and it also includes consumption in kind. It is available by quintiles in each region and each type of area. The average income in the first quintile is 3,900 Birr, whereas it reaches 24,200 Birr for households in the fifth quintile. We use average income per capita as a measure of income in our analysis.

Ethnic diversity. Data on ethnic diversity comes from the 2007 Population and Housing Census of Ethiopia conducted by the CSA (see CSA, 2008). We compute the share of each ethnic group (93 in the country) in the total population of every woreda. Similar to Schiff (2015), our measure of ethnic diversity is the inverse of a Herfindahl index computed with these shares. At the country-level, Oromos are the most represented ethnic group, with 34.6% of the total population.

Expenditure weights. Aggregate price indexes require information on expenditure weights across product categories. We use expenditure weights reported in the 2011 Ethiopian Household Consumption Expenditure survey (HCE) conducted by the CSA (see CSA, 2012). We have expenditure weights for 19 semi-aggregated categories. These categories are related to food and daily expenditures. The products for which we do not have weights are thus those that are purchased less frequently. In particular, we do not take into account housing. As compared to the standards in the literature, we do not think this issue is a major one here. Indeed, several papers on the US use housing prices alone as a measure of the local cost of living (Moretti, 2013; Monras, 2018), with housing expenses representing 33% of US households' expenses. Here, the products and services for which we have reliable information account for 62.5% of overall expenses of Ethiopian consumers. Moreover, housing only represents 9.9% of total household expenditures in Ethiopia (see CSA, 2012). We are thus confident the set of products and services for which information on expenditure weights is available allows for a proper measure of local cost of living. Information is available separately for rich and poor households. As depicted in Figure 3.4, for both poor and rich, Cereals represent the highest expenditure share, and Oil seed represents the lowest.

Figure 3.4: ***Weights in terms of expenditures***



Note: Each bar represents the share of an item in the total expenditures dedicated by a household for the selected products. The blue bars represent the weights for poor consumers, the green bars represent the weights for rich consumers, the red bars represent the average weights.

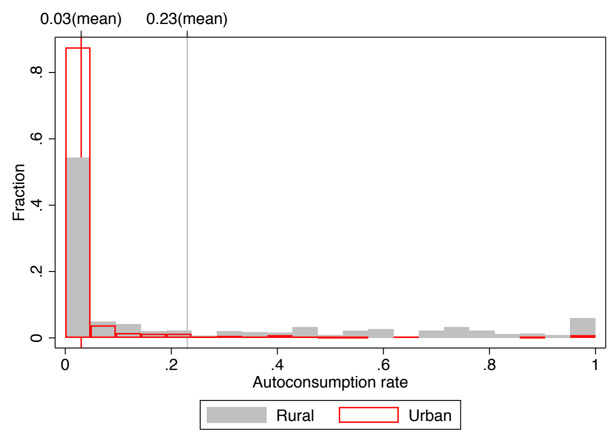
Elasticities of substitution. To compute the CES price index described in section 3.4, we will need elasticities of substitution between varieties within semi-aggregated categories. For most of the 19 semi-aggregated categories, we use import demand elasticities estimated by Broda et al. (2017). More precisely, we assign HS codes for which estimated elasticities are available to one of the 19 semi-aggregated categories for which we have expenditure shares, and take the average elasticity. Ethiopia is not available in this dataset. We thus use the elasticities estimated for Togo and Madagascar instead because these countries are the closest to Ethiopia in terms of GDP per capita and linguistic proximity for which information is available. We try two options: consider the elasticities of Togo, or take the average elasticity between Togo and Madagascar. This has little impact on the results. Cigarettes and Tobacco have the highest substitution elasticity (84.83), and Alcohol has the lowest (1.59). These elasticities are for tradable goods. For restaurants, we use the elasticity of 8.8 estimated by Couture (2014) for the US. [15]

Auto-consumption. In some robustness checks, we account for auto-consumption, which could be correlated with both prices and product availability on the one hand, and city size and remoteness, which are our variables of interest, on the other. We compute data on auto-consumption using the third wave (2016) of the Ethiopian socioeconomic survey conducted by the CSA and the World Bank Living Standards Measurement Study-Integrated Surveys on Agriculture (LSMS-ISA) team (see CSA, 2018b). Obviously, auto-consumption is a concern for food products mainly, and information on auto-consumption is thus available only for these products. We define auto-consumption as the share of household consumption (in terms of quantity) that comes from its own production. These data are available for each product-region-area triplet, where area refers to urban versus rural areas. “Dairy products” (e.g., cow milk) and “Cereals unmilled” (e.g., sorghum) are the product categories with the highest share of auto-consumption in the country (average of 38% and 24%, respectively). Not surprisingly, the auto-consumption rate is higher in rural areas with an average of 23% compared to only 3% in urban areas. Note, however, that these averages hide large discrepancies at the household level, as shown in Figure 3.5. Auto-consumption is below 5% for almost all of the households living in urban districts. For households in rural districts, the distribution of auto-consumption shares is bi-model with a (large) peak

15. Unfortunately, no estimate is available for Ethiopia or for a close developing country. However, this 8.8 for restaurants is close to the average elasticity of substitution measured across the product categories taken into account in the price index, which is equal to 8.84.

around 0 and a (small) peak close to 1.

Figure 3.5: ***Auto-consumption***



Note: Auto-consumption rates for food products computed from the 2015 household survey (LSMS data). The plain gray bars represent the distribution of auto-consumption shares for households living in rural districts. The red empty bars are for households living in urban districts. The red and grey vertical lines indicate the average auto-consumption rate for urban and rural households, respectively.

Migration data. We use migration data to investigate the effects of spatial differences in terms of cost of living on internal migration flows. We use migration data from the 2013 national labour force survey (see CSA, 2014). In this context, migrants are individuals who moved from a region to another, less than five years prior to survey data collection. Internal migration is measured by asking individuals about the number of years they have continuously lived in the the region (rural or urban). If the answer was less than five years, questions are asked about where the person come from , in which region the person lived before and the reason for leaving (see, e.g., Bundervoet, 2018).

Employment opportunities and surface area of regions. To investigate the determinants of migration flows, we need a measure of regional employment opportunities at the destination. For this purpose, we use 1 minus the regional unemployment rate as a proxy for employment opportunities. We use data on unemployment rates from the

2005 national labor force survey (see CSA, 2006).[16]

The regions in Ethiopia have different sizes. Beyond total population size, surface area (and thus population density) could matter for migration decisions. We use data on the surface area of regions from the 2007 population and housing census of Ethiopia.

3.2.3 Geographic variables

A key part of our analysis is about the relationship between distance to other cities (or remoteness) and the local cost of living. We build our distance measures as follows.

Distance and travel time between cities. We first need to assess how far each city in our dataset is from the other cities. Distance is computed thanks to the Stata package GEODIST. Travel time is computed using the Stata package GEOROUTE (Weber and Péclat, 2016). The travel times to other cities provided by this package are missing for 10 cities. For these cities, we predict travel time based on bilateral distance.[17] We use a similar strategy to compute the distance and travel time to the capital city (Addis-Ababa) and to the main international trade corridor (Kombolcha, through which shipments from and to Djibouti transit).[18]

Remoteness. For each city, we aggregate the information on bilateral distances with the 105 other cities by building a remoteness index, which is the average travel time to the other Ethiopian cities:

$$\text{remote}_c = \frac{1}{105} \sum_{j \neq c} \text{travel time}_{cj},$$

16. In the migration literature, previous employment rates are often used as determinant of current migration flows (see, e.g., Crozet, 2004). Consequently, we use 2005 unemployment rates to compute regional employment rates.

17. Travel time is regressed on a polynomial of degree 7 of distance as well as fixed effects for origin-destination pairs of regions. For the 8,730 pairs of cities for which we have both forms of information, the $R-squared$ is close to 90%. Travel time is then predicted for all the city-pairs involving the 10 destinations for which information on travel time is missing.

18. But in this case, a linear fit with region fixed effects is enough.

where remote_c is the remoteness index for city c, and travel time_{cj} is the travel time between cities c and j. In robustness checks, we use a population-weighted version of this remoteness index.

Travel costs. Part of our analysis relies on the travel cost paid by a consumer to buy a product that is not available in her own town. To compute travel costs between towns c and j, we use the following formula:

$$\text{travel\_cost}_{cj} = \text{distance}_{cj} \times \text{cost\_per\_km}.$$

The previous paragraph explains the computation of distance between any pair of cities. More difficult to obtain is a reliable measure of the travel cost per kilometer. Estimates of transport costs in Ethiopia are available, but they are based on the cost of transporting products by truck from the location of production or port of entry to a given destination. We are instead interested in the cheapest transport mode for a consumer who wants to travel to a location where the variety is available. Interviews with people from the field suggest the cheapest mode of transport across Ethiopian cities is van taxis and intercity buses. We use the intercity bus fare of 0.1235 birr/km reported in the management of the commercial road transport survey (see Addis Ababa Chamber of Commerce and Sectoral Associations, 2009). [19]

From this figure, we are able to compute the price of consuming in city c a product j from city c' as follows:

$$p_{jcc'} = p_{jc'} + 2\text{travel\_cost}_{cj}.$$

3.3 Spatial dispersion of individual prices and product availability

In this section, we assess the "macro" determinants of local individual prices and product availability. More specifically, we investigate whether city-size and geographic remoteness affect the price of a product when it is available, and the probability that

19. Note this cost is not product-specific, because it measures the cost for a consumer to travel to a location where the product is available.

this product is available in that city. The regression we estimate is thus the following:

$$\text{y}_{pc} = \alpha \text{pop}_c + \beta \text{remote}_c + \omega_p + \varepsilon_{pc}, \tag{3.1}$$

where y_{pc} is either the log median price quote of product p in city c in 2015 or a dummy equal to 1 when p is missing every month of 2015 in city c. City-size is measured by the log population in c. Our measure of remoteness is the log (unweighted) average time to reach all the other Ethiopian cities in the database from c. Finally, ω_p is a product fixed effect, and ε_{pc} is the error term.

Before getting to the baseline results, we provide in Table 3.1 a variance decomposition of local prices and of the dummy identifying product unavailability. The product fixed effects explain most of the dispersion in prices, which simply reflects that different products have different average prices. Once we take into account the product fixed effects, the share of the remaining variation explained by city fixed effects is equal to 6.37%. The last row of Table 3.1 shows once product-region fixed effects are accounted for the R-squared rises to 32.86%. The picture is qualitatively similar (though quantitatively different) for the dummy identifying missing products. Thus, the correlation between local prices/product availability and city size and remoteness might not be the same across products. We discuss this heterogeneity at length in our analysis of the macro-determinants of local prices. This heterogeneity will also be important when aggregating the micro-data to examine the effect of city-size and remoteness on local price indexes.

3.3.1 Baseline results

Table 3.2 presents our baseline results. As can be seen from regression (1), remoteness and city-size are significant determinants of the local prices of available products. Large and more remote cities are more expensive than the others. When we control for the product category-level correlation between prices and income per capita (which should capture quality differences across cities), the coefficient on our remoteness proxy is

Table 3.1: ***Variance decomposition***

	Log p_{pc}	$\mathbb{1}_{\text{missing}_{pc}}$
Product fixed effect	97.4%	35.5%
City fixed effect	6.37%	12.07%
Product-Region fixed effect	32.86%	21.85%

Notes: p_{pc} is the price quote of product p in city c and $\mathbb{1}_{Miss_{pc}}$ is a dummy equal to 1 if product p is missing in city c. The table reports the R^2 of regressions including the fixed effects mentioned in the first column. For city and product-region fixed effects, the dependent variable is the price or the dummy $\mathbb{1}_{Miss_{pc}}$ net of the product fixed effects.

even higher (even though not statistically different from the one in the first column), whereas the coefficient on city-level population is reduced but remains positive and highly significant. The pattern of the coefficient on population size is consistent with large cities also being the wealthiest ones. Several explanations could account for this positive association between price and city-size. The distribution costs could be higher in larger cities due to fancier distribution chains or higher land rents for example; products sold in larger cities could also be less likely to be produced locally, especially food products.

Column (3) displays the results on product availability. The probability that a product is missing is significantly higher in remote cities and significantly lower in large cities, with elasticities equal to 10.5% and -7.2%, respectively.

Table 3.2: ***Macro determinants of micro prices and availability***

	Log p_{pc}		$\mathbb{1}\text{missing}_{pc}$
	(1)	(2)	(3)
Ln Remoteness	0.119^a	0.134^a	0.105^b
	(0.030)	(0.031)	(0.052)
Ln Population	0.040^a	0.028^a	-0.072^a
	(0.004)	(0.005)	(0.009)
Per capita income	No	Yes	No
Observations	28,183	28,183	42,506

Notes: The dependent variables are the log price (column 1), the log price net of its product category-level correlation with the local per-capita income (column 2), and a dummy equal to 1 if the product is unavailable in the city (column 3). These dependent variables are defined in the *product* × *city* dimension. Standard errors clustered at the city level in parentheses, [a], [b] and [c] respectively denote significance at the 1, 5, and 10% levels.

3.3.2 Robustness checks

In Table 3.3, we check the robustness of our baseline empirical regularities. More specifically, we propose four types of checks. First, the Ethiopian population is composed of various ethnic groups, but many cities are characterized by the presence of a (very) majoritarian one. If ethnic groups have very specific tastes, more ethnically diverse cities could be cities where varieties are more likely to be available (e.g., Schiff, 2015), and this diversity could also affect the price at which they are sold. We thus control for the inverse of a Herfindahl index based on the share of the various ethnic groups in the population. Second, remoteness might not only be related to distance to all other cities, but more specifically to distance to the capital city, Addis-Ababa, or to the most important commercial routes, namely, Kombolcha, which is on the corridor to Djibouti. We thus control for distance to these two cities. Third, we use a weighted average distance to all other Ethiopian cities (using population as weights) as an alternative measure of remoteness. Finally, we drop Addis-Ababa from the regression sample.

In a nutshell, controlling for ethnic diversity does not change anything. The distances to Addis-Ababa and to the import corridor are not significant for the price of the available varieties or for the probability that a product is missing, and their introduction does not affect the coefficient on population size and remoteness. The weighted average distance to other cities is positively related to prices but not significantly related to product availability. What matters for product availability is thus how far a city is from the other cities, not from the large ones only. Finally, dropping Addis-Ababa from the sample does not significantly change the results.

Table 3.3: ***Macro determinants of micro prices and unavailability: Robustness checks***

	Log p_{pc}					$\mathbb{1}_{\text{missing}_{pc}}$				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Ln Remoteness	0.134^a	0.136^a			0.141^a	0.105^b	0.103^c			0.116^b
	(0.031)	(0.032)			(0.031)	(0.052)	(0.053)			(0.052)
Ln Population	0.028^a	0.028^a	0.027^a	0.027^a	0.026^a	-0.072^a	-0.072^a	-0.068^a	-0.072^a	-0.075^a
	(0.005)	(0.005)	(0.006)	(0.005)	(0.005)	(0.009)	(0.009)	(0.010)	(0.009)	(0.010)
Ln Ethnic diversity index		-0.008					0.007			
		(0.014)					(0.026)			
Ln Travel time to import corridor			-0.012					0.017		
			(0.007)					(0.010)		
Ln Travel time to Addis-Ababa			0.008					0.005		
			(0.008)					(0.010)		
Ln Remoteness (weighted)				0.054^b					0.058	
				(0.027)					(0.038)	
Per-capita income			Yes					No		
Addis-Ababa	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	No
Observations	28,183	28,183	27,697	28,183	27,810	42,506	42,506	41,704	42,506	42,105

Notes: The dependent variables are the log price (columns 1 to 5) net of its product category-level correlation with the local per-capita income and a dummy equal to 1 if the product is unavailable (columns 6 to 10). These dependent variables are defined in the *product* × *city* dimension. Standard errors clustered at the city-level in parentheses, ^{a}, ^{b} and ^{c} respectively, denote significance at the 1, 5, and 10% levels.

