Investigating Product Market Integration in the Southern African Development Community: A Price-Based Approach

By
Neil Balchin

Thesis Presented for the Degree of
Doctor of Philosophy

In the School of Economics
University of Cape Town

June 2015

Supervisors: Professor Lawrence Edwards
&
Asha Sundaram

Abstract

This thesis extends the price-based empirical literature on border effects and product market integration to the Southern African Development Community (SADC). The analysis draws on a unique integrated dataset of district-level monthly average retail product prices spanning a number of districts in Botswana, Malawi, South Africa and Zambia. The thesis is comprised of four main chapters.

The first main chapter provides an empirical analysis of the extent to which product prices are integrated within and between the four SADC countries. The results reveal large and persistent absolute deviations from the law of one price both within and between each of the countries over the period from 2006 to 2009. Price deviations are found to be higher between SADC countries than within the individual countries, although there is considerable heterogeneity in the magnitude of these deviations across products. On average, absolute price deviations between-country pairs are smaller for countries adjacent to each other and for countries that share common membership in the Southern African Customs Union (SACU). Simple econometric estimates based on the standard regression approach in the literature show that absolute price deviations between district pairs in the region increase the further apart the districts are from each other; and are 11.8% higher, on average, between districts separated by a national border. Overall, there is no clear evidence that product prices in the SADC region became more integrated between 2006 and 2009 (although product prices between the Common Market for Eastern and Southern Africa countries did become more integrated over this period), despite the liberalization of tariffs under the SADC Protocol on Trade.

The second main chapter critically evaluates the standard empirical methodology used to estimate border effects. The evaluation identifies several different sample selection effects that bias estimates of distance and border effects in the existing literature, and demonstrates the sensitivity of estimates of transaction costs and border effects for the SADC region to these sample selection biases using quantile regressions. The results show that the standard pooled OLS estimates reported in much of the existing literature (and in the first main chapter) suffer from a sample selection problem which biases the estimated distance and border coefficients downwards relative to the true cost of trade. The chapter also demonstrates the impact of two additional product and distance sample selection biases that are not dealt with in the Borraz et al. (2012) application of the quantile regression approach. Finally, it shows that not accounting for variation in within-product quality across districts results in omitted variable bias which raises the estimated distance

and border coefficients. The chapter proposes novel extensions to the quantile regression methodology to allow for the analysis of cross-country border effects; and to account for sample selection bias arising due to product and distance sample selection effects.

The third main chapter applies the modified quantile regression methodology to precisely estimate average and individual border effects in the SADC region for the 2006 to 2009 period. The quantile regression results show the effects of borders in raising price dispersion in the region are generally comparatively lower between the SACU (23.1%) and Common Market for Eastern and Southern Africa (26.6%) countries compared to the remaining SADC country pairs (36.2%). There are also clear differences in the magnitude of border effects for contiguous (24.1%) and non-contiguous (41.1%) SADC countries, providing evidence of incremental border effects as products are traded across multiple borders. While the magnitude of the border effect estimates are sensitive to the estimation technique employed, the ranking of the border effects across different regional trade agreements and for contiguous versus non-contiguous countries is robust across different specifications. Finally, the results in the third main chapter reveal that, on average, there was little change in the magnitude of border effects in the SADC region between 2006 and 2009, despite the accelerated liberalization of tariffs on intra-SADC trade over this period.

The final main chapter of the thesis unpacks the contribution of preferential tariffs to the South Africa-Zambia border effect over the period from 2002 to 2009. This addresses the lack of studies in the literature of the direct contribution of tariffs to border effects on international relative prices. The estimation results reveal an almost perfect pass-through of preferential tariffs onto domestic prices in Zambia. They also show that preferential tariffs account for a significant portion of the border effect. After accounting for the role of preferential tariffs, the additional increase in prices in the Zambian market caused by the border effect falls from 29.4% to 16.1%. This general result is qualitatively the same even after accounting for variation in the intensity with which products are traded between the two countries. Even so, a simple analysis of the trend in the South Africa-Zambia border effect indicates that the impact of crossing the border in raising Zambian retail prices actually increased between 2002 and 2009. The results do suggest, however, that the increase in the border effect over this period would have been more substantial in the absence of the phasing down of preferential tariffs.

On balance, the evidence presented in this thesis indicates that markets in the SADC region remain fragmented, with little sign of greater product market integration either within or between countries. This is despite the explicit policy focus on trade reform that has accompanied the introduction of the SADC Protocol on Trade. Trade liberalization alone appears not to be sufficient in generating greater product market integration within the region.

Acknowledgements

I would like to express my sincere gratitude to my supervisors, Professor Lawrence Edwards and Asha Sundaram, for their guidance and tireless efforts to assist me at each stage of my PhD research, from the conceptualization of the research topic right through to the finalization of the thesis. Their feedback and comments on various chapters and drafts of the thesis have made an immense contribution to the research, and their willingness to answer my queries and assist me in solving problems encountered during the course of the research have played a crucial role in supporting the eventual completion of the study. I have benefited greatly from their expertise, guidance and vast knowledge of the subject matter.

I am also indebted to my family and my girlfriend, Mardi Smit, for encouraging me to pursue full-time studies towards a PhD degree and for their unwavering love and support over the past three years. My parents, in particular, deserve special mention for instilling in me an appreciation of the importance of education. The financial and personal sacrifices that they made in order to ensure that I received the best possible schooling and tertiary education created the perfect platform from which to embark on doctoral studies.

I am most grateful to Statistics South Africa, Professor Neil Rankin, Dale Mudenda and Mamello Nchake for facilitating access to the unpublished Consumer Price Index micro-data for South Africa, Botswana, Malawi and Zambia. This data has formed the basis for the empirical research undertaken in this thesis. I would also like to thank Dale Mudenda for constructing the data series on Zambia's Most Favoured Nation tariffs and preferential tariff offers to South Africa that are used in this thesis.