In the context of a developing country such as Ethiopia, another important issue that could undermine our benchmark results is auto-consumption. For food products in particular, households sometimes produce their own consumption. The phenomenon is particularly important in rural areas (see section 3.2.2). If products are produced directly by those who consume them, they might be unavailable on the market without necessarily being unavailable for consumption. For a subset of food products in our database, we have information on the share of auto-consumption in total households' consumption by region and type of area (rural or urban). After reproducing the benchmark results for the subset of observations for which auto-consumption is available, we thus directly introduce auto-consumption in the regression. Table 3.4 shows auto-consumption is not statistically significantly related to prices or product availability, and controlling for it does not affect the results.

Table 3.4: ***The role of auto-consumption***

	Log p_{pc}			$\mathbb{1}_{\text{missing}_{pc}}$		
	(1)	(2)	(3)	(4)	(5)	(6)
Ln Remoteness	0.134^a	0.104^b	0.102^b	0.105^b	0.175^a	0.176^a
	(0.031)	(0.041)	(0.041)	(0.052)	(0.039)	(0.039)
Ln Population	0.028^a	0.032^a	0.028^a	-0.072^a	-0.047^a	-0.044^a
	(0.005)	(0.007)	(0.008)	(0.009)	(0.006)	(0.007)
Auto-consumption rate			-0.046			0.023
			(0.029)			(0.022)
Per capita income		yes			no	
Observations	28,183	8,672	8,672	42,506	12,370	12,370

Notes: The dependent variables are the log price (columns 1 to 3) net of its product category-level correlation with the local per-capita income, and a dummy equal to 1 if the product is unavailable in the city (columns 4 to 6). These dependent variables are defined in the *product* × *city* dimension. Standard errors clustered at the city-level in parentheses, [a], [b] and [c] respectively denote significance at the 1, 5, and 10% levels.

All in all, our benchmark results are thus very robust: large cities are places where products and services are more expensive but also more likely to be available, whereas remote cities are places where products and services are more expensive and less likely to be available, in line with the idea that these cities are more difficult to cater to.

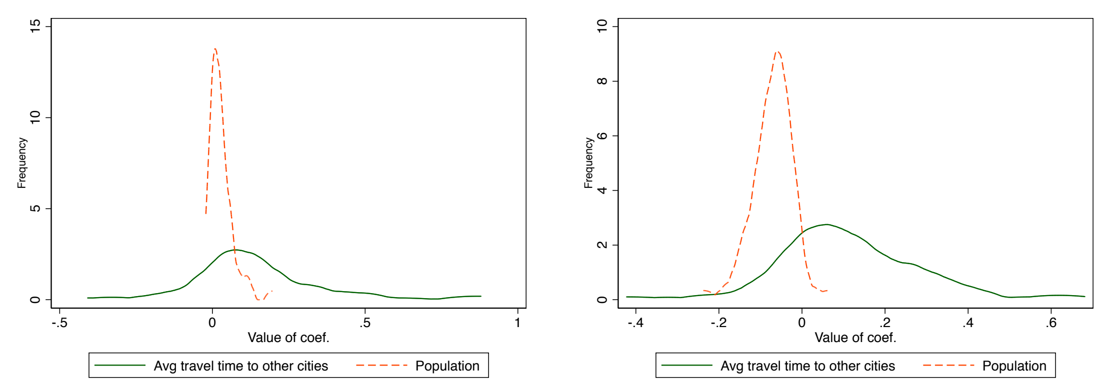
3.3.3 Heterogeneity

To go beyond the average correlations of our baseline results, we run the regressions of columns (2) and (3) of Table 3.2 separately for the 55 product categories present in our sample. We summarize the results in Figure 3.6. Panel (a) shows that for almost all products, prices are higher in larger cities. The coefficient on population is equal to 2.8% on average, with a standard deviation of 3.9%. It is positive for 44 out of 55 products and positive and significant at the 10% level at least for 24 of them. The coefficient on remoteness is equal to 14.5% on average (standard deviation of 22.4%.). It is positive for 47 out of 55 products, and positive and significant for 21 of them.

Regarding product availability, panel (b) of Figure 3.6 shows that for almost all products, the probability that it is missing is lower in larger cities. The average coefficient on population size is equal to -7.1% with a standard deviation of 4.9%. This coefficient is negative for 53 out of 55 product categories, and negative and significant for 46 of them. The coefficient on remoteness is equal to 11.3% on average with a standard deviation of 17.9%. It is positive for 43 out of 55 product categories and positive and significant for 16 of them.

All in all, these figures largely confirm the findings of Table 3.2: large cities have a greater access to products and services but are also more expensive, whereas remote cities suffer from both lower availability and higher prices. However, they also clearly demonstrate that a significant amount of heterogeneity exists in the relationship between product prices/product availability and the economic geography of cities.

Figure 3.6: ***Product price, product availability, remoteness, and city size***



(a) Product price

(b) Probability to be missing

Note: Both graphs present the density distributions of the coefficients obtained when running the regressions of columns (2) and (4) of Table 3.2 separately for the 55 product categories.

To try to dig deeper into this heterogenity, we focus on four specific broad categories:

foreign products, services, food products, and barcode products (see section 3.2.1 for the definitions). We identify the observations corresponding to these types of products thanks to adequate dummies that we interact with our variables of interest.

As Table 3.5 shows, we detect no significant heterogeneity for imported products. The prices of services increases more in large cities as compared to other products, whereas the probability that they are missing increases less with remoteness. This type of heterogeneity is consistent with services being less tradable than goods. The presence of food products is less intensely related to the city's population size; this finding is well in line with the fact that many food products are produced in rural areas where cities are smaller (see our discussion on auto-consumption). Note the negative effect of remoteness on product availability is stronger for food products, perhaps due to lower transportability of food products, especially fresh ones, in the context of a developing country such as Ethiopia. Finally, the price of barcode products and their availability are on average less related to cities' population size compared to the other products. This finding could be explained by the fact that barcode products are sold by larger firms with country-wide presence. However, remoteness does increase the price of these products.

Table 3.5: ***Heterogeneity***

	Log p_{pc}					$\mathbb{1}_{\text{missing}_{pc}}$				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Ln Remoteness	0.138^{a}	0.131^{a}	0.114^{a}	0.137^{a}	0.112^{a}	0.111^{b}	0.116^{b}	0.076	0.107^{b}	0.098
	(0.032)	(0.030)	(0.029)	(0.033)	(0.031)	(0.052)	(0.053)	(0.061)	(0.053)	(0.067)
Ln Population	0.029^{a}	0.025^{a}	0.024^{a}	0.030^{a}	0.022^{a}	-0.071^{a}	-0.069^{a}	-0.085^{a}	-0.073^{a}	-0.084^{a}
	(0.005)	(0.005)	(0.004)	(0.005)	(0.005)	(0.009)	(0.009)	(0.011)	(0.009)	(0.011)
Ln Remoteness × Imported	-0.052				-0.034	-0.069				-0.065
	(0.046)				(0.041)	(0.042)				(0.040)
Ln Population × Imported	-0.010				-0.005	-0.011				-0.004
	(0.007)				(0.006)	(0.007)				(0.006)
Ln Remoteness × Services		0.033			0.052		-0.102^{a}			-0.084^{c}
		(0.041)			(0.046)		(0.036)			(0.043)
Ln Population × Services		0.025^{a}			0.028^{a}		-0.029^{a}			-0.015^{c}
		(0.006)			(0.007)		(0.007)			(0.008)
Ln Remoteness × Food			0.049		0.054			0.079^{b}		0.059
			(0.033)		(0.035)			(0.038)		(0.042)
Ln Population × Food			0.010		0.012^{c}			0.035^{a}		0.033^{a}
			(0.006)		(0.006)			(0.006)		(0.006)
Ln Remoteness × Barcode				-0.043	-0.026				-0.020	-0.016
				(0.040)	(0.038)				(0.039)	(0.040)
Ln Population × Barcode				-0.033^{a}	-0.027^{a}				0.013^{b}	0.019^{a}
				(0.007)	(0.006)				(0.006)	(0.006)
Observations	28,183	28,183	28,183	28,183	28,183	42,506	42,506	42,506	42,506	42,506

Notes: The dependent variables are the log price (columns 1 to 5) net of its product category-level correlation with the local per capita income and a dummy equal to 1 if the product is unavailable in the city (columns 6 to 10). These dependent variables are defined in the *product* × *city* dimension. Standard errors clustered at the city-level in parentheses. ^{a}, ^{b}, and ^{c}, respectively, denote significance at the 1, 5, and 10% levels.

3.4 Cost of living across Ethiopian cities

We have shown that, on average, prices are higher in large cities and in remote cities, whereas products are more likely to be available in densely populated and/or less remote locations. However, using this micro-analysis to draw any conclusion on the relationship between city size, geography, and the cost of living is hard. Indeed, we have shown that behind these average relationships, a significant amount of heterogeneity exists across products; the overall penalty that consumers in larger and more remote locations pay will then depend on the distribution of their expenditure shares across products. Moreover, no simple way exists to account for both the price of available products and product availability when measuring the cost of living. To make some progress along these two dimensions, one needs to put some structure on the preferences of consumers to derive a measure of the local cost of living that is consistent with these preferences.

We provide the structure in this section by deriving two alternative measures of the local cost of living. We then revisit our empirical micro-evidence using these local price indexes as dependent variables.

3.4.1 Measurement

We propose two local price indexes.

Feenstra-CES price index.

For the first one, we follow Handbury and Weinstein (2015) and compute the spatial version of the price index proposed by Feenstra (1994). This index assumes consumers have CES preferences nested in a Cobb-Douglas utility function. More precisely, their utility function is as follows:

$$U = \prod_{g \in G} \left(\sum_{j \in J_g} q_j^{\frac{\sigma_g - 1}{\sigma_g}} \right)^{\alpha_g \frac{\sigma_g}{\sigma_g - 1}}, \tag{3.1}$$

where G denotes the set of product-categories consumed, J_g the set of varieties consumed in product-category g, q_j the quantity of each product j consumed, σ_g the elasticity of substitution between varieties in product-category g, and α_g the Cobb-Douglas preference parameter for product-category g.

With such preferences, and assuming consumers are utility-maximizers, we can show the local exact price index in location c (EPI_c) is a geometric average of category-level indexes, the weights being the (constant) share of each category in households' consumption α_g. Category-level indexes have two parts: (1) an intensive part (SPI_c) that is a standard price index that tracks the price gap across products in location c compared to a location of reference, (2) and extensive part (VA_c) that measures the utility cost of unavailable products in location c compared to the same reference location. Here, we assume the reference location is a fictitious city where all the products of all the product-categories are available at a price equal to the median price observed in 2015 across all Ethiopian cities.

Then, the formulas are as follows:

$$\begin{aligned} EPI_c &= \Pi_{g\in G}\left(SPI_{gc}\times VA_{gc}\right)^{\alpha_g} && (3.2)\\ SPI_{gc} &= \Pi_{j\in J_{gc}}\left(\frac{p_{jc}}{p_{jE}}\right)^{w_j} && (3.3)\\ VA_{gc} &= \left(\frac{\sum_{j\in J_{gc}} x_j}{\sum_{j\in J_g} x_j}\right)^{\frac{1}{1-\sigma_g}}, && (3.4)\end{aligned}$$

where p_{jc} is the individual price of good j in city c, and p_{jE} is the median price of good j across Ethiopian cities, J_{gc} is the set of products in category g that are available in city c (which size is also noted J_{gc}), w_j is the log-ideal Sato-Vartia weight (based on the share of product j in consumers's expenditures for product-category g both at the city- and at the national- levels), x_j is the total expenditures for product j of the nationally representative consumer, so that $\left(\frac{\sum_{j\in J_{gc}} x_j}{\sum_{j\in J_g} x_j}\right)$ represents the share of the products of product-category g that are available in c in the representative consumer's overall consumption of product category g.

Data limitations imply some components of the exact price index cannot be directly measured or estimated. We thus make the following assumptions to overcome this issue:

- We do not have the details of the weights for products j within product categories g. We thus assume even weights within each product-category $w_j = 1/J_g$ and $x_j = 1/J_g$. Product categories instead have different weights based on survey data, as explained in section 2.2.
- A key parameter is σ_g, the elasticity of substitution between varieties within a given product-category. We do not have the data to estimate these elasticities, and no estimates are available for Ethiopia. We thus use the elasticity estimated

by Broda et al. (2017) for the two countries the most similar to Ethiopia, which are present in their dataset (see section 3.2).

Travel Price Index.

In any version of a Feenstra (1994) price index, the consumption of unavailable varieties is assumed to be zero, or put differently, the implicit price of missing varieties tends to infinity. This feature makes sense in the context of Feenstra (1994), who aims to measure the evolution of the price index over time: a variety that is not available at time t cannot be consumed in t. In a spatial context, however, consumers may travel around to shop products that are not available in their hometown. In an alternative index, we thus assume that if the product is not available in their hometown, consumers choose to purchase the product from the cheapest location (including travel costs) where it is available. Formally, the price $\tilde{p}_{jc}$ of product j in city c is given by:

$$\tilde{p}_{jc} = \begin{cases} p_{jc} & if\ j \in J_{gc} \\ Min\left(p_{j1} + 2\text{travel\_cost}_{c1}, p_{j2} + 2\text{travel\_cost}_{c2}, ..., p_{jC} + 2\text{travel\_cost}_{cC}\right) & if\ j \notin J_{gc}, \end{cases}$$

where p_{jc} is the price of product j in city c, J_{gc} is the set of products available in city c, and travel_cost$_{cc'}$ is the travel cost between cities c and c'.

The travel price index (TPI) is then given by:

$$TPI_c = \Pi_{g \in G} \left(\Pi_{j \in J_g} \left(\frac{\tilde{p}_{jc}}{p_{jE}} \right)^{w_j} \right)^{\alpha_g} \tag{3.5}$$

where J_g is the entire set of products included in category g considered in the analysis. As for the CES-Feenstra index, due to data limitations, we assume even weights within each category $w_{jc} = 1/J_g$.

Facts on the price of unavailable products. Before turning to the determinants of the price index, we present in Table 3.6 an analysis of the determinants of the price of the cheapest accessible variety as defined above. Columns (1) and (2) only include the price of the cheapest variety when the product is missing locally, whereas as a robustness check, columns (3) and (4) include all of the observations. The cheapest price at which a missing variety is available in the other cities is higher for more remote locations, which is not surprising, because those locations face, by definition, higher travel costs to the rest of the country. This price is also higher in large cities. This finding implies that if a product is not available in a large city, consumers do not have access to cheap alternatives in the nearby cities. Instead, they have to pay a large trade costs to get access to the product in another (distant) city. When we include in the sample all of the observations, as expected, results are qualitatively similar to those of columns (1) and (2) of Table 3.6.