Lastly, I am extremely grateful for the funding provided by the National Research Foundation, the John Davidson Educational Trust and the University of Cape Town. Without the financial support of these organizations I would not have been in a position to pursue doctoral studies on a full-time basis.

Table of Contents

List of Acronyms

CFA	*Communauté Financiére Africaine*
COMESA	Common Market for Eastern and Southern Africa
CPI	Consumer Price Index
EAC	East African Community
EU	European Union
FTA	Free Trade Agreement
HS	Harmonized System
LOP	Law of One Price
MFN	Most Favoured Nation
NAFTA	North American Free Trade Agreement
NTB	Non-tariff Barrier
OECD	Organization for Economic Co-operation and Development
OLS	Ordinary Least Squares
PPP	Purchasing Power Parity
REC	Regional Economic Community
SACU	Southern African Customs Union
SADC	Southern African Development Community
SSA	Sub-Saharan Africa
VAT	Value Added Tax
WTO	World Trade Organization

List of Tables

List of Figures

1. General Introduction

1.1 Background and Motivation

Regional integration has long been seen as a means to boost economic growth, reduce poverty and foster social development. In Africa, regional integration is particularly important given its geography and the size of countries on the continent. There are a relatively large number of landlocked countries in Africa, and many countries have low population densities and income levels, generally small domestic markets, poor transportation infrastructure, and dispersed and often disconnected rural communities. These challenges translate into high trade costs that inhibit trade flows and the movement of products within and between countries. The result is often disconnected and fragmented product markets in Africa (Hartzenberg, 2011; World Bank, 2012). This has highlighted the need for policies that drive regional integration on the continent, and is a key motivation behind the recent emphasis on promoting regional integration in Africa.

Regional integration can produce a number of benefits for Africa. Through the absence of (or very low) barriers to commerce and the movement of goods, product market integration in Africa can serve as a mechanism to lower trade and transaction costs. It can also improve firms' access to affordable inputs, essential skills and services. At the same time, regional integration facilitates deeper specialization in production through economies of scale, thereby fostering the development of regional production chains and integrated production networks. These factors can raise productivity and lead to greater diversification in production and exports. As a result, more integrated product markets are expected to experience higher volumes of bilateral trade, stronger competition and greater convergence in relative prices.

Higher levels of product market integration can also have important welfare implications in Africa. From a consumer perspective, stronger competition stemming from more integrated markets means that consumers benefit from lower prices and access to a wider array of goods and services. At the same time, as product markets become more integrated firms are likely to benefit from access to a wider range of suppliers or distributors from which to source inputs into production. Furthermore, the presence of well-integrated markets is likely to afford firms access to a significantly larger market in which to sell their products and services. Finally, better-integrated markets may be more resilient to economic shocks and, thereby, contain the impact of adverse shocks on economic growth and employment.

Significant progress has been made over the past two decades in advancing the broader vision of regional integration within the Southern African Development Community (SADC). This has been evident in notable reductions in tariffs on intra-SADC trade; a process that accelerated following the introduction of the SADC Protocol on Trade in 2000.[1] It is not clear, however, whether these tariff reductions have been accompanied by greater integration of product markets in the region.

One way to assess the extent to which product markets are integrated is to examine the volume of trade flows between markets. In SADC, the volume of intra-regional trade remains relatively low. Furthermore, trade flows within the region are highly concentrated both in terms of product coverage – the majority of intra-SADC trade occurs in the form of trade in resource-based products (Gillson, 2012) – and the dominance of trade between the Southern African Customs Union (SACU) countries (Behar & Edwards, 2011).[2]

Yet, relying on trade flows to determine the extent to which product markets are integrated can be misleading since trade volumes may be an endogenous outcome of market integration, and are also influenced by unrelated factors such as government expenditure, exchange rates, donor funding and rules of origin (Edwards & Rankin, 2012). The latter are particularly onerous in the SADC context and may indirectly affect trade flows in the region (Erasmus et al., 2004).

Market integration can also be measured along a number of alternative dimensions, including the behaviour of relative prices or price levels. Among these alternatives, techniques that measure price differentials between countries have increasingly been used. Prices should provide a good indication of changes in product market integration, even in cases where trade does not occur, since the potential for arbitrage is the key driver in determining the extent to which prices diverge (Parsley & Wei, 2002). According to Bradford and Lawrence (2004), analysing price differentials is "the most plausible" method to capture the effects of 'invisible barriers' to trade and market integration.

[1] The signatories to the Protocol agreed on schedules to phase down tariff barriers over a 12-year period beginning in 2000. The 12 SADC Member States to originally ratify the SADC Protocol on Trade in 2000 were: Botswana, Lesotho, Madagascar, Malawi, Mauritius, Mozambique, Namibia, South Africa, Swaziland, Tanzania, Zambia and Zimbabwe. Angola acceded to the Protocol in 2003 but, to date, has opted not to implement it.
[2] Levels of intra-SADC trade have been rising gradually. Moreover, the level of intra-SADC trade is actually relatively high when the geography of the region and the limited size of its individual markets are taken into account (Faroutan & Pritchett, 1993; Behar & Edwards, 2011).

Despite these benefits, there is a shortage of studies using disaggregated product price level data to examine product market integration in developing regions, particularly in Africa. Most disaggregated price-based studies of product market integration have focused on developed countries in North America and Europe (for instance, Mathä, 2003; Engel et al., 2005; Broda & Weinstein, 2008; Horváth et al., 2008; Wolszczak-Derlacz, 2008). Furthermore, of the small number of price-based studies that have examined product market integration in Africa, little attention has been afforded to the SADC region. The existing studies have either focused primarily on Central, East or West Africa (Versailles, 2012; Brenton et al., 2014; Aker et al., 2014), or have been limited to SACU countries when looking exclusively at Southern Africa (Nchake, 2013).