Table 3.6: ***Macro-determinants of the price of the cheapest accessible variety.***

	Log $\tilde{p}_{pc}$			
	(1)	(2)	(3)	(4)
Ln Remoteness	0.581^a	0.644^a	0.331^b	0.356^b
	(0.111)	(0.132)	(0.141)	(0.170)
Ln Population	0.027^b	0.031^b	0.047^a	0.059^a
	(0.011)	(0.015)	(0.015)	(0.019)
Sample	Missing		All	
Purged price	No	Yes	No	Yes
Observations	14,323	14,323	42,506	42,506
R-squared	0.93	0.91	0.93	0.89

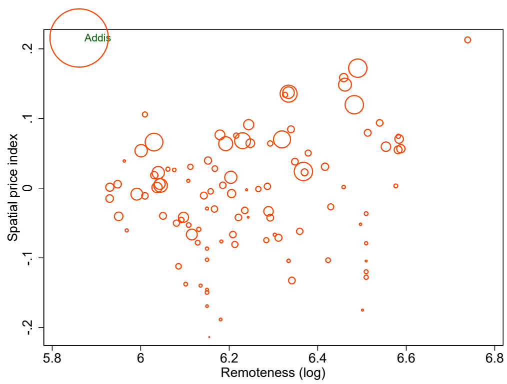
Notes: When a product is not available in a given city, the alternative price is the price of the cheapest accessible variety (plus the travel costs).Standard errors in parentheses are clustered in the $region \times market-type$ dimension ($market-type = \{urban, rural\}$). [a], [b] and [c] respectively denote significance at the 1, 5 and 10% levels.

3.4.2 Results

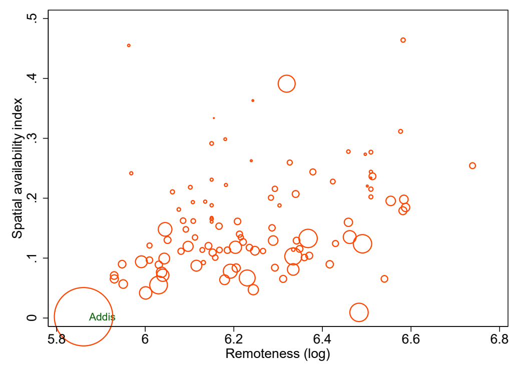
Baseline results. We first propose a visual inspection of the relationship between the cost of living as measured by our prices indexes (and their various components) and city size and remoteness. The graphs in Figure 3.7 plot the level of the various components of the price index against the degree of remoteness of the city; the size of the circle is proportional to the population of the city it represents. Panel (a) relates to the "intensive" component of the CES-price index, that is how expensive available varieties are compared to the median price quote observed in Ethiopia. Two main messages emerge: once expenditure shares are accounted for, large cities are still more expensive (big circles are at the top of the graph), and remote cities also still tend to exhibit higher prices (with the notable exception of Addis-Ababa). Regarding the "extensive" component of the price index in panel (b), when expenditure shares and elasticities of substitution are taken into account, large and less remote cities appear to be those in which consumers have access to a greater variety of products, which translates into a lower value of the extensive component of the price index. The combination of the two components gives us the exact price index. It appears on panel (c). The relationship with size is no longer obvious, due to the compensating effects of city size on the level of prices and on product availability. However, remote cities still do have a higher cost of living, due to higher prices and lower product variety. Finally, when using the travel price index in panel (d), the picture is quite similar to the one observed for the Exact Price Index. No clear correlation exists between city size and cost of living, and the correlation with remoteness is positive (but slightly more pronounced).

The graphical observations are largely confirmed by the econometric analysis reported in Table 3.7. The EPI is the product of the intensive and the extensive components. In large cities, the intensive component is higher (the weighted price index is higher), whereas the availability index is lower, because more products are available. The two effects cancel out so that city size does not significantly influence the cost of living as measured by the EPI. On the other hand, remote locations exhibit both higher intensive and higher extensive components of the EPI, due to the detrimental effect of remoteness on both the price of available products and on product availability. Results with

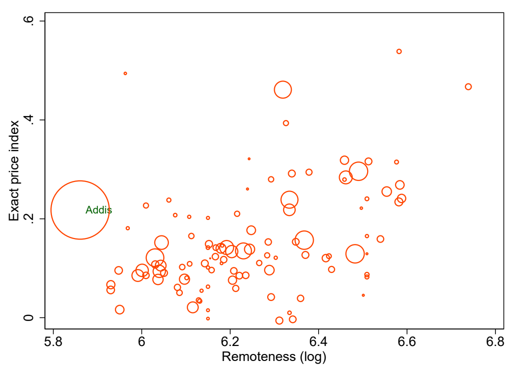
Figure 3.7: ***The geography of prices and product availability***



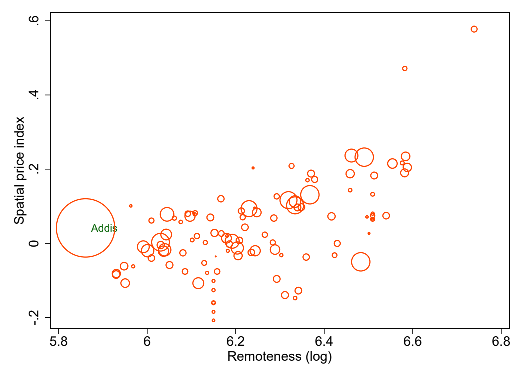
(a) ***SPI***



(b) ***VA***



(c) ***EPI***



(d) ***TPI***

Notes: Each dot is an Ethiopian city. The size of the circle is proportional to the city's population size. The log of the intensive part of the price index is in panel (a). The log of the extensive component is in panel (b). The exact price index in panel (c) is the sum of the indexes presented in panels (a) and (b). The log of the travel price index is in panel (d).

the TPI are very similar. These results hold in specifications including additional control variables and alternative measures of remoteness as shown in Table 3.8.[20] More specifically, controlling for the ethnic diversity of the city does not change the correlation between the price index and size and remoteness of the city. Measuring remoteness with a weighted average of bilateral distances between cities does not change the results either. In columns (5) and (6), we find little evidence that the time to an import corridor or the travel time to Addis-Ababa explains spatial differences in terms of the cost of living. The empirical regularities are not driven by Addis-Ababa as shown in columns (7) and (8). Using different weighting schemes (in terms of households' expenditure shares) for urban and rural cities does not affect the results either (columns 9 and 10).

Table 3.7: ***City-level regressions-local consumer price index***

	Ln SPI	Ln VA	Ln EPI	Ln TPI
	(1)	(2)	(3)	(4)
Ln Population	0.046^a	-0.042^a	0.004	0.015
	(0.007)	(0.007)	(0.011)	(0.012)
Ln Remoteness	0.108^b	0.127^a	0.235^a	0.429^a
	(0.045)	(0.029)	(0.054)	(0.084)
Observations	105	105	105	105
R-squared	0.46	0.46	0.17	0.39

The dependent variable is the log spatial price index (Ln *SPI*) in columns 1, the log spatial availability index (Ln *VA*) in columns 2, the log exact price index (Ln *EPI*=Ln *SPI* + Ln *VA*) in column 3, and the alternative spatial price index in column 4. Robust standard errors are in parentheses. ^{a}, ^{b}, and ^{c} respectively, denote significance at the 1%, 5%, and 10levels.

20. The results are also robust to purging individual prices from their correlation with income per capita as shown in Appendix (Table 3.14).

Table 3.8: ***City-level regressions-local consumer price index : robustness checks***

	EPI	TPI	EPI	TPI	EPI	TPI	EPI	TPI	EPI_UR	TPI_UR
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Ln Population	0.004	0.014	0.004	0.016	0.008	0.022	-0.000	0.013	-0.006	0.017
	(0.012)	(0.012)	(0.012)	(0.017)	(0.012)	(0.018)	(0.012)	(0.013)	(0.011)	(0.012)
Ln Remoteness	0.247^a	0.456^a					0.250^a	0.438^a	0.178^a	0.384^a
	(0.048)	(0.075)					(0.055)	(0.086)	(0.044)	(0.080)
Ln of Ethnic diversity index	0.033	0.074^b								
	(0.032)	(0.030)								
Ln Remoteness (weighted)			0.102^c	0.246^a						
			(0.052)	(0.072)						
Ln travel time to import corridor					-0.025	-0.016				
					(0.025)	(0.029)				
Ln travel time to Addis-Ababa					0.024	0.059^c				
					(0.022)	(0.030)				
Observations	105	105	105	105	105	105	104	104	105	105
R-squared	0.18	0.45	0.05	0.21	0.06	0.13	0.18	0.39	0.11	0.35
Addis-Ababa	Yes	Yes	Yes	Yes	Yes	Yes	No	No	Yes	Yes
Weight				common					urban/rural	

Notes: The dependent variable is the exact price index (EPI) in odd columns, the travel price index (TPI) in the even columns, and the price indexes for urban areas (EPI_UR and TPI_UR) are reported in columns 11 and 12. Remoteness is measured as a population-weighted distance of the city to all Ethiopian cities. Standard errors in parentheses are clustered in the $region \times market-type$ dimension ($market-type = \{urban, rural\}$). ^{a}, ^{b} and ^{c} respectively denote significance at the 1%, 5%, and 10% levels.

Rich and poor. We finally compare the impact of remoteness and city size along the income distribution. More specifically, we compute the cost-of-living across Ethiopian cities for households in different quintiles of the income distribution.[21] Doing so requires information on expenditure shares for representative households at different points of the income distribution. We build a series of indexes for rich and poor households and then compute the gap in the cost of living between rich and poor across locations. We present the results in Table 3.9. The main finding is that the cost of living of poor households deteriorates more strongly with remoteness than the cost of living of richer households. This seems to be driven by the fact that poor households spend a higher fraction of their income on products that are both more expensive when available and more often not directly available in their city.

21. We define poor and rich households as those in the first and fifth quintiles of the income distribution, respectively.

Table 3.9: ***Difference in the spatial index of rich and poor***

	gap_SPI		gap_VA	gap_EPI		gap_TPI	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Ln average distance to other cities	0.024[a]	0.025[a]	0.018[c]	0.042[a]	0.043[a]	0.074[a]	0.112[a]
	(0.006)	(0.006)	(0.010)	(0.010)	(0.010)	(0.013)	(0.026)
Ln Population	0.003[c]	0.001	-0.003[b]	-0.000	-0.002	-0.000	-0.003
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.003)	(0.005)
Prices	Raw	Purged	-	Raw	Purged	Raw	Purged
Observations	105	105	105	105	105	105	105
R-squared	0.15	0.12	0.19	0.19	0.20	0.35	0.31

Notes: The dependent variable is the difference in the spatial price index between poor and rich households. (*gap_SPI*) in column 1, and 2, the difference in the spatial availability index between rich and poor households (*gap_VA*) in column 3, the difference in the exact price index between rich and poor households (*gap_EPI*) in column 4, and 5, the difference in the travel spatial price index of poor and rich households (*gap_TPI*) in column 6, and 7. It is available across regions for urban and rural markets. All these variables are in Log. Robust standard errors are in parentheses. [a], [b] and [c] respectively denote significance at the 1, 5 and 10% levels.

3.4.3 Nominal income versus real income

We now ask whether cost-of-living spatial variations across Ethiopian regions have implications for the measurement of income inequality. We have access to a measure of nominal income per capita for 18 regions. These regions are the 10 administrative regions of Ethiopia, and for eight of them, we have information on income per capita in urban and rural areas (Addis-Ababa and Dire Dawa being only urban).

We compare the ranking of regions in terms of nominal income per capita with their ranking in terms of real income per capita. We use three alternative deflators to compute real income, namely, the standard price index (*SPI*) that does not account for product availability, and the two price indexes (EPI and TPI) that do account for product availability.

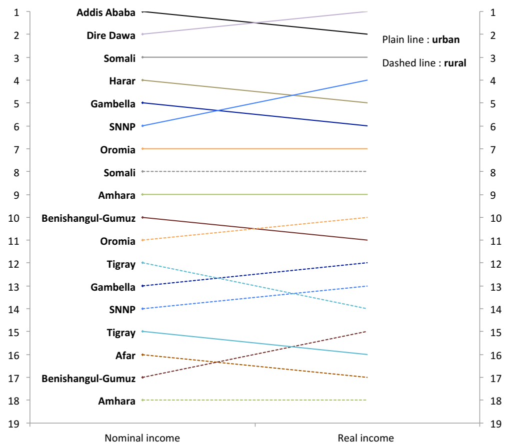
We report the exercise in Figure 3.8, panels (a) to (c). A visual inspection shows that changes in the ranking of regions if we use real income instead of nominal income. The Kendall rank correlation between nominal income and real income ranges from .74 and .90 depending on the index used as a deflator. Accounting for differences in the cost of living is thus essential to get a complete picture of spatial income inequality.

In particular, deflating with a standard price index that does not account for product availability has less of an impact on the ranking than the other two indexes (Kendall rank correlation of .9 in panel (a) vs .82 and .74 in panels (b) and (c)). Interestingly, we see that the ranking of the two wealthiest regions (Addis-Ababa and Dire-Dawa) changes when deflated by the standard price index, whereas it is stable when the deflator is an index accounting for product availability. Therefore, the price level of available products is higher in Addis-Ababa than in Dire Dawa, which more than compensates for the income differential between the two places. However, once the impact of unavailable products is properly measured, the cost of living in Addis-Ababa decreases relative to Dire Dawa, and real income in Addis-Ababa remains higher than in Dire Dawa.

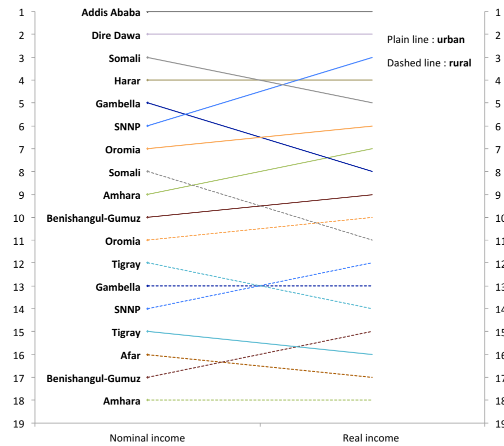
We also see the poorest regions in terms of nominal income are rural regions. These regions are not compensated for the lower nominal income by lower cost of living, although the decision to deflate or not the nominal income influences the ranking among these regions, which highlights the differences in product availability across them. [22]

22. We complement this subsection by analyzing the regional disparity of nominal and real income. See Figures 3.9 , 3.10, 3.11 and 3.12 in the appendix.