This thesis draws on newly available microeconomic price data to address the lack of price-based studies of product market integration in Africa, and the SADC region in particular. Specifically, the thesis uses highly disaggregated retail price data collected at the district/city level (hereafter referred to as districts)[3] for a selection of four SADC countries – Botswana, Malawi, South Africa and Zambia – to assess the extent to which product markets are integrated within and between these countries. The thesis thus contributes to the empirical literature by analysing product market integration in an understudied region. A further contribution is to use an alternative price-based approach rather than the focus in the existing literature on trade flows between SADC countries to measure integration in the region.

1.1.1 Prices and product market integration

Progress towards greater market integration should be reflected in lower price differentials for similar products across markets. This assertion is derived from the theoretical benchmark of the Law of One Price (LOP), which states that for any good *i*:

$$P_i = EP_i^* \tag{1}$$

where P_i is the domestic currency price, E is the home currency price of foreign currency and P_i^* is the foreign currency price. This framework can be used to consider variation in prices within and between countries (the nominal exchange rate, E, is equal to one in the within country case).

In essence, the LOP implies that identical goods should sell for the same price (when measured in a common currency) in different markets. It is the potential for arbitrage that underpins the

[3] The data represents average monthly product prices by district. The majority of the districts are defined by the main city in that particular area.

equalization of prices across markets as suggested by the LOP. To illustrate this concept, assume that the final retail prices of an identical good *i* in markets *j* and *k* are given by $p_{i,j}$ and $p_{i,k}$, respectively.[4] If there are no transaction costs associated with transport and other barriers to trade between the two markets, then the possibility for profitable arbitrage will exist unless the absolute value of the difference in the price of product *i* between the two markets, $|p_{i,j} - p_{i,k}|$, is equal to zero. In the event that the price differential is not equal to zero, and in the absence of any trade costs, arbitrage will occur until the difference in prices between the two markets is eliminated.

In reality, however, there are likely to be a variety of transaction costs associated with trading across markets. These include, for example, the cost to transport a good from one market to another, as well as the cost of taxes and other barriers to the free movement of goods. The existence of trade costs causes deviations from the LOP that have implications for the integration of markets. Baulch (1997) formalises the relationship between trade costs, inter-market price differentials and product market integration through a parity bounds model that distinguishes between three possible trade regimes that correspond to different levels of market integration. In this spatial model, trade costs "determine the parity bounds within which the prices of a homogenous commodity in two geographically distinct markets can vary independently" (Baulch, 1997: 479).

When two markets engage in trade, they will be spatially integrated if the price in the importing market equals the price in the exporting market plus the cost incurred in moving the good between the two markets (Baulch, 1997). This relationship is depicted as:

$$p_{i,j} + \tau_{i,jk} = p_{i,k} \tag{2}$$

where $p_{i,j}$ and $p_{i,k}$ are the prices of product *i* in markets *j* and *k*; and $\tau_{i,jk}$ represents the ad valorem equivalent of all transaction costs involved in trading product *i* between the two markets.

In this case, trade will result in a one-for-one movement in prices in markets *j* and *k* (Gopinath et al., 2011). The magnitude of deviations from the LOP are exactly equal to the cost of all transaction costs involved in trading an identical product between the two markets, such that the spatial arbitrage conditions are binding and:

$$|p_{i,j} - p_{i,k}| = \tau_{i,jk} \tag{3}$$

[4] The final retail prices embody the prices of both traded and non-traded inputs (such as labour or rent) so that the final retail price of good *i* in location *j* can be decomposed as: $P_{i,j} = (P_{NT}^{\alpha} - P_{T}^{1-\alpha})$ where P_{NT} and P_T are the prices of non-traded and traded inputs, respectively.

In contrast, there will be no incentive to trade product i between markets j and k when:

$$p_{i,j} + \tau_{i,jk} > p_{i,k} \quad (4)$$

In this alternative trade regime, structural differences in prices between the two markets are less than the transaction costs associated with trade between them, and the spatial arbitrage conditions are *not* binding:

$$|p_{i,j} - p_{i,k}| < \tau_{i,jk} \quad (5)$$

Importantly, both the trade regimes represented in equations (3) and (5) are potentially consistent with integrated markets.[5] In the case of the latter, for example, markets may be integrated in the sense that structural price differentials are stable (Ravallion, 1986). In the scenario presented in equation (5), while prices in the two markets are determined by local conditions, the size of the trade cost imposes a bound on the magnitude of the difference in prices between the two markets. Thus, in integrated markets product prices are interrelated and linked by the size of the trade cost.

In contrast, there is a third type of trade regime in which price differences exceed trade costs and the spatial arbitrage conditions are violated, such that:

$$|p_{i,j} - p_{i,k}| > \tau_{i,jk} \quad (6)$$

In this scenario, the size of the price gap means that it would be profitable for traders to purchase product i in the cheaper location and incur the transaction costs required to sell it in the higher priced location. However, there exists some barrier to trade that prevents arbitrage. This barrier breaks the link between price gaps and trade costs.[6,7] In this case, the two product markets are segmented. Product prices in each market are determined entirely by local market conditions and the size of the price gap is unrelated to the cost of trade between the two markets.

---

[5] Whether or not the second trade regime – reflected in equation (5) – is consistent with market integration will depend on the spatial location of production and consumption. In instances in which production and consumption are specialized such that product i is produced in a different geographic location from where it is consumed, only the first trade regime will be consistent with market integration (Bauch, 1997). If, however, production and consumption do not occur in distinct geographical locations, then the second regime will also be consistent with integrated markets.

[6] There could, for example, be government controls on the movement of specific goods between markets.