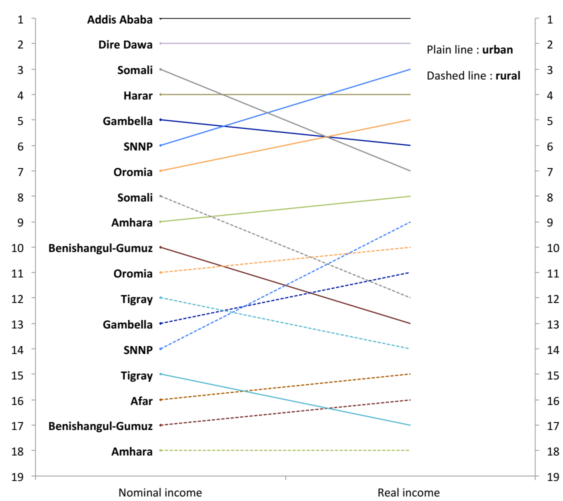
Figure 3.8: ***Spatial ranking of income: nominal versus real***



(a) ***SPI***



(b) ***EPI***



(c) ***TPI***

Notes: Each line in the graphs links the ranking of a region in terms of nominal income to its ranking in terms of real income. In panel (a) the spatial price index is used to deflate nominal income, the exact price index is used in panel (b), and the travel price index is used in panel (c).

3.5 Internal migration and cost of living

In this section, we show that properly measuring the local cost of living greatly improves our understanding of internal migration patterns in Ethiopia. A key insight from the economic geography literature is that spatial differences in terms of real income determine migration patterns across regions (Krugman, 1991a). Real income in such models featuring love for variety is nominal income deflated by a local ideal price index, which accounts for product availability. Due to the lack of data, the literature has not yet investigated the impact of prices and product availability on internal migration patterns. We here try to fill this gap. We first motivate our empirical model. We then show product availability contributes substantially to the internal movements of people in Ethiopia. We eventually discuss how cost of living intersects with economic geography in determining migration flows.

3.5.1 Migration equations

We use migration data from the 2013 national labor force survey. Unlike price data, migration flows are available at the region level, not the city level. We thus run our analysis at the region level, and all of the socio-economic explanatory variables included in the analysis are weighted averages of city-level characteristics, using as weights the share of each city in the population of the region where it is located. [23]

In Ethiopia, employment opportunities are the most often cited motivation for migrating. Land scarcity at the origin town and cost of living are also part of the underlying reasons for migrating (Bundervoet, 2018). In addition to these determinants, we control for other standard determinants such as bilateral distance, or income and population size in the destination region (Crozet, 2004). In the end, the empirical model we estimate to uncover the determinants of internal of migration flows is as follows:

23. For instance, $\text{distance}_{od} = \sum_{i \in o} \sum_{j \in d} \omega_i \times \omega_j \times \text{distance}_{ij}$ where indexes i and j stand for cities in regions o and d respectively, ω_i is the share of city i in the population of region o, ω_j is the share of city j in the population of region d. distance_{ij} is the distance between city i and city j.

$$\begin{aligned}\text{mig}_{od} =&\beta_1 \text{Ln distance}_{od} + \beta_2 \text{Ln emp}_d + \beta_3 \text{Ln pop}_d + \beta_4 \text{Ln area}_d \\ &+ \beta_5 \text{Ln income}_d + \beta_6 \text{ Ln price index}_d + \gamma_o + \varepsilon_{od},\end{aligned} \tag{3.1}$$

where mig_{od} stands for the share of migrants from region o (origin) who have decided to move to region d (destination), distance_{od} is the distance between region o and region d and is a standard proxy for migration costs between o and d, emp_d is the employment rate in region d and proxies for employment opportunities at destination, and pop_d and area_d are respectively the total population and the surface area of d and proxy for the size of the destination region. Our main variables of interest are income_d and price index_d and stand for the nominal income and cost of living in d.

Our specification also contains origin fixed effects γ_o, allowing us to abstract from any consideration as to why some regions have migration flows that are more diversified across regions than others. The coefficient associated with each explanatory variable is estimated comparing migration flows from a given region o with each of the possible destinations d. Note that our specification is similar to one where the dependent variable would be the log number of migrants from o to d, instead of the share of migrants from o going to d.

3.5.2 Results

Table 3.10 presents our main results. Column (1) corresponds to the estimation of equation 3.1 excluding the price index from the regressors. Most of the variables have the expected sign. People are less likely to migrate to distant regions, and are more likely to move to larger places in terms of population and surface area. Our proxy for employment opportunities is not significant, which might reflect the difficulty in measuring job opportunities with (the opposite of) the unemployment rate in a country where informality is still pervasive. All else equal, regions with a higher nominal income per capita appear to be more attractive to migrants.

We then add in columns (2) to (4) different measures of price indexes that capture the spatial differences in terms of cost of living. More specifically, when we control for the standard price index (SPI) , which only accounts for the price of available varieties, more expensive regions appear to attract fewer migrants. However, controlling for SPI barely affects the other coefficients, and does little to improve the predictive power of our empirical model. The conclusions are different once we control for the exact price index (EPI), and thus for product availability on top of the price of available varieties. As the results in column (3) of Table 3.1 show, the coefficient on EPI is strongly negative and significant, and the inclusion of this price index increases the R-squared of the regression by 8 p.p. (i.e. 20%). Measuring local cost of living thanks to TPI instead of EPI yields very similar results. All else equal, the price of consumed goods and services and their availability thus appear to matter both statistically and economically to migrants when they make their location decisions.

To assess the interplay between the cost of living, economic geography, and migration decisions, we reproduce in columns (5) to (8) the first four columns of Table 3.10, but we now also control for the geographic remoteness of the destination region. As column (5) shows, remote regions attract fewer migrants. Once we control for the price of available varieties in column (6), remoteness remains significantly (and negatively) related to migration, but the coefficient is slightly smaller in absolute value, and the coefficient on population is slightly boosted; these movements in the coefficients are coherent with larger and remote cities/regions being more expensive places to live, so that not controlling for prices tends to downward bias the coefficients on population size and geographic remoteness.

The most interesting results in our view are in column (7). In this specification, we control for both product availability and the price of available varieties thanks to the exact price index (EPI). EPI is, as in column (3), highly negative and significant. Migrants prefer going to places where the price index is lower, that is, with more numerous and less expensive products and services. Moreover, controlling for it causes a sharp increase in the (negative) coefficient associated with remoteness, which becomes statistically indistinguishable from 0, and a decrease in the (positive) coefficient on population, which remains significantly positive. These patterns align with our previous results showing that small and remote places have a narrower set of available products and services. Actually, the negative correlation between migrations flows and remoteness of the destination region vanishes once the exact price index is included, suggesting the correlation is entirely driven by the effect of remoteness on the local cost of living. Using TPI instead of EPI yields qualitatively similar results.

Having an adequate measure of local cost of living that accounts for both prices and product availability is thus crucial to understand internal migration patterns in a developing country such as Ethiopia.

Table 3.10: ***Migration, income per capita and prices***

	Migration share							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Ln distance	-0.091^a	-0.094^a	-0.100^a	-0.087^a	-0.079^a	-0.083^a	-0.094^a	-0.090^a
	(0.024)	(0.025)	(0.021)	(0.021)	(0.021)	(0.022)	(0.021)	(0.023)
Ln destination employment rate	0.069	-0.115	-0.265^b	-0.019	0.218^c	0.062	-0.165	-0.075
	(0.108)	(0.113)	(0.109)	(0.097)	(0.124)	(0.113)	(0.131)	(0.130)
Ln destination pop	0.046^a	0.049^a	0.032^a	0.027^a	0.035^a	0.038^a	0.029^a	0.027^a
	(0.010)	(0.010)	(0.008)	(0.008)	(0.008)	(0.009)	(0.007)	(0.008)
Ln destination area	0.031^a	0.030^a	0.034^a	0.036^a	0.030^a	0.029^a	0.034^a	0.037^a
	(0.009)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Ln income per capita at destination	0.146^b	0.138^b	0.123^b	0.133^b	0.143^b	0.137^b	0.125^b	0.132^b
	(0.062)	(0.061)	(0.057)	(0.055)	(0.058)	(0.058)	(0.057)	(0.055)
Ln destination remoteness					-0.204^a	-0.184^a	-0.081	0.058
					(0.056)	(0.052)	(0.050)	(0.074)
Ln SPI at destination		-0.411^b				-0.317^c		
		(0.182)				(0.163)		
Ln EPI at destination			-0.649^a				-0.571^a	
			(0.125)				(0.117)	
Ln TPI at destination				-0.499^a				-0.576^a
				(0.116)				(0.166)
Observations	110	110	110	110	110	110	110	110
R-squared	0.47	0.49	0.55	0.54	0.51	0.51	0.55	0.54
Adjusted R-squared	0.39	0.40	0.47	0.46	0.42	0.42	0.47	0.46
Origin fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The dependent variable is the share of migrants from a given region who have decided to move to another region. ^{a}, ^{b}, and ^{c} respectively denote significance at the 1%, 5% and 10% levels. SPI is the spatial price index, EPI the exact price index, and TPI the travel price index.

3.6 Conclusion

This paper documents geographical discrepancies in the price and availability of products purchased by consumers across Ethiopian cities. We find the cost of living is higher in remote cities because consumers pay higher prices and have access to a narrower set of products and services. All else equal, consumers in large cities pay higher prices but have access to a broader set of products and services. The price effect and the availability effect cancel each other, so that cost of living that accounts for both prices and availability is not related to city size.

One implication of our results is that to have a proper view of income inequality, nominal income should be deflated by local price index, and the choice on the price index matters for the ranking of regions we obtain.

Last, we show that differences in cost of living across Ethiopian regions is also instrumental to understand the destination choice of internal migrants in Ethiopia. Interestingly, differences in cost of living driven by better availability of products and services explain a significant fraction of internal migration patterns. Expanding the availability of products and services might thus be an effective tool for cities to attract migrants or retain their workforce.

3.7 Appendix to Chapter 3

3.7.1 Additional tables on the data

Table 3.11: ***List of Product Categories***

Beverages - Alcoholic	Fuel And Power	Vegetables (Fresh)
Beverages - Non Alcoholic	Glass Ware	
Bread And Other Prepared Foods	Household Operation	
Bed Sheet (Non-Patterned Bahir Dar)	Jewellery	
Bed Sheet (Patterned Kombolcha)	Livestock	
Cereals Milled	Meat	
Cereals Unmilled	Medical Care	
Chairs, Tables, Etc.	Metal Ware	
Cigarettes	Miscellaneous Goods And Services	
Clothing	Oil Seeds	
Coffee, Tea, Chat And Buck-Thorn Leaves	Oils And Fats	
Communication	Other Food Items	
Construction Materials	Other Household Equipment	
Cost Of Milling	Other Medical Expenses	
Cost Of Tailoring	Personal Care	
Diary Products And Egg	Personal Effects	
Domestic Service	Plastic Ware (Local Made)	
Draught Animals	Potatoes, Other Tubers And Stems	
Earthen Ware	Prepared Cereal Products	
Education	Pulses Milled Or Split	
Equipment And Accessories	Pulses Unmilled	
Farm Equipment (Hand Made)	Ready-Made / For Adults	
Farm Equipment (Industrial Product)	Ready-Made / For Children	
Fish And Fish Products	Spices	
Food Taken Away From Home	Straw, Bamboo And Others	
Footwear (Men And Women)	Tobacco	
Fruits Fresh	Transport	

Table 3.12: ***Product List***

'Katikalla'-Lt	Durrah-Kg	Internet Service-10hrs/month
'Tej' (Mead)-Lt	Hulled Barley-Kg	Mobile Apparatus(Nokia6200)
'Tella'-Lt	Maize (White)-Kg	Mobile call from Tel-Period
Araki (Local)-900cc	Oats -Kg	Telephone Charge (with town)-Period
Beer (Bedele)-330cc	Rice (Imported) -Kg	Telephone Charge to Addia Ababa -Period
Beer (Harar)-330cc	Sorghum Red-Kg	Telephone Line Installation Charge-Once
Beer (Meta Abo)-330cc	Sorghum White-Kg	Bricks (25cm x 12cm x 6cm)
Brandy (Local)-900cc	Sorghum Yellow-Kg	Cement/Bag/(Local)-50Kg
Cognac (Local)-900cc	Wheat Black (Red)-Kg	Chipwood (125cm x 250cm x 8mm)
Gin (Local)-900cc	Wheat Mixed-Kg	Coarse Aggregate Gravel-Meter cube
Saris Wine (Normal)-750cc	Wheat White-Kg	Corrugated Iron Sheet (.2mm)
Ambo Mineral Water-500cc	Book Shelves Wanza (3 Shelves) no Door No	Door made of iron
Coca Cola/Fanta-300cc	Chairs Wanza (Hand Made) Varnished	Floor Board 4m Length
Mineral Water-Lt	Chairs Wanza (Machine Made) Varnished. No	Gutter No 33 -Meter
Pepsi Cola/Mirinda-300cc	Chest of Drawer	Hollow Concrete Block(15x20x40 cm Cube)No
'Dabo' (Traditional Ambasha)-350gm	Cupboard Wanza (2 doors) Varnished	Iron Pipe 6mt. (1/2 inch Wide) Local
'Dabo' (Traditional Sheleto)-350gm	Double Bed Wanza (120cm) Varnished	Key (With hand)
'Enjera' ('Teff' Mixed)-325gm	Sofas (Complete)	Lime-Kg
Biscuits -150gm	Table Wanza (Hand made) Not Varnished. No	Lime/Jeso(Local) -Kg
Bread Wheat (Bakery) -350gm	Table Wanza (Machine made) Varnished	Mega Paints-4Kg
Bed sheet (Non-patterned Bahir Dar)-1.90m x 2.50m	Gissila-Packet	Nail (7cm - 12cm)-Kg
Bed sheet (Patterned Kombolcha)-1.90m x 2.50m	Marlboro-Packet	Nail With Cape-Kg
Bed Cover(Patterned Kombolcha)	Nyala -Packet	Nefas seleke Paints-4Kg
Blanket Woolen(Debre B.)-160cm x 220cm. No	Rothmans/England/ -Packet	Sand -Meter cube
Curtains-meter	Abujedid(Akaki/Bahir Dar)91cm. Meter	Stone for House Construction-Meter cube
Mattress-Sponge (A.A Foam) 120cm	Abujedid(Komb./Arba Min.)150cm Meter	Wall Paints-Super (Fluid) Normal -4Kg
Towel-Local (Kombolcha)	Cotton-Kg	Wall Paints-Super (Fluid) Plastic-4Kg
'Furno Duket' Locally Processed-Kg	Deriya-Meter	Water-Meter cube
'Teff' Black (Red)-Kg	Hisufi-Meter	Water Tanker,Roto (1 meter cube)
'Teff' Mixed-Kg	Jersi-Meter	Window Glass (50cm x 50cm x 3mm)
'Teff' White -Kg	Kashemire-Meter	Wood for House Construction('Atana')
Barley Mixed-Kg	Kefai-Meter	Yewellel Nitaf(Cement Made Tile)
Barley White-Kg	Khaki(Akaki) -Meter	Yewellel Nitaf(Plastic Made Tile)
Durrah-Kg	Khaki(S-10,000 Twil)150cm-Meter	Cereals-100Kg
Maize (White)-Kg	Nylon(Mojo)-Meter	Pepper Whole-100Kg
Oats-Kg	Polyster(Arba Minch/Awasa)-Meter	Pulses-100Kg
Sorghum-Kg	Poplin(Dire Dawa)105cm-Meter	Khaki/Teteron Suit (Boys)
Wheat Mixed-Kg	Poplin(Komb./Arba Minch)150cm-Meter. Meter	Khaki/Teteron Suit (Men)
Wheat White-Kg	Tetron(A.Minch)-Meter	Woolen Suit (Men)
'Teff' Black (Red)-Kg	Wool-England 100%-Meter	Camel Milk-Lt
'Teff' Mixed-Kg	'Chat'-Kg	Cheese Cottage-Kg
'Teff' White-Kg	Buck Thorn Leaves-Kg	Cow Milk (Unpasteurized)-Lt
African Millet-Kg	Coffee Beans-Kg	Cow Milk (pasteurized)-Lt
Barley Black-Kg	Coffee Leaves-Kg	Egg (Traditional)-Dozen
Barley Mixed-Kg	Coffee Whol-Kg	Goat Milk-Lt
Barley White-Kg	Malt-Barley-Kg	Powdered Milk (Me&My)-450gm
Barley for Beer-Kg	Malt-Wheat-Kg	Yoghurt (Traditional)-Lt

Maid Servant-Month	Cup of Milk	Neck Laces (6gm 18 carat Local)
Salary for Guard-Month	Cup of Tea	Rings (4gm 18 carat Local)
Unskilled Service (Daily Laborer)-Day	Anbessa Leather Shoes Men(Local)Pair	Seiko-21 Jewels Automatic (Men)
Donkey	Boots for Men Plastic (Local)-Pair	Seiko-21 Jewels Automatic (Women)
Horse	Canvas Shoes (China) Men-Pair	Bull (2-4 Years)
Mule	Canvas Shoes (Local) Men-Pair	Cock (Indigenous)
'Jebena' Medium Size	Cost of Mending Shoes sole(Men)-Pair	Cow (4 Years and Above)
'Mitad' (Griddle of Clay)	Leather Shoes Men (Croft)-Pair	Goat (10-15Kg)
Cup for Coffee (China)	Plastic Shoes (Local) Women-Pair	Heifer (2-4 Years)
Local Stove '(Lakech)'	Plastic shoe(Childern)-Pair	Hen (Indigenous)
Plastic tile-Meter	Sendel Plastic Shoes (Imp.)-Pair	Ox (4 Years and Above)
Plate Clay (imported)	Shoe leather,Children(Imported)-Pair	Sheep (10-15Kg)
Water Pot	Shoe leather,Children(Local)-Pair	Beef-Kg
Ball Point-Bic England	Shoe leather,Women(China)-Pair	Camel Meat-Kg
Day School Fee-Private(Grade 9-10)-Month	Shoe leather,Women(Local)-Pair	Amoxacillin(500mg)-16 caps
Day School Fee-Public(Grade 7 & 8)-Month	Shoe sendel(Plastic) Childern-Pair	Ampicillin(250 mg) Local-56 caps
Exercise Book (50 Leaves) Local	Slippers Sponge Adult (China)-Pair	Asprin (300 mg) Local-20 pills
Night College Fee-Private-Credit/hr	Walking Shoes(Imp.) Non-Leather-Pair	Bactrim(480mg) Local -30 pills
Night School Fee-Government(Gr.9-10)-Month	Avocado-Kg	Chloramphenicol(250mg) Local-56 caps
Night School Fee-Private(Grade9-10)-Month	Banana-Kg	Cough Syrup (Efadykse) Local-125cc
Pencil (China)	Cactus-Kg	Fasider Table-1 pill
School Uniform Fee	Grapes-Kg	Insulin(Lente)-Buttle
Cassette Recorded Original(Local Music)No	Lemon -Kg	Magnesium Oxide -30 Pills
Expense for Photograph(Passport size)4Pho	Mango-Kg	Mezel(250mg) Local-30 caps
Newspaper (Addis Zemen)	Orange-Kg	Paracitamole(500mg) Local -20 Pills
Radio Set Philips 3 Band	Papaya-Kg	Penicillin injection(4 Mu. Local)-Buttle
T.V. Set Philips 21 inch(Colored)	Tangerine-Kg	Tetracycline (250 mg) Local-56 caps
Tape Recorder National (2 Speaker)	Buthane Gas (Shell)-12,5Kg	Vermox(100mg) Local-12 pills
Theater Enterance Fee-Once	Candles	Cooking Pan Medium (Local)
VCD Cassette rent	Charcoal-Kg	Electric 'Mitad' Aluminium
VCD-Player(Mayato Japan)	Diesel-Lt	Permuze(Japan)
'Digr'	Dung Cake-Kg	Refregrator
'Erfe '	Electric-Kwatt	Tray (Nickel) Medium N45 Local
'Kember'	Eveready Drycell	Charge for Money Transfer-Once
'Mofer'	Fire Wood-Meter cube	Coffin (Medium Quality)
Plough	Kerosine-Lt	Photocopy-Per page
Sickel	Matches-Box	Wedding Invitation Card-Per page
'Gejera'	Glass for Tea (Duralex)	Castor Beans-Kg
Pick Axe ('Doma')	Detergent (Omo)-50gm	Ground Nut Shelled-Kg
Sickel	Detergent(Zahira)-50gm	Linseed Red-Kg
Fish Fresh-Kg	Dry Cleaning (Suit Men)	Linseed White-Kg
Sardines (Imported)-125gm	Hard Soap (Imported)-200gm	Niger Seed-Kg
'Fasting Meal Without fish-One Meal	Hard Soap (Local)-200gm	Rape Seed-Kg
'Key Wot Yebeg/Yefyel'-One meal	Incense-Kg	Sesame Seed Red-Kg
'Key Wot Yebere'One meal	Sandal Wood	Sunflower-Kg
'Yebeg Kikil'-One meal	Toilet Paper (Mamko)-Roll	Butter Unrefined-Kg
'Yebeg Tibs'-One meal	Bracelet 20gm (18 carat Local)	Cooking Oil (Imported)-Lt
Cup of Coffee	Earrings (4gm 18 carat Local) -Pair	Cooking Oil (Local)-Lt

Vegetable Butter (Sheno & Shady)-Kg	Lentils Split-Kg	Jeans trouser and Jacket
Canned Tomato (Local)-410gm	Mixed Pulses Milled-Kg	Kemise(for childeren)
Dry Yeast(Baking powder)-350gm	Peas Milled-Kg	Shirt Long Sleeved(Imported)Boys
Honey-Kg	Peas Split-Kg	Socks (Imported) Cotton-Pair
Salt-Kg	Peas Split(Roasted)-Kg	Sweater (England) for Girls
Sugar-Kg	Vetch Milled-Kg	Sweater (Local) for Boys
’Kuraz’ Small Local Kerosine Lamp	Vetch Split(Roasted)-Kg	T-Shirt
Electric Bulb Philips(40/60 Watt)	Chick Peas-Kg	Basil Dry-Kg
Flash Light	Fenugreek(Green-Kg)	Black Cumin(Local)-Kg
Kerosine Lamp	Haricot Beans-Kg	Black Pepper(Local)-Kg
Bed Charge (Private-Per day	Horse Beans-Kg	Cardamon(Local)-Kg
Bed Charge (gov.)-Per day	Lentils-Kg	Chillies Whole-Kg
Doctor’s Fee (Government)-Per visit	Lima Beans-Kg	Cinnamon(Imported)-Kg
Doctor’s Fee (Private)-Per visit	Peas Green(dry)-Kg	Cloves(Imported)-Kg
Injection (Service Charge)-Once	Peas Mixed-Kg	Ginger Dry(Local)-Kg
X-Ray(For TB)-Once	Peas White-Kg	Ginger Wet(Local)-Kg
Barbery (Mens Hair Cut)	Soya Beans-Kg	Long Pepper(Local)-Kg
Blade-INDIA	Vetch-Kg	Pepper Whole-Kg
GIV Toilet Soap -90gm	’Gabi’	Tumeric Flour(Local)-Kg
Hair Dressing (Modern)	’Kemisna Netela’	White Cumin(Bishop’s Weed) Local -Kg
Modes(Disposable napkins-Packet)	’Netela’	’Sefed’
Parafin Hair Oil -330cc	Geldem	Sack 100Kg Capacity
Perfume-100cc	Jeans Trouser	’Gaya’-Kg
Shaving Machine (medium)	Jogging Suit(sport tuta)	Air Plane (To Addis-Ababa) -Trip
Shoe Polish(Black/Brown)-Once	Khaki Jacket	Animal Transport fare-Trip
Zenith Hair Oil(Liquids Form)-330cc	Khaki Short	Benzene-Lt
Zenith Hair Oil(Non-Liquids Form)-330cc	Leather Jacket	Bus Fare (per km)
Belt (Local) Hand Made	Mekremia	Bus Fare (within Town)-Tarif
Belt (Local) Machine Made	Nylon Dress	Car Washing and Greasing -Trip
Hand Bag (Imported Synthetic)	Pants(for men)	Cart Fare-Trip
Umbrella-Men Medium (Local)	Polyester Suit	Motor Oil (Mobil)-Lt
Umbrella-Women Medium (Imported)	Polyester skirt	Taxi Fare-Trip
Bucket (20 Litres)	Shash (Imported)	Beet Root-Kg
Jerrycan (20 Litres)	Shirts Long Sleeved (Imported)	Cabbage-Kg
’Bula’ -Kg	Shirts Long Sleeved (Local)	Carrot
’Kocho’ (Unprocessed)-Kg	Shirts Short Sleeved (Imported)	Cauliflower-Kg
Potato-Kg	Singlets (Local) White	Ethiopian Kale-Kg
Sweet Potato-Kg	Socks (Cotton) Imported-Pair	Garlics-Kg
’Dube’ Flour-Kg	Sweater (Local) Men	Green Peas-Kg
’Fafa’ Flour-Kg	Sweater (Local) Women	Leaks-Kg
Macaroni (Local) Without Egg-Kg	Sweater-Men (Imported)	Lettuce-Kg
Pastini -Kg	Sweater-Women (Imported)	Onions-Kg
Spaghetti (Local) Without Egg-Kg	T-Shirts	Pepper Green-Kg
Chick Peas Milled-Kg	Tetron Trouser	Pumpkin-Kg
Chick Peas Split(Roasted)-Kg	Under Wear China	Spinach-Kg
Fenugreek Milled-Kg	Woolen Suit	Tomatoesv
Horse Beans Milled-Kg	Baby Cloths(Complete)	Tea Leaves(Local)-100gm
Horse Beans Split(Roasted)-Kg	Jeans Trouser	

Table 3.13: ***Barcode products***

Category	Product
BEVERAGES - ALCOHOLIC	Beer (Bedele)
BEVERAGES - ALCOHOLIC	Beer (Meta Abo)
BEVERAGES - ALCOHOLIC	Beer (Harar)
BEVERAGES - ALCOHOLIC	Saris Wine (Normal)
CIGARETTES	Gissila
CIGARETTES	Nyala
CIGARETTES	Rothmans/England/
CIGARETTES	Marlboro
COMMUNICATION	Mobile Apparatus(Nokia6200)
EDUCATION	Ball Point-Bic England
BEVERAGES - NON ALCOHOLIC	Ambo Mineral Water
BEVERAGES - NON ALCOHOLIC	Coca Cola/Fanta
BEVERAGES - NON ALCOHOLIC	Pepsi Cola/Mirinda
DIARY PRODUCTS AND EGG	Powdered Milk (Me&My)
PREPARED CEREAL PRODUCTS	’Fafa’ Flour
HOUSEHOLD OPERATION	Detergent (Omo)
HOUSEHOLD OPERATION	Detergent(Zahira)
HOUSEHOLD OPERATION	Toilet Paper (Mamko)
OTHER HOUSEHOLD EQUIPMENT	Electric Bulb Philips(40/60 Watt)
CONSTRUCTION MATERIALS	Water Tanker,Roto (1 meter cube)
FUEL AND POWER	Eveready Drycell
FUEL AND POWER	Buthane Gas (Shell)
EQUIPMENT AND ACCESSORIES	Newspaper (Addis Zemen)
EQUIPMENT AND ACCESSORIES	Radio Set Philips 3 Band
EQUIPMENT AND ACCESSORIES	T.V. Set Philips 21 inch(Colored)
JEWELLERY	Seiko-21 Jewels Automatic (Women)
JEWELLERY	Seiko-21 Jewels Automatic (Men)
PERSONAL CARE	GIV Toilet Soap
PERSONAL CARE	Zenith Hair Oil(Liquids Form)
PERSONAL CARE	Zenith Hair Oil(Non-Liquids Form)
TRANSPORT	Motor Oil (Mobil)

3.7.2 Additional results

Table 3.14: ***City-level regressions-local consumer price index (adjusted prices)***

	Ln SPI	Ln VA	Ln EPI	Ln TPI
	(1)	(2)	(3)	(4)
Ln Population	0.032[a]	-0.042[a]	-0.010	-0.007
	(0.008)	(0.007)	(0.012)	(0.018)
Ln Remoteness	0.125[b]	0.127[a]	0.252[a]	0.547[a]
	(0.052)	(0.029)	(0.063)	(0.115)
Observations	105	105	105	105
R-squared	0.26	0.46	0.18	0.37

Notes: The dependent variable is the log spatial price index (Ln *SPI*) in columns 1, the log spatial availability index (Ln *VA*) in columns 2, the log exact price index (Ln *EPI*=Ln *SPI* + Ln *VA*) in column 3, the alternative spatial price index in column 4. Robust standard errors in parentheses with [a], [b] and [c] respectively denoting significance at the 1, 5 and 10% levels.

Table 3.15: ***Migration and real income per capita***

	Migration share							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Ln destination employment rate	0.069	0.003	-0.013	0.054	0.218[c]	0.147	0.111	0.141
	(0.108)	(0.098)	(0.094)	(0.100)	(0.124)	(0.116)	(0.113)	(0.115)
Ln distance	-0.091[a]	-0.091[a]	-0.090[a]	-0.086[a]	-0.079[a]	-0.080[a]	-0.081[a]	-0.081[a]
	(0.024)	(0.023)	(0.022)	(0.022)	(0.021)	(0.021)	(0.020)	(0.021)
Ln destination area	0.031[a]	0.032[a]	0.037[a]	0.038[a]	0.030[a]	0.031[a]	0.035[a]	0.035[a]
	(0.009)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Ln destination pop	0.046[a]	0.046[a]	0.036[a]	0.032[a]	0.035[a]	0.035[a]	0.029[a]	0.029[a]
	(0.010)	(0.009)	(0.008)	(0.009)	(0.008)	(0.007)	(0.007)	(0.008)
Ln income per capita at destination	0.146[b]				0.143[b]			
	(0.062)				(0.058)			
Ln destination remoteness					-0.204[a]	-0.194[a]	-0.165[a]	-0.125[b]
					(0.056)	(0.055)	(0.055)	(0.059)
Ln real income per capita at destination (SPI)		0.160[a]				0.152[a]		
		(0.055)				(0.052)		
Ln real income per capita at destination (EPI)			0.197[a]				0.178[a]	
			(0.050)				(0.049)	
Ln real income per capita at destination (TPI)				0.201[a]				0.172[a]
				(0.048)				(0.052)
Observations	110	110	110	110	110	110	110	110
R-squared	0.47	0.48	0.51	0.51	0.51	0.51	0.53	0.52
Adjusted R-squared	0.39	0.40	0.43	0.43	0.42	0.43	0.45	0.44
Origin fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The dependent variable is the share of migrants from a given region who have decided to move to another region. [a], [b] and [c] respectively denoting significance at the 1, 5 and 10% levels. SPI is the spatial price index, EPI the exact price index and TPI the travel price index.