[7] In this sense, larger differences between inter-market price gaps and trade costs are consistent with the presence of more substantial barriers to product market integration (Brenton et al., 2014).

1.2 Thesis Objectives

The primary objective of this thesis is to examine the extent to which product markets are integrated between countries in the SADC region using disaggregated product price level data. This is done in order to address the shortage of price-based studies on product market integration in developing countries, and the lack of studies for the SADC region in particular. The thesis is structured around four specific objectives.

Objective 1

The first main objective is to provide an empirical assessment of the extent to which product prices are integrated both within and between SADC countries. This is undertaken in Chapter 2 of the thesis with three different objectives in mind. The first is to compare within-country price dispersion against price dispersion between countries in the SADC region. This is done by computing mean absolute price deviations across traded products within and between Botswana, Malawi, South Africa and Zambia. It is also achieved by exploring the extent to which there is heterogeneity in price dispersion across products within and between countries in the region. A second objective of the descriptive analysis is to understand the importance of contiguity and common membership in additional regional trade agreements for product price integration in the SADC region.

The third objective of the research in Chapter 2 is to test whether certain key stylized facts in the international literature on the sources of price dispersion and the determinants of product price integration hold in the SADC context. The main objective in this regard is to empirically assess the importance of distance, borders and membership in regional trade groupings for product price integration between countries in the SADC region. This is done using standard econometric approaches in the price-based literature on product market integration. Doing so allows for a direct comparison of the average border effect in the SADC region with border effect estimates reported in the existing literature for different regions within Africa as well as in other parts of the world.

Objective 2

The second core objective of the research is to critically evaluate the standard empirical methodology used to estimate border effects. This is undertaken in Chapter 3 through the identification of a number of sample selection biases in the existing literature. The sensitivity of transaction cost and border effect estimates to these sample selection biases is then tested using a quantile regression methodology and data for the SADC region. The findings from the sensitivity tests are used as the basis for devising extensions to the quantile regression approach. These extensions are designed to facilitate analysis of cross-country border effects using quantile regressions, and to address additional product and distance sample selection effects that are not dealt with in the Borraz et al. (2012) application of the approach.

Objective 3

The third main objective of this thesis – which is addressed in Chapter 4 – is to precisely estimate the effect of national borders on product price dispersion in the SADC region using the appropriate method identified in Chapter 3. By dealing explicitly with sample selection biases in the existing literature, the thesis provides more accurate estimates of the magnitude of specific country border effects in the region as well as the size of the average SADC border effect. It also considers whether differences in product characteristics influence the way in which distance and borders affect relative prices. The focus in this regard is on differences between perishable and non-perishable products.

Using the precisely estimated border effects, the research also evaluates whether there have been changes in the level of product market integration in the SADC region over time. This is examined by assessing whether there was any change in the size of the average and individual border effects in the SADC region over the period between 2006 and 2009. The goal here is to determine whether border effects in the region have declined over time, which would be indicative of falling border-related barriers to product market integration.

Objective 4

The final core objective of this thesis, which is addressed in Chapter 5, is to analyse the tariff contribution to border effects in the SADC region. The analysis adopts a production-consumption model to focus on products exported from South Africa into the Zambian market. The objective is to establish whether Zambian tariffs on South African imports affect domestic prices in Zambia. The research also seeks to determine whether the effects of national borders and tariffs on cross-country relative prices vary according to the intensity with which products are traded across the border.

1.3 Relevance and Contribution of the Thesis

Recent studies have highlighted a host of shortcomings associated with empirical analyses of product market integration that are based either on trade flows between countries or on price indices (Hillberry, 2002; Crucini et al., 2005a; Engel et al., 2005; Broda & Weinstein, 2008; Hillberry & Hummels, 2008; Burstein & Jaimovich, 2012; Edwards & Rankin, 2012; Inanc & Zachariadis, 2012). These shortcomings stem from compositional issues and aggregation problems that hamper the effective use of trade volume or price indices as the basis for measuring product market integration. Many of these problems can, however, be avoided by using comparisons of the actual prices of homogenous products in different markets as a metric for assessing product market integration.

This, coupled with the generally improved availability of microeconomic price level data has spaked a growing number of disaggregated price-based studies of product market integration in developed countries (for instance, Mathä, 2003; Bradford and Lawrence, 2004; Engel et al., 2005; Broda & Weinstein, 2008; Crucini et al., 2008; Horváth et al., 2008; Wolszczak-Derlacz, 2008; Gopinath et al., 2010; Burstein and Jaimovich, 2012). In contrast, little attention has been given to such analysis in developing regions. This has been largely due to data limitations related to a lack of high-frequency price level data for narrowly defined goods in developing countries. In particular, very little research has investigated product market integration in Africa using actual price level data. Moreover, the few existing price-based studies on Africa have generally focused on the East, Central and West African regions (Versailles, 2012; Aker et al., 2014; Brenton et al., 2014), meaning that product market integration in Southern Africa – and the SADC in particular – is understudied.[8]

This thesis addresses the lack of price-based studies of product market integration in the SADC region. In doing so, it makes a number of important contributions to the empirical literature. The first contribution is the construction of a unique integrated dataset of district level monthly average retail prices spanning a number of districts in four SADC countries: Botswana, Malawi, South Africa and Zambia. The dataset is constructed from raw price data collected by statistical agencies

[8] One notable exception is Nchake's (2013) study of three SACU countries.

in each country for the purposes of computing the consumer price index (CPI). The time variation and district disaggregation in the data allows for an analysis of product market integration both within and between countries and over time; while the disaggregation down to the level of narrowly defined products makes it possible to account for potential heterogeneity across products in the extent of market integration. Furthermore, the selection of countries in the dataset spans three different regional trade groupings: SADC, SACU and COMESA, thereby making it possible to explore product market integration between countries that are members of different regional trade and monetary arrangements. The results of the research thus contribute to the understanding of product market integration in developing countries and the importance of different instruments of regional integration.

The second main contribution of the thesis is to provide empirical evidence on whether the key stylised facts in the burgeoning theoretical and empirical literature on product price integration, border effects and product market integration hold in the SADC context. This helps to align theoretical predictions on the relationship between product market integration and traditional trade costs, national borders, geography, and regional trade and monetary agreements with the reality in developing countries in Southern Africa.

A third contribution is to provide insight into a number of different dimensions of border effects on product market integration. To date, the bulk of the empirical research in this area has maintained a narrow geographic focus, concentrating on industrialised countries. The multi-country nature of the price data used in this thesis means that it is possible to estimate and compare the impact of specific national borders on product market integration in the developing SADC region, as well as to examine whether these effects are magnified by crossing multiple borders (a so-called incremental border effect).

The results presented in this thesis also provide insight into product heterogeneity in border effects. To date, the price-based literature on border effects has largely assumed homogeneity in border effects regardless of the type of product or the extent to which the products are actually traded across borders. The results of this thesis provide insight into whether national border effects in the SADC region vary according to the type of product being traded across the border (for instance, perishable versus non-perishable products or highly homogenous versus less homogenous products). Similarly, this thesis contributes to the international literature by providing new insight into the relationship between import intensity and national border effects on relative prices.

The fourth major contribution of this thesis is to advance the understanding of the relationship between tariffs and border effects. Most disaggregated price-based studies in the existing literature have focused on the pass-through effects of tariffs onto domestic prices or on the relationship between tariffs and distance-related trade costs, and the majority have been confined to an analysis of these relationships *within* countries (Nicita, 2009; Cherkaoui, 2011; De Loecker et al., 2012; Bas & Strauss-Kahn, 2013). Few studies have examined the relationship between tariffs and border-related trade costs or *international* trade costs more generally, and no studies have looked directly at the potential contribution of tariffs in generating border effects on international relative prices. This thesis makes an important contribution in extending the literature in this direction by providing insight into the contribution of preferential tariffs to the border effect between South Africa and Zambia.

In doing so, it also deals with some of the methodological concerns that affect other studies of tariffs and product market integration. First, it produces more accurate estimates by accounting for the spatial relationship between production and consumption in the SADC region. Second, it draws on product-level tariff data, whereas many other studies rely on aggregate data in the form of average tariff rates across products. The analysis in this thesis is thus able to exploit variation in tariffs at the product level. Third, by using data on prior agreed tariff phase downs, the analysis deals with the causality concerns that affect studies using aggregate tariff data. In analyses based on aggregate data it is difficult to exclude product-specific effects influencing tariffs. Fourth, the analysis in this thesis accounts for the possibility that the marginal effect of tariffs on the border effect may vary according to the intensity with which products are traded across the border.

More generally, the thesis makes several other important methodological contributions to the empirical literature. The empirical results on border effects and product market integration reported in most studies, including those on Africa, are obtained using standard empirical approaches in the literature. This thesis presents a critical evaluation of these standard empirical approaches. It highlights a range of sample selection effects that may bias estimates of trade costs inferred from standard approaches and, thereby, affect the accuracy of border effect estimates reported in the existing literature. The sensitivity of transaction cost and border effect estimates to these various sample selection biases is tested in this thesis using actual product price data for the

SADC region. In doing so, the thesis provides insight into the magnitude and direction of biases on distance and border effect estimates obtained using specifications that do not account for these sample selection effects.

Methodological extensions are proposed in this thesis that are designed to address these sample selection effects and produce more accurate estimates of border effects obtained using disaggregated product price level data. The thesis extends the application of the quantile regression approach first introduced by Borraz et al. (2012) to study the impact of cross-country border effects. Further novel extensions to the quantile regression methodology are developed in this thesis to account for product and distance sample selection biases. By applying these extensions to an analysis of border effects in the SADC region, the results presented in this thesis contribute to the empirical literature by providing more precise estimates of border effects in Africa.

The period of analysis in this thesis coincides with the significant liberalization of tariffs on intra-SADC trade under the SADC Protocol on Trade (which commenced in 2000). While there is some evidence that volumes of intra-SADC trade have grown since 2000, little is known about changes in product market integration over this period. The observed variation in trade flows may be consistent with the same degree of product market integration. From a policy perspective, the price-based analysis in this thesis allows for an investigation of whether the implementation of the SADC Protocol on Trade has coincided with greater integration of product markets in the region. In this respect, the findings in this thesis advance the understanding of the links between trade policy and product market integration in Africa.

1.4 Thesis Structure

The remainder of the thesis is structured as follows. Chapter 2 presents a descriptive analysis of product price integration within and between Botswana, Malawi, South Africa and Zambia. This chapter explores trends in market integration within and between these countries over time and across products. It also presents a more formal econometric analysis of the relationship between product price integration and distance, national borders and regional trade agreements using the standard empirical approach in the price-based literature on product market integration.

Chapter 3 presents a critical evaluation of the standard empirical approach in the literature. Four selection effects are identified in the chapter that may bias estimates of the impact of trade costs on product market integration that are obtained using standard empirical approaches in the literature. The estimation biases arising due to these sample selection effects are then demonstrated using actual product price level data for the four SADC countries. Extensions to the quantile regression approach are proposed in the chapter to deal with these selection effects in order to produce more accurate border effect estimates using disaggregated product price data.

Chapter 4 implements the extensions to the quantile regression methodology introduced in Chapter 3 in order to precisely quantify the effect of national borders on product market integration in the SADC region. Both average and individual country border effects on product price dispersion between countries in the region are estimated; and trends in the specific border effects are analysed over the 2006 to 2009 period.