Table 3.16: ***Migration, income per capita and adjusted prices***

	Migration share							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Ln destination employment rate	0.069	-0.059	-0.245[b]	0.023	0.218[c]	0.096	-0.103	0.004
	(0.108)	(0.115)	(0.110)	(0.099)	(0.124)	(0.112)	(0.124)	(0.131)
Ln distance	-0.091[a]	-0.094[a]	-0.101[a]	-0.089[a]	-0.079[a]	-0.082[a]	-0.092[a]	-0.090[a]
	(0.024)	(0.025)	(0.022)	(0.021)	(0.021)	(0.022)	(0.021)	(0.023)
Ln destination area	0.031[a]	0.030[a]	0.033[a]	0.035[a]	0.030[a]	0.029[a]	0.032[a]	0.036[a]
	(0.009)	(0.009)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Ln destination pop	0.046[a]	0.048[a]	0.033[a]	0.027[a]	0.035[a]	0.036[a]	0.028[a]	0.027[a]
	(0.010)	(0.010)	(0.008)	(0.008)	(0.008)	(0.008)	(0.007)	(0.008)
Ln income per capita at destination	0.146[b]	0.121[c]	0.083	0.088	0.143[b]	0.120[c]	0.091	0.085
	(0.062)	(0.066)	(0.062)	(0.059)	(0.058)	(0.063)	(0.060)	(0.063)
Ln destination remoteness					-0.204[a]	-0.201[a]	-0.127[b]	0.023
					(0.056)	(0.054)	(0.049)	(0.074)
Ln adjusted SPI at destination		-0.254				-0.236		
		(0.164)				(0.150)		
Ln adjusted EPI at destination			-0.548[a]				-0.463[a]	
			(0.121)				(0.107)	
Ln adjusted TPI at destination				-0.388[a]				-0.410[a]
				(0.104)				(0.147)
Observations	110	110	110	110	110	110	110	110
R-squared	0.47	0.48	0.53	0.54	0.51	0.51	0.54	0.54
Adjusted R-squared	0.39	0.39	0.45	0.46	0.42	0.42	0.46	0.45
Origin fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The dependent variable is the share of migrants from a given region who have decided to move to another region. [a], [b] and [c] respectively denoting significance at the 1, 5 and 10% levels. SPI is the spatial price index, EPI the exact price index and TPI the travel price index. Incomes are deflated using prices which are adjusted for spatial differences in term of products quality.

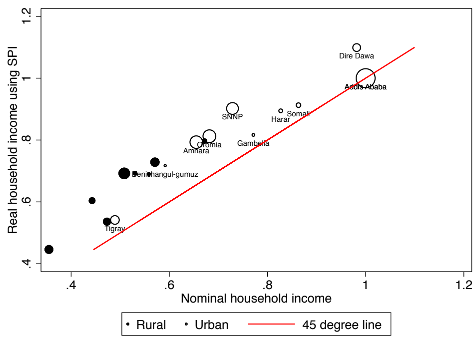
Table 3.17: ***Migration and real per capita income (adjusted prices)***

	Migration share							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Ln destination employment rate	0.069	0.000	-0.015	0.068	0.218[c]	0.149	0.118	0.148
	(0.108)	(0.099)	(0.094)	(0.101)	(0.124)	(0.116)	(0.113)	(0.116)
Ln distance	-0.091[a]	-0.092[a]	-0.091[a]	-0.086[a]	-0.079[a]	-0.081[a]	-0.082[a]	-0.081[a]
	(0.024)	(0.023)	(0.022)	(0.022)	(0.021)	(0.021)	(0.020)	(0.021)
Ln destination area	0.031[a]	0.031[a]	0.036[a]	0.038[a]	0.030[a]	0.030[a]	0.034[a]	0.035[a]
	(0.009)	(0.008)	(0.008)	(0.008)	(0.008)	(0.007)	(0.008)	(0.008)
Ln destination pop	0.046[a]	0.046[a]	0.037[a]	0.030[a]	0.035[a]	0.035[a]	0.029[a]	0.028[a]
	(0.010)	(0.009)	(0.008)	(0.008)	(0.008)	(0.007)	(0.007)	(0.008)
Ln income per capita at destination	0.146[b]				0.143[b]			
	(0.062)				(0.058)			
Ln destination remoteness					-0.204[a]	-0.202[a]	-0.176[a]	-0.117[b]
					(0.056)	(0.055)	(0.055)	(0.058)
Ln real income per capita at destination (adjusted SPI)		0.142[a]				0.138[a]		
		(0.050)				(0.047)		
Ln real income per capita at destination (adjusted EPI)			0.175[a]				0.161[a]	
			(0.046)				(0.045)	
Ln real income percapita at destination (adjusted TPI)				0.180[a]				0.155[a]
				(0.040)				(0.043)
Observations	110	110	110	110	110	110	110	110
R-squared	0.47	0.48	0.50	0.51	0.51	0.51	0.52	0.52
Adjusted R-squared	0.39	0.40	0.42	0.44	0.42	0.43	0.44	0.44
Origin fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

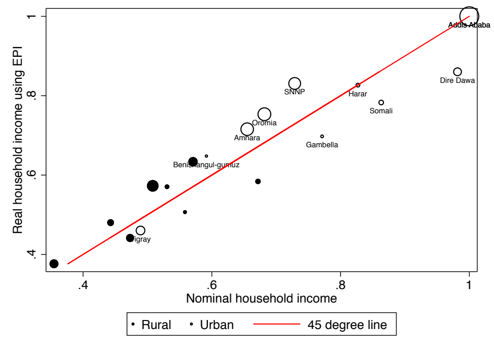
Notes: The dependent variable is the share of migrants from a given region who have decided to move to another region. [a], [b] and [c] respectively denoting significance at the 1, 5 and 10% levels. SPI is the spatial price index, EPI the exact price index and TPI the travel price index. Incomes are deflated using prices which are adjusted for spatial differences in term of products quality.

3.7.3 Additional figures

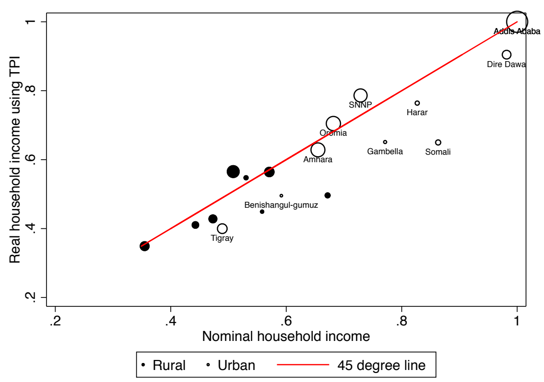
Figure 3.9: ***Regional disparity of income***



(a) ***SPI***



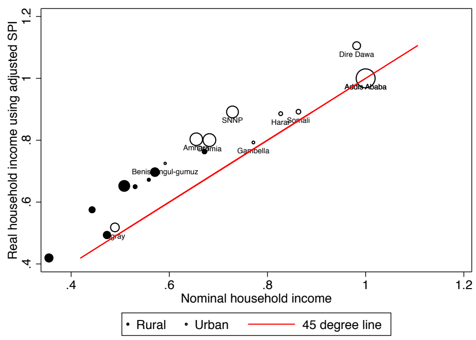
(b) ***EPI***



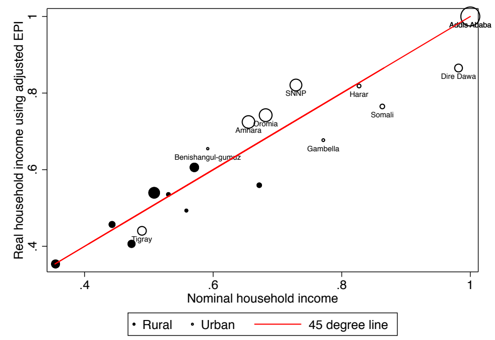
(c) ***TPI***

Notes: This graph presents the spatial disparity of nominal and real income. Incomes are normalized using Addis-Ababa as reference. Each circle represents an Ethiopian Region. The size of the circle is proportional to region's size. In panel (a) the spatial price index is used to deflate nominal income, the exact price index is used in panel (b), and the travel price index is used in panel (c).

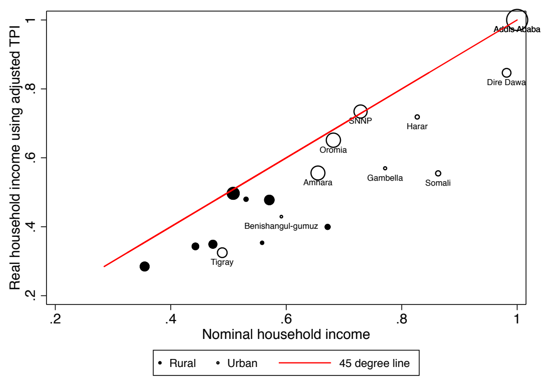
Figure 3.10: ***Regional disparity of income (adjusted prices)***



(a) ***SPI***



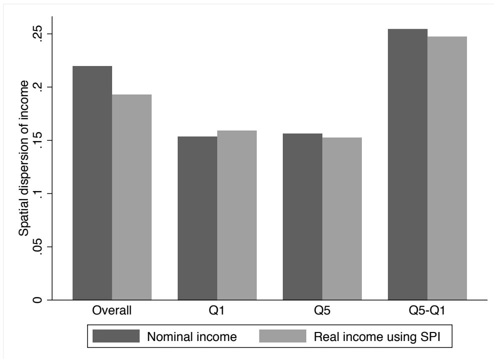
(b) ***EPI***



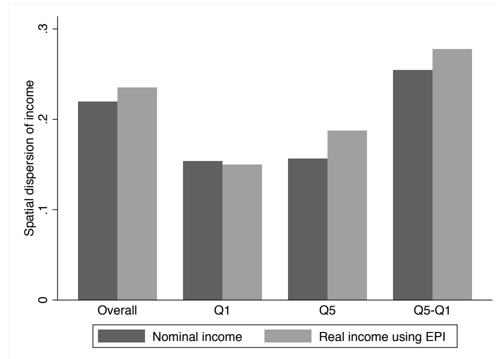
(c) ***TPI***

Notes: This graph presents the spatial disparity of nominal and real income. Each circle represents an Ethiopian Region. The size of the circle is proportional to region's size. Incomes are normalized using Addis-Ababa as reference and are deflated using prices which are adjusted for spatial differences in terms of product quality. In panel (a) the spatial price index is used to deflate nominal income, the exact price index is used in panel (b), and the travel price index is used in panel (c).

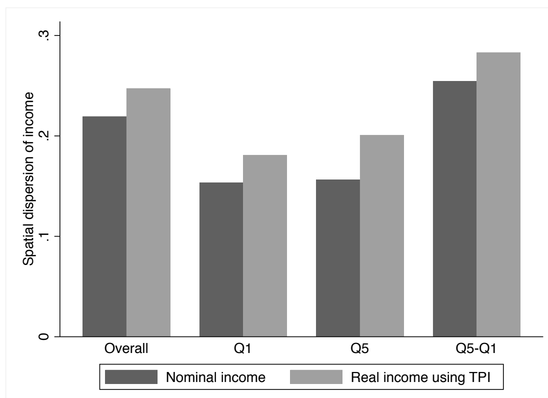
Figure 3.11: ***Coefficient of variation of income***



(a) ***SPI***



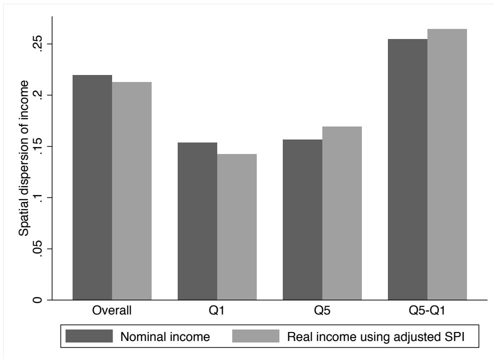
(b) ***EPI***



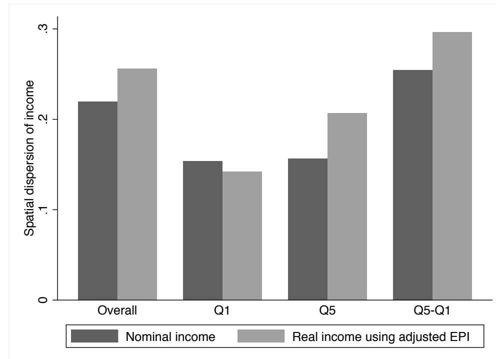
(c) ***TPI***

This graph describes the dispersion of nominal and real income using the coefficient of variation as measure. Q1 and Q5 represent respectively the first and fifth quintiles of income. Incomes are deflated using prices which are adjusted for spatial differences in terms of product quality. In panel (a) the spatial price index is used to deflate nominal income, the exact price index is used in panel (b), and the travel price index is used in panel (c).