Chapter 5 examines the relationship between tariffs and product market integration in the SADC region for the period between 2002 and 2009. The chapter considers the role played by preferential tariffs on Zambian imports from South Africa in contributing to the South Africa-Zambia border effect on relative prices. The sensitivity of the core empirical results to the potential influence of SACU Most Favoured Nation (MFN) tariffs on product prices in South Africa and Botswana, as well as to variation in import intensity across traded products, is also established.

Chapter 6 concludes the thesis. The policy implications of the core results presented in the thesis are discussed, and suggestions are made for future research in this area.

2. A Disaggregated Analysis of Product Price Integration in the Southern African Development Community

2.1 Introduction

Price-based studies of product market integration are rare in the international literature. The price-based research that has been undertaken in this area has focused primarily on industrialised countries (see for instance Engel et al., 2005; Crucini et al., 2008; Gopinath et al., 2010; Burstein and Jaimovich, 2012) rather than developing regions, with noticeably limited attention given to Africa. This has been primarily due to data limitations – particularly the lack of high-frequency data on narrowly defined goods – which have precluded research on product price dispersion and product market integration within and between African countries.[9]

As a result, little is known about product price integration within and between African economies. In addition, with the exception of studies by Versailles (2012) for members of the East African Community (EAC), and Brenton et al. (2014) for a selection of African countries including some members of the SADC and the EAC, there remains a shortage of evidence on the impact of mechanisms designed to enhance regional integration – such as regional trade and monetary agreements – on product market integration in Africa. This is in spite of the obvious importance of studying market integration in the African context, where transport costs are high and other market rigidities such as poor infrastructure, regulatory barriers and inefficient border controls hamper the flow of goods and may segment markets (Portugal-Perez & Wilson, 2008; Versailles, 2012).

This chapter advances the empirical literature by using highly disaggregated retail price data for districts in four SADC countries – Botswana, Malawi, South Africa and Zambia – to address the lack of price-based studies of product market integration in developing regions, and Africa in particular. The main objective of this chapter is to present an empirical analysis of the extent to which product prices are integrated within and between the SADC countries. The analysis is structured around the following specific objectives:

- To compare mean price dispersion within and between SADC countries and over time.

[9] Notable exceptions are Edwards and Rankin (2012); Versailles (2012); Aker et al. (2014); Brenton et al. (2014); and Nchake, Edwards and Rankin (2014).

- To analyse price deviations at the individual product level within and between the SADC countries.

- To assess the importance of distance, borders and membership in regional trade groupings for product price integration in the SADC region.

- To investigate the degree of price convergence over time within the sample of SADC countries.

The remainder of the chapter is structured as follows. Section 2.2 outlines the basic theoretical concepts that underpin the analysis of product price integration. This is followed in section 2.3 by a discussion of the key findings in the relevant empirical literature. Sections 2.4 and 2.5 explain the main features of the price data and the conceptual framework and empirical methodology used in the analysis, respectively. Section 2.6 presents empirical results on the extent to which product prices are integrated within and between SADC countries. Finally, section 2.7 concludes and discusses the implications of the empirical results for trade policy in the region.

2.2 Theoretical Background

In the absence of barriers to arbitrage between markets, buyers faced with the option to purchase similar goods in different markets will purchase a good from the market in which it is priced the lowest, subject to the cost of transportation to their home market (Bradford & Lawrence, 2004). Thus, arbitrage forces should ensure that prices for similar goods converge across well-integrated markets, as the absence of such convergence would entice buyers to purchase equivalent goods at a lower price in a different market (Knetter & Slaughter, 2001; Rogers & Smith, 2001; Parsley & Wei, 2002; Engel et al., 2005). In this context, market integration is expected to reduce the size of deviations from the LOP and result in countries facing similar relative prices for traded goods (Knetter & Slaughter, 2001).

In practice, however, a variety of factors lead to deviations from the LOP. Price dispersion across locations arises due to the presence of direct barriers to arbitrage or as a result of cost differences, variation in market structure and characteristics, or differences in the application and size of sales taxes.

Distance between locations is one of the most widely cited direct barriers to arbitrage in the theoretical literature (Dumas, 1992; Engel & Rogers, 1996; Crucini et al., 2005a; Bergin & Glick, 2007). Transportation costs, as well as costs associated with market discovery and network creation are all expected to rise as the distance between markets increases (Anderson et al., 2013). In addition, markets that are more geographically dispersed are likely to face less similar cost structures, which may be reflected in important differences in the relative prices of non-traded services or inputs into production as well as variation in relative productivity, leading to greater dispersion in product prices (Engel & Rogers, 1996; Redding & Venables, 2004). On the other hand, demand shocks might be correlated geographically, resulting in lower price dispersion between markets situated in close proximity (Anderson et al., 2013).

Cost differences between markets can drive a wedge between product prices in different markets. For instance, variation in factor prices across markets or firms – such as rental costs or differences in relative wages paid to labour – is likely to result in product price differentials. Heterogeneity in market structure could also generate price differentials across locations. For example, variation in demand conditions across markets might be reflected in differences in the size of good- and location-specific mark-ups over cost. Similarly, differing levels of competition in individual markets can prompt firms to apply different profit margins, price discrimination or pricing-to-market strategies, leading to dispersion in prices for otherwise identical goods.

Differences in market or consumer characteristics across locations can have similar effects. For example, variation in income levels, language and cultural differences, or heterogeneity in the density of ethnic networks across markets may generate price dispersion (Aker et al., 2014; Anderson et al., 2014).

For a variety of reasons, these factors are expected to be magnified in the case of markets located in different countries. For instance, additional transaction costs generated by the presence of political boundaries can hinder arbitrage and drive a wedge between prices across countries. These transaction costs take the form of direct costs stemming from barriers to trade (such as tariffs or quotas), NTBs (such as bureaucratic red tape) and other trade restrictions (Rogoff, 1996; Engel & Rogers, 1996; Rogers & Smith, 2001; Borraz, 2006); or non-pecuniary transaction costs such as exchange rate risk (Anderson et al., 2014). Alternatively, domestic policies that discriminate against foreign goods, either directly or inadvertently, can inhibit trade across borders. In this sense, hidden barriers such as subsidies, lax antitrust enforcement, health and safety standards and regulations, or cumbersome customs procedures could be present across countries.