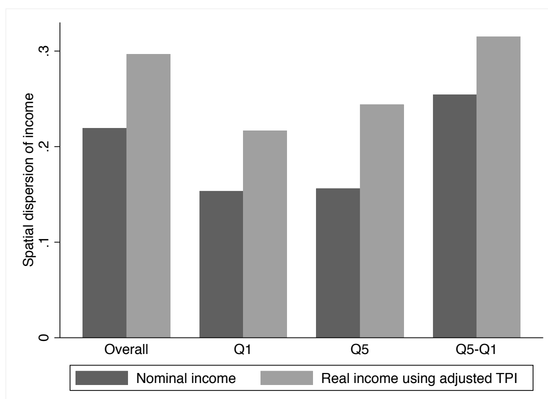
Figure 3.12: ***Coefficient of variation of income (adjusted prices)***



(a) ***SPI***



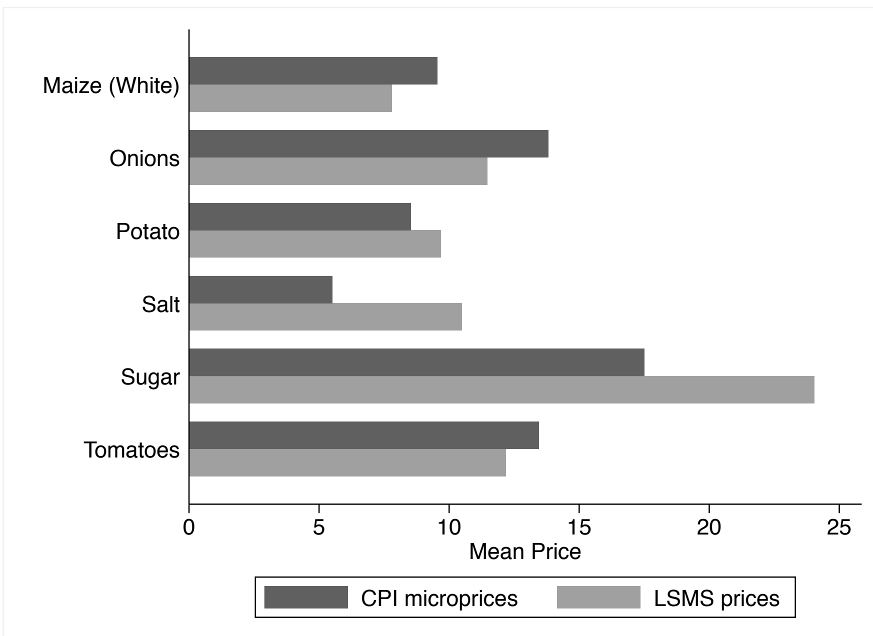
(b) ***EPI***



(c) ***TPI***

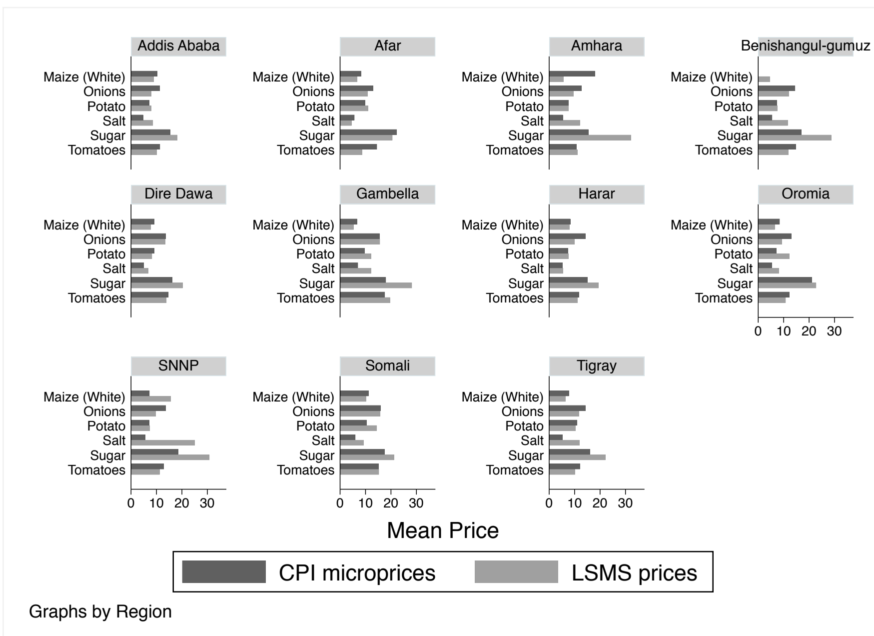
Notes: This graph de the dispersion of nominal and real income using the coefficient of variation as measure. Q1 and Q5 represent respectively the first and fifth quintiles of income. In panel (a) the spatial price index is used to deflate nominal income, the exact price index is used in panel (b), and the travel price index is used in panel (c).

Figure 3.13: ***Prices comparison: CPI versus LSMS***



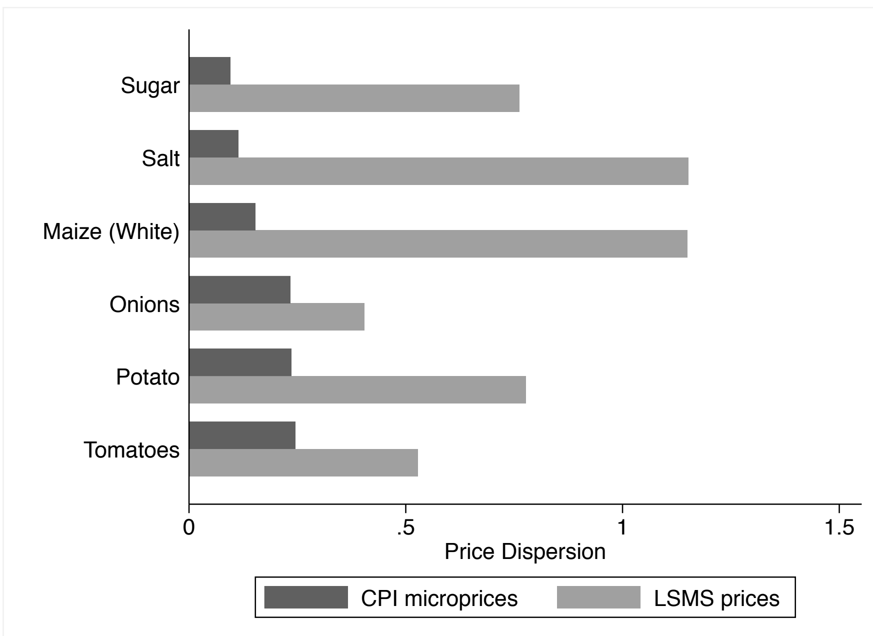
Notes: This graph compares prices for six major items using CPI data and data from the World Bank Living Standards Measurement Study (LSMS).

Figure 3.14: ***Prices comparison per region: CPI versus LSMS***



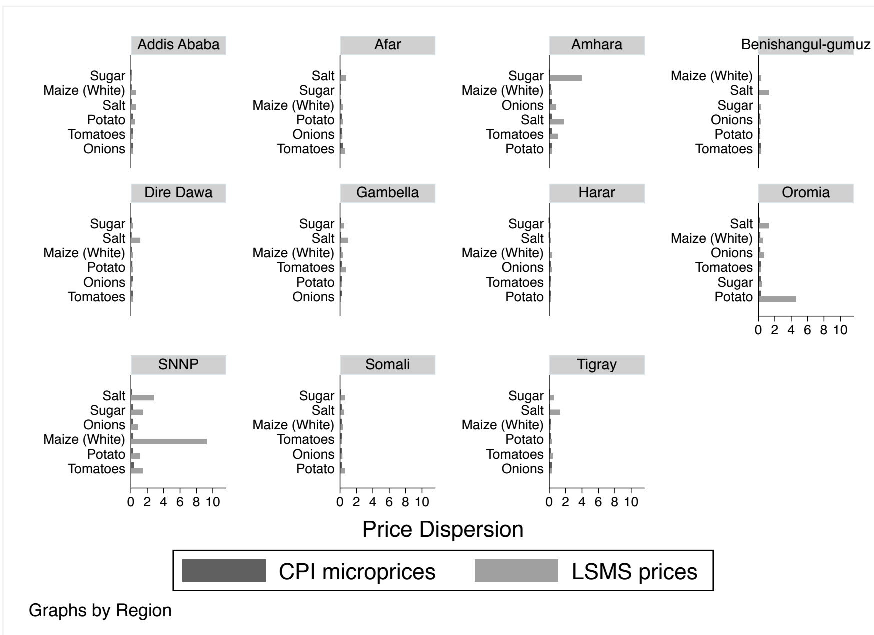
Notes: This graph compares for each region, the price of six major items using CPI data and data from the World Bank Living Standards Measurement Study (LSMS).

Figure 3.15: ***Price dispersion : CPI versus LSMS***



Notes: This graph compares the coefficient of variation of prices for six major items using CPI data and data from the World Bank Living Standards Measurement Study (LSMS).

Figure 3.16: ***Price dispersion per region : CPI versus LSMS***



Notes: This graph compares the coefficient of variation of prices for six major items using CPI data and data from the World Bank Living Standards Measurement Study (LSMS). We compute these coefficients for each region.

CONCLUSION

Cette thèse propose trois chapitres liés à la géographie de l'activité économique dans les pays africains. Elle a pour objectif de répondre aux questions suivantes: (i) la méthode d'estimation des coûts au commerce interne à partir des différences spatiales de prix est-elle valide? (ii) quelle est la taille des coûts au commerce interne au Rwanda et quels sont les effets d'une meilleure connectivité routière sur ces coûts? (iii) qu'est-ce qui détermine la différence spatiale du coût de la vie en Éthiopie et quelles en sont les implications? Le premier chapitre évalue la méthode d'estimation des coûts au commerce interne à partir de la différence spatiale des prix. Le second chapitre utilise cette méthode pour mesurer la taille des coûts au commerce interne au Rwanda; puis analyse l'effet d'une meilleure connectivité routière sur ces coûts. Le dernier chapitre examine les déterminants de la différence spatiale du coûts de la vie à travers les villes en Éthiopie et son implication en termes de flux migratoires.

Trois séries de résultats se dégagent de cette thèse. En premier lieu, l'estimation des coûts au commerce interne à partir de la différence spatiale des prix évalue correctement les coûts réels. En effet, les résultats montrent que l'élasticité des coûts au commerce interne par rapport à la distance estimée à partir de variation spatiale des prix est similaire à l'élasticité des coûts au commerce interne observée. Ce résultat reste valable lorsque l'on considère des mesures alternatives de distance.
En second lieu, cette thèse montre que les coûts au commerce interne au Rwanda sont importants. En effet, ils sont 10 fois plus élevés qu'aux États-Unis, 3 fois plus élevés qu'en Éthiopie et 2 fois plus élevés qu'au Nigéria. Par ailleurs, une meilleure connectivité routière se traduit par des coûts au commerce interne plus faibles et une plus grande disponibilité des biens.
En dernier lieu, s'agissant de la dispersion spatiale du coût de la vie à travers les villes en Éthiopie; les résultats indiquent qu'elle est liée à l'éloignement des villes mais pas

systématiquement à leurs tailles. En outre, cette différence de coût de la vie explique bien les flux migratoires. Les migrants quittent les villes où le coût de la vie est élevé pour s'installer dans les villes à plus faibles coûts de la vie.

Cette thèse apporte trois contributions à la littérature économique. D'abord, elle évalue une mesure indirecte des coûts au commerce interne avec une mesure directe. Cet exercice n'avait pas encore été effectué dans la littérature sur les coûts au commerce interne qui est assez récente. Ensuite, elle documente la littérature sur la magnitude des coûts au commerce interne dans les pays africains ainsi que sur le rôle que jouent les infrastructures dans un tel contexte. Enfin, cette thèse enrichie la littérature sur la mesure du coût de la vie en proposant des indices agrégés de coûts de la vie qui prennent à la fois les prix et la disponibilité des produits. En particulier, cette thèse propose un indice du coût de la vie qui intègre le fait que les consommateurs peuvent se déplacer pour acquérir des biens lorsque ceux-ci sont indisponibles dans leur ville.

Pour étendre le travail présenté dans cette thèse, nous envisageons trois pistes de recherche futures. La première piste consiste à évaluer les conséquences des infrastructures routières sur l'activité économique et donc à élargir l'analyse effectuée dans le deuxième chapitre de cette thèse où nous nous sommes appesantis sur les effets des infrastructures routières sur les coûts au commerce. Une attention particulière sera accordée à la localisation des activités économiques et ses corollaires en termes de distribution des salaires et des revenus. Dans la seconde piste de recherche, nous comptons analyser plus en détail le lien entre les migrations internes et le coût de la vie. Il s'agira notamment de mettre l'accent sur les migrations entre les villes et non plus seulement entre les régions comme dans le troisième chapitre de cette thèse. Les pays africains sont en général caractérisés par des niveaux de corruption élevés. Cette corruption endémique est concomitante avec des coûts au commerce élevés. Ce constat nous emmène à la troisième piste de recherche qui explorera et quantifiera l'impact de la corruption sur les coûts au commerce dans le contexte africain.

BIBLIOGRAPHY

[1] Addis Ababa Chamber of Commerce and Sectoral Associations, *The Management of Commercial Road Transport in Ethiopia,* May 2009.

[2] AFDB, "Infrastructure and Urban Development: Annual Report," *Infrastructure and Urban Development Department, African Development Bank, Abidjan*, 2018.

[3] Aggarwal, Shilpa, Brian Giera, Dahyeon Jeong, Jonathan Robinson, and Alan Spearot, "Market Access, Trade Costs, and Technology Adoption: Evidence from Northern Tanzania," NBER Working Papers 25253, National Bureau of Economic Research, Inc November 2018.

[4] Agnosteva, Delina E, James E Anderson, and Yoto V Yotov, "Intra-national trade costs: Assaying regional frictions," *European Economic Review*, 2019, *112*, 32–50.

[5] Albrecht, Lukas and Trevor Tombe, "Internal trade, productivity and interconnected industries: A quantitative analysis," *Canadian Journal of Economics/Revue canadienne d'économique*, 2016, *49* (1), 237–263.

[6] Allen, Treb and Costas Arkolakis, "The welfare effects of transportation infrastructure improvements," Technical Report, National Bureau of Economic Research 2019.

[7] Anderson, James E and Eric Van Wincoop, "Gravity with gravitas: a solution to the border puzzle," *The american economic review*, 2003, *93* (1), 170–192.

[8] ____ and ____ , "Trade costs," *Journal of Economic literature*, 2004, *42* (3), 691–751.

[9] Ansu, Y, M Mcmillan, P Pade, and D Wilem, "Promoting Manufacturing in Africa," *African Transformation Forum 2016*, 2016.

[10] Atkin, David and Dave Donaldson, "Who's getting globalized? Intra-national trade costs and world price pass-through in South Asia and Sub-Saharan Africa," *Manuscript, http://www. econ. yale. edu*, 2012.

[11] ____ and ____ , "Who's Getting Globalized? The Size and Implications of Intra-national Trade Costs," Technical Report, National Bureau of Economic Research 2015.

[12] ——, Benjamin Faber, and Marco Gonzalez-Navarro, "Retail Globalization and Household Welfare: Evidence from Mexico," *Journal of Political Economy*, 2018, *126* (1), 1–73.

[13] Baum-Snow, Nathaniel, Loren Brandt, J Vernon Henderson, Matthew A Turner, and Qinghua Zhang, "Roads, railroads, and decentralization of Chinese cities," *Review of Economics and Statistics*, 2017, *99* (3), 435–448.

[14] Behrens, Kristian and Marc Brown, "Transport costs, trade, and geographic concentration: Evidence from canada," *In progress, Université du Quebec à montréal*, 2016.

[15] Berg, Claudia N, Uwe Deichmann, Yishen Liu, and Harris Selod, "Transport Policies and Development," *The Journal of Development Studies*, 2017, *53* (4), 465–480.

[16] Berger, Thor and Kerstin Enflo, "Locomotives of local growth: The short-and long-term impact of railroads in Sweden," *Journal of Urban Economics*, 2017, *98*, 124–138.

[17] Blankespoor, Brian, Théophile Bougna, Rafael Garduno-Rivera, and Harris Selod, *Roads and the geography of economic activities in Mexico*, The World Bank, 2017.

[18] Blimpo, Moussa P, Robin Harding, and Leonard Wantchekon, "Public investment in rural infrastructure: Some political economy considerations," *Journal of African Economies*, 2013, *22* (suppl_2), ii57–ii83.

[19] Brenton, Paul, Alberto Portugal-Perez, and Julie Régolo, *Food prices, road infrastructure, and market integration in Central and Eastern Africa*, The World Bank, 2014.

[20] Broda, Christian and David E Weinstein, "Understanding international price differences using barcode data," Technical Report, National Bureau of Economic Research 2008.

[21] ——, Joshua Greenfield, and David E. Weinstein, "From groundnuts to globalization: A structural estimate of trade and growth," *Research in Economics*, 2017, *71* (4), 759 – 783. SPECIAL ISSUE IN HONOR OF 40 YEARS OF THE DIXIT-STIGLITZ MODEL.

[22] Bryan, Gharad and Melanie Morten, "The Aggregate Productivity Effects of Internal Migration: Evidence from Indonesia," *Journal of Political Economy*, 2019, *127* (5), 2229–2268.

[23] Bulow, Jeremy I and Paul Pfleiderer, "A note on the effect of cost changes on prices," *Journal of political Economy*, 1983, *91* (1), 182–185.

[24] Bundervoet, Tom, *Internal Migration in Ethiopia: Evidence from a Quantitative and Qualitative Research Study*, World Bank, 2018.

[25] Chen, Natalie and Dennis Novy, "On the measurement of trade costs: direct vs. indirect approaches to quantifying standards and technical regulations," *World Trade Review*, 2012, *11* (3), 401–414.

[26] Couture, V, "Valuing the consumption benefits of urban density," *Working Paper*, 01 2014, pp. 1–50.

[27] Crozet, Matthieu, "Do migrants follow market potentials? An estimation of a new economic geography model," *Journal of Economic Geography*, 2004, *4* (4), 439–458.

[28] CSA, *Report on the 2005 National Labour Force Survey*, Statistical bulletin, May 2006.

[29] ——, *The 2007 Population and Housing Census of Ethiopia*, Statistical Report at Country Level, 2008.

[30] ——, *Household consumption and expenditure (HCE) survey 2010/11*, Analytical report, October 2012.

[31] ——, *Population Projection of Ethiopia for All Regions At Wereda Level from 2014-2017,* August 2013.

[32] ——, *Analytical report on the 2013 National Labour Force Survey*, Statistical bulletin, June 2014.

[33] ——, *The 2015/16 Ethiopian Household Consumption-Expenditure(HCE) Survey, Results For Country Level*, Statistical Report, January 2018.

[34] ——, *Ethiopia - Socioeconomic Survey 2015-2016, Wave 3*, Report, April 2018.

[35] de Rochambeau, Golvine, "Is Liberia Getting Globalized? Estimating Intranational Trade Costs in Liberia," *Working paper*, 2018.

[36] —— and Jonas Hjort, "Intra-national trade costs and economic isolation," *International Growth Centre Blog*, 2017.

[37] de Soyres, François, Alen Mulabdic, Siobhan Murray, Nadia Rocha, and Michele Ruta, "How much will the Belt and Road Initiative reduce trade costs?," *International Economics*, 2019, *159* (C), 151–164.

[38] Deaton, Angus and Alessandro Tarozzi, "Prices and poverty in India," *Princeton, July*, 2000.

[39] Demir, Banu and Beata Javorcik, "Forensics, Elasticities and Benford's Law," CEPR Discussion Papers 12798, C.E.P.R. Discussion Papers 2018.

[40] Dercon, Stefan and John Hoddinott, "Livelihoods, growth, and links to market towns in 15 Ethiopian villages," Technical Report 2005.

[41] Diewert, W. Erwin and Robert C. Feenstra, "Estimating the Benefits of New Products: Some Approximations," in "Big Data for 21st Century Economic Statistics" NBER Chapters, National Bureau of Economic Research, Inc, March 2019.

[42] Donaldson, D, A Jinhage, and E Verhoogen, "Beyond borders: Making transport work for African trade," *IGC Growth Brief Series 009. London: International Growth Centre*, 2017.

[43] Donaldson, Dave, "Railroads of the Raj: Estimating the impact of transportation infrastructure," *American Economic Review*, 2018, *108* (4-5), 899–934.

[44] ___ and Richard Hornbeck, "Railroads and American economic growth: A ?market access? approach," *The Quarterly Journal of Economics*, 2016, *131* (2), 799–858.

[45] Duranton, Gilles and Matthew A Turner, "The fundamental law of road congestion: Evidence from US cities," *American Economic Review*, 2011, *101* (6), 2616–52.

[46] Eaton, Jonathan and Samuel Kortum, "Technology, geography, and trade," *Econometrica*, 2002, *70* (5), 1741–1779.

[47] Engel, Charles and John H Rogers, "How wide is the border?," Technical Report, National Bureau of Economic Research 1994.

[48] Faber, Benjamin, "Trade integration, market size, and industrialization: evidence from China's National Trunk Highway System," *Review of Economic Studies*, 2014, *81* (3), 1046–1070.

[49] Fafchamps, Marcel, *Market institutions in sub-Saharan Africa: Theory and evidence*, MIT press, 2003.

[50] ___ and Forhad Shilpi, "Determinants of the Choice of Migration Destination," *Oxford Bulletin of Economics and Statistics*, 2013, *75* (3), 388–409.

[51] ___ , Ruth Hill, Aliziki Kaudha et al., "The transmission of international commodity prices to domestic producers," Technical Report, EconWPA 2004.

[52] FAO, IFAD, WHO WFP, and UNICEF, "The State of Food Insecurity in the World 2018. Building climate resilience for food security and nutrition," 2018.

[53] ——, ——, and ——, "The State of Food Insecurity in the World 2019. Safeguarding against economic slowdowns and downturns," 2019.

[54] Feenstra, Robert C, "New Product Varieties and the Measurement of International Prices," *American Economic Review*, March 1994, *84* (1), 157–177.

[55] Feenstra, Robert C., Mingzhi Xu, and Alexis Antoniades, "What is the Price of Tea in China? Towards the Relative Cost of Living in Chinese and U.S. Cities," *The Economic Journal*, 2019.

[56] Ferré, Céline, Francisco H.G. Ferreira, and Peter Lanjouw, "Is There a Metropolitan Bias? The relationship between poverty and city size in a selection of developing countries," *World Bank Economic Review*, 2012, *26* (3), 351–382.

[57] Freund, Caroline L and Nadia Rocha, "What constrains Africa's exports?," *Staff working paper ERSD*, 2010, (2010-07).

[58] Gollin, Douglas, David Lagakos, and Michael E. Waugh, "The Agricultural Productivity Gap," *The Quarterly Journal of Economics*, 2014, *129* (2), 939–993.

[59] ——, Martina Kirchberger, and David Lagakos, "In Search of a Spatial Equilibrium in the Developing World," CEPR Discussion Papers 12114, C.E.P.R. Discussion Papers June 2017.

[60] Gunning, Jan Willem, Pramila Krishnan, Andualem T Mengistu et al., "Fading Choice: Transport Costs and Variety in Consumer Goods," Technical Report, Centre for the Study of African Economies, University of Oxford 2018.

[61] Handbury, Jessie, "Are Poor Cities Cheap for Everyone? Non-Homotheticity and the Cost of Living Across U.S. Cities," *SSRN Electronic Journal*, 01 2013.

[62] —— and David E Weinstein, "Goods prices and availability in cities," *The Review of Economic Studies*, 2015, *82* (1), 258–296.

[63] Hartmann, O and E Asebe, "Road Transport Industry in East Africa," *SSATP, unpublished document*, 2013.

[64] Hausman, Jerry, "Valuation of New Goods under Perfect and Imperfect Competition," in "The Economics of New Goods," National Bureau of Economic Research, Inc, 1996, pp. 207–248.

[65] Hering, Laura and Rodrigo Paillacar, “Does Access to Foreign Markets Shape Internal Migration? Evidence from Brazil,” *World Bank Economic Review*, 2016, *30* (1), 78–103.

[66] Hsieh, Chang-Tai and Enrico Moretti, “Housing Constraints and Spatial Misallocation,” *American Economic Journal: Macroeconomics*, 2019, *11* (2), 1–39.

[67] Hummels, David, “Toward a Geography of Trade Costs,” *Citeseer*, 2001.

[68] Irarrazabal, Alfonso, Andreas Moxnes, and Luca David Opromolla, “The tip of the iceberg: a quantitative framework for estimating trade costs,” *Review of Economics and Statistics*, 2015, *97* (4), 777–792.

[69] Jedwab, Remi and Adam Storeygard, “The Average and Heterogeneous Effects of Transportation Investments: Evidence from sub-Saharan Africa 1960-2010,” Discussion Papers Series, Department of Economics, Tufts University 0822, Department of Economics, Tufts University 2017.

[70] Judge, George and Laura Schechter, “Detecting Problems in Survey Data Using Benford’s Law,” *The Journal of Human Resources*, 2009, *44* (1), 1–24.

[71] Kondylis, Florence and Valerie Mueller, “Economic consequences of conflict and environmental displacement,” in “International Handbook on Migration and Economic Development,” Edward Elgar Publishing, 2014.

[72] Krishnan, Pramila and Peng Zhang, “Restricting Trade and Reducing Variety: Evidence from Ethiopia,” Technical Report 2018.

[73] Krugman, Paul, “Increasing returns and economic geography,” *Journal of political economy*, 1991, *99* (3), 483–499.

[74] Krugman, Paul R, *Geography and trade*, MIT press, 1991.

[75] Lagakos, David, Ahmed Mushfiq Mobarak, and Michael E. Waugh, “The Welfare Effects of Encouraging Rural-Urban Migration,” NBER Working Papers 24193, National Bureau of Economic Research, Inc 2018.

[76] ——, Mushfiq Mobarak, and Michael E Waugh, “The Welfare Effects of Encouraging Rural-Urban Migration,” *Working Paper*, 2020.

[77] Leamer, Edward E, “A flat world, a level playing field, a small world after all, or none of the above? A review of Thomas L. Friedman’s The World is Flat,” *Journal of Economic Literature*, 2007, *45* (1), 83–126.

[78] Limao, Nuno and Anthony J Venables, "Infrastructure, geographical disadvantage, transport costs, and trade," *The World Bank Economic Review*, 2001, *15* (3), 451–479.

[79] Matsa, David A, "Competition and product quality in the supermarket industry," *The Quarterly Journal of Economics*, 2011, *126* (3), 1539–1591.

[80] Maza, Adolfo and José Villaverde, "Interregional migration in Spain: a semiparametric analysis," *Review of Regional Studies*, 2004, *34* (2), 156–171.

[81] Ministry of Water and Energy, *Urban Water Supply Access Plan (Part III)*, The Federal Democratic Republic of Ethiopia, December 2011.

[82] Monras, Joan, "Economic Shocks and Internal Migration," CEPR Discussion Papers 12977, Center for Economic Policy Research 2018.

[83] Moretti, Enrico, "Real Wage Inequality," *American Economic Journal: Applied Economics*, January 2013, *5* (1), 65–103.

[84] Moulton, Brent R, "An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Unit," *The Review of Economics and Statistics*, May 1990, *72* (2), 334–338.

[85] Muller, Christophe, "Prices and living standards: evidence for Rwanda," *Journal of Development Economics*, 2002, *68* (1), 187–203.

[86] Nagothu, Udaya Sekhar, *Food security and development: country case studies*, Routledge, 2014.

[87] Nakamura, Emi and Jón Steinsson, "Five facts about prices: A reevaluation of menu cost models," *The Quarterly Journal of Economics*, 2008, *123* (4), 1415–1464.

[88] Olken, Benjamin A, "Monitoring corruption: evidence from a field experiment in Indonesia," *Journal of political Economy*, 2007, *115* (2), 200–249.

[89] ___ and Patrick Barron, "The simple economics of extortion: evidence from trucking in Aceh," *Journal of Political Economy*, 2009, *117* (3), 417–452.

[90] Parsley, David C and Shang-Jin Wei, "Explaining the border effect: the role of exchange rate variability, shipping costs, and geography," *Journal of International Economics*, 2001, *55* (1), 87–105.

[91] Porteous, Obie, "High trade costs and their consequences: An estimated dynamic model of African agricultural storage and trade," *American Economic Journal: Applied Economics*, 2019, *11* (4), 327–66.

[92] Prager, Jean-Claude, *Le Grand Paris Express: les enjeux économiques et urbains*, Economica, 2019.

[93] Raballand, Gael and Supee Teravaninthorn, “Transport Prices and Costs in Africa: A Review of the International Corridors,” *World Bank Publications*, 2009.

[94] Rajashekar, Anirudh, Marion Richard, and Dimitri Stoelinga, “The economic geography of Rwanda,” 2019, *C-38415-RWA-1*.

[95] Rancourt, Marie-Ève, François Bellavance, and Jarrod Goentzel, “Market analysis and transportation procurement for food aid in Ethiopia,” *Socio-Economic Planning Sciences*, 2014, *48* (3), 198–219.

[96] Ravallion, Martin and Dominique Van De Walle, “Urban-rural cost-of-living differentials in a developing economy,” *Journal of Urban Economics*, 1991, *29* (1), 113–127.

[97] Redding, Stephen J and Matthew A Turner, “Transportation costs and the spatial organization of economic activity,” in “Handbook of regional and urban economics,” Vol. 5, Elsevier, 2015, pp. 1339–1398.

[98] Rizet, Christophe and Henri Gwét, “Une comparaison internationale des prix du camionnage Afrique, Asie du sud-est, Amérique centrale,” *Recherche-Transports-Sécurité*, 1998, *60*, 69–85.

[99] Roberts, Mark, Martin Melecky, Théophile Bougna, and Yan Xu, “Transport corridors and their wider economic benefits: A quantitative review of the literature,” *Journal of Regional Science*, 2020, *60* (2), 207–248.

[100] Rochambeau, Golvine De, “Essays on Development Economics.” PhD dissertation, Columbia University 2018.

[101] Schiff, Nathan, “Cities and product variety: evidence from restaurants,” *Journal of Economic Geography*, 2015, *15* (6), 1085–1123.

[102] Sequeira, Sandra, “Corruption, trade costs, and gains from tariff liberalization: evidence from Southern Africa,” *The American Economic Review*, 2016, *106* (10), 3029–3063.

[103] Sotelo, Sebastian, “Domestic trade frictions and agriculture,” *Journal of Political Economy*, 2020, *128* (7).

[104] Storeygard, Adam, “Farther on down the road: transport costs, trade and urban growth in sub-Saharan Africa,” *The Review of Economic Studies*, 2016, *83* (3), 1263–1295.

[105] Sutton, John and Nebil Kellow, "An enterprise map of ethiopia, international growth centre," 2014.

[106] Timmins, Christopher, "Estimating spatial differences in the Brazilian cost of living with household location choices," *Journal of Development Economics*, 2006, *80* (1), 59–83.

[107] Weber, Sylvain and Martin Péclat, "GEOROUTE: Stata module to calculate travel distance and travel time between two addresses or two geographical points," Statistical Software Components, Boston College Department of Economics November 2016.

[108] ——, Martin Péclat et al., "GEOROUTE: Stata module to calculate travel distance and travel time between two addresses or two geographical points," *Statistical Software Components*, 2016.

[109] WFP and CSA, "Ethiopia-comprehensive food security and vulnerability analysis (CFSVA)," 2018.

[110] WTO, "World Trade Report," *World Trade Organization, Geneva*, 2004.

[111] ——, "World Trade Statistical Review," *World Trade Organization, Geneva*, 2016.

[112] Young, Alwyn, "Inequality, the Urban-Rural Gap, and Migration," *The Quarterly Journal of Economics*, 2013, *128* (4), 1727–1785.