There is also likely to be greater heterogeneity in tastes, language, culture or social networks across locations in different countries, which can serve to segment markets and contribute to price dispersion across countries (Rogers & Smith, 2001; Bradford & Lawrence, 2004; Borraz, 2006). Similarly, levels of heterogeneity in price mark-ups or relative productivity shocks between cities located across national borders might be higher in comparison to more homogenous locations within a country, leading to greater cross-border price dispersion (Engel & Rogers, 1996).

There may also be barriers to the movement of factors such as labour and capital across borders (Rogers & Smith, 2001; Borraz, 2006). This also implies that factor markets are likely to be more integrated within as opposed to between countries. Similarly, markets for non-traded services such as marketing are expected to be more integrated within countries than across countries separated by a border. This further contributes to greater dispersion in prices between than within countries.

Finally, nominal exchange rate variability in cases where final goods prices are sticky in local currency terms can generate movements in the good-level real exchange rate and cause cross-border prices of similar goods to fluctuate in line with the exchange rate (Engel & Rogers, 1996; Rogers & Smith, 2001; Engel & Rogers, 2004; Engel et al., 2005; Borraz, 2006). This would result in greater price dispersion across countries than within countries. Price stickiness is likely to be higher where markets are more segmented (Engel & Rogers, 1996).

2.3 Empirical Literature

Increasing focus has been placed in the empirical literature on the use of price-based metrics to measure product market integration. The existing body of empirical research in this area can be organised into two broad segments based on the nature of the data used. Early studies tended to utilise aggregate price indices as the basis for investigating questions related to market integration. More recently, the focus has shifted to studies that draw on disaggregated data on actual price levels to analyse price dispersion and product market integration.

2.3.1 Aggregate purchasing power parity studies using price indices

Much of the early focus in the empirical literature involved studies using price indices to test for aggregate purchasing power parity (PPP) in relative terms. For relative PPP to hold, which is expected in the long-run for integrated markets, the rate of growth in domestic and foreign prices (converted into the domestic currency) must be equal (Edwards & Rankin, 2012). In studies based on price indices, volatility in aggregate relative prices (measured as the standard deviation of the log of aggregate relative prices between two markets or locations) is most commonly used to measure deviations from PPP. Greater volatility in aggregate relative prices across locations is indicative of markets that are not well integrated.

A number of these studies find that there is significant variation in product prices around their long-run PPP means, with evidence of large and volatile short-run deviations from PPP (Frenkel, 1981; Krugman, 1987; Wei & Parsley, 1995; Rogoff, 1996; Asplund & Friberg, 2001). On balance, the price-based evidence stemming from these studies suggests that consumer goods prices remain dispersed internationally and have not converged to the extent expected given the level of globalization in the world economy (Rogoff, 1996; Rogers & Smith, 2001). Furthermore, the available evidence suggests that where convergence in prices has occurred, it has generally happened at a slow pace.[10] In Africa, the evidence from aggregate PPP studies is mixed, but does provide some support for increased price integration within Sub-Saharan Africa (SSA) (Holmes, 2000; Nagayasu, 2002; Bahmani-Oskooee & Gelan, 2006; Chang et al., 2006).

In seeking to explain the observed volatility and persistence of PPP deviations, a large body of empirical studies have utilised price indices to estimate and explain cross-border price differentials. Most studies employ CPI data to estimate the following core function in order to explain volatility in relative prices between locations:

$$V(\Delta P_{j,k}) = \alpha + \beta D_{j,k} + \delta B_{j,k} + \theta X_{j,k} + \sum_{m=1}^{n} \gamma_m L_m + \varepsilon_{j,k} \quad (1)$$

where $V(\Delta P_{j,k})$ is the standard deviation of the change in the log of the relative price between locations *j* and *k*; $D_{j,k}$ is the log of the distance between the two locations; $B_{j,k}$ is a dummy variable equal to one if the two locations are separated by a political border; $X_{j,k}$ is a vector of additional controls; and $\varepsilon_{j,k}$ is the regression error. The standard equation typically includes a dummy variable for each location in the sample, L_m; thereby allowing the standard deviation of prices to vary from

[10] Rogoff (1996) notes that deviations from PPP tend to dampen out at a rate of 15% per year.

location to location. In essence, studies using the core specification in equation (1) examine the effect on price volatility of transaction and trade costs (mostly proxied in the literature by distance, $D_{j,k}$), border costs (associated with trading across a defined geographical boundary), and other factors (such as nominal exchange rate variability, language and relative wages).

While these studies typically find that distance between locations affects the variability of prices, they also tend to find that distance alone cannot account for the observed price volatility across markets (see, for instance, Engel & Rogers, 1996; Rogers & Smith, 2001). Beyond the role of distance, a central finding in many studies is that international price dispersion significantly exceeds intranational dispersion in prices. Focusing primarily on industrialised countries, this has been demonstrated empirically through greater variability in price indices between locations in different countries in comparison to across locations in the same country, even after accounting for the distance between markets and other location-specific factors (Engel, 1993; Rogers & Jenkins, 1995; Engel & Rogers, 1996; Rogers & Smith, 2001). The 'unexplained' component of the cross-border price volatility observed in these studies has been labelled the 'border effect'.

In a seminal paper, Engel and Rogers (1996) use CPI data for 23 cities in the United States and Canada across 14 categories of consumer goods and attempt to explain the volatility of relative prices across United States and Canadian cities. They show empirically that both distance and international borders contribute to volatility in prices across cities, but find that the effect of the international border is larger. The latter is reflected in markedly higher variation in the prices of similar goods across cities in the two countries compared to cities located an equal distance apart in the same country. To quantify the "width" of the United States-Canada border, the authors calculate a distance-equivalent effect of the border. Their estimates suggest that two cities located in Canada and the United States respectively would need to be 75,000 miles apart to generate the observed level of price volatility by distance alone.

A number of studies have applied the basic approach employed in Engel and Rogers (1996) to different geographical regions, including Europe, Central Asia, and broader coverage of the North American Free Trade Agreement (NAFTA) countries (Rogers & Smith, 2001; Engel & Rogers, 2001; Parsley & Wei, 2001; Grafe et al., 2008). The vast majority of these studies focus on industrialised countries, and their results tend to support Engel and Rogers' (1996) finding of very large border effects, and smaller, but significant, distance effects on the volatility of relative prices.

Several studies using price indices have expanded the range of control variables included in $X_{j,k}$ in equation (1) in an attempt to better isolate the effect of the border. A finding common to many studies is that accounting for nominal exchange rate variability significantly reduces the estimated border effect (Engel & Rodgers, 1996 and 2001; Parsley & Wei, 2001; Borraz, 2006; Grafe et al., 2008). In addition, Parsley and Wei (2001) show that variation in unit shipping costs is responsible for a notable portion of cross-country volatility in relative prices. Notably, however, both Parsley and Wei (2001) and Rogers and Smith (2001) find that the border effect is largely unrelated to variability in relative wages between the United States and Japan and across the NAFTA countries, respectively.

Shortcomings of studies that use price indices

The initial use of aggregate price indices in the empirical literature was primarily due to a lack of disaggregated data on actual product prices. Importantly, however, studies that rely on price indices to analyse price dispersion suffer from a number of shortcomings. At a fundamental level, price indices can only be used to compare rates of inflation across locations and not to examine differences in price levels (Engel et al., 2005).[11] In addition, price indices are typically calculated using expenditure weights from household surveys, meaning that even if the absolute LOP holds, in the presence of variation in these weights the aggregate LOP will not hold. Additionally, when price indices are used, it is not possible to conclude whether observed changes in deviations in relative prices signify price convergence or divergence across countries unless PPP holds in the base year (Rogoff, 1996; Knetter & Slaughter, 2001; Edwards & Rankin, 2012).

For cross-country analyses, the use of price indices to draw inferences about the impact of distance and borders on relative prices is particularly problematic. Given that CPI data is typically based on sub-indices of fairly broad categories of goods, which may not be standardized across countries, evidence of dispersion in relative prices between countries based on price indices may reflect variation in the product and quality composition of the indices, rather than actual price differences for common products (Burstein & Jaimovich, 2012). These compositional differences may bias estimates of distance- and border-related trade costs because, in the presence of variation in the composition of goods and their underlying weights in price indices across locations, "unless all individual prices within the index move together, price indices will appear to deviate across space

[11] In studies based on price indices, volatility in relative prices (measured as the standard deviation of the log of relative prices between two markets or locations) is most commonly used to measure price dispersion. Greater volatility in relative prices across locations is indicative of markets that are not well integrated.

and borders simply due to the fact that the underlying weights and goods are different." (Broda and Weinstein, 2008: 3-4)

Additionally, the use of price indices can induce aggregation bias and overstate cross-country dispersion in relative prices by collapsing within-country volatility in relative prices and preserving the variation arising from cross-country differences (Evans, 2001; Broda & Weinstein, 2008; Chahrour & Stevens, 2012). In these instances, the use of aggregate price indices rather than disaggregated goods prices can lead to large and less precise estimates of border effects (Engel et al., 2005; Imbs et al., 2005; Grafe et al., 2008; Versailles, 2012; Brenton et al., 2014; Aker et al., 2014). Alternatively, by aggregating microeconomic price data to form an index, it is possible that positive and negative price deviations for different goods included in the index will cancel each other out, thereby averaging out potential instances of cross-sectional variation (Wolszczak-Derlacz, 2008; Schwartz, 2012). In these cases, estimates of the impact of trade costs on price differentials will be underestimated (Crucini et al., 2005a; Inanc & Zachariadis, 2012).

Another problem, which relates to both the composition and aggregation of price indices, is that most price indices are composed of both traded and non-traded goods. By aggregating the prices of these goods together into a single index, it is only possible to estimate a single border effect for both broad categories, even though the effect of political boundaries on price variability may differ considerably across traded and non-traded goods. Furthermore, the price of non-traded inputs is itself a function of distance to markets and from input suppliers, meaning that the inclusion of non-traded goods prices in the dependent variable as well as distance as an explanatory variable may bias the estimated border effect coefficient.

2.3.2 Empirical evidence on the LOP across markets within and between countries based on disaggregated price level data

In contrast to price indices, the use of disaggregated price level data in studies of price dispersion confers a number of advantages. Goldberg and Knetter (1997) stress that it is necessary to use price level data to properly understand LOP deviations. Fundamentally, the use of disaggregated data on actual product prices allows for a comparison of differences in the price *levels* of homogenous goods across locations. This makes it possible to measure price dispersion in terms of absolute deviations from the LOP. It also means that it is possible to account for potential heterogeneity across products in empirical analyses of price dispersion.

Recognising the limitations of price indices, there has been a clear shift in the empirical literature towards studies that use microeconomic data on actual product prices to analyse dispersion in relative prices. Studies covering price levels for a large number of countries have shown that global price dispersion has been uneven over time (Knetter & Slaughter, 2001; Bergin & Glick, 2007). For instance, in analysing price dispersion across a sample of 70 countries Bergin and Glick (2007) find that global price dispersion followed a U-shaped pattern over the period between 1990 and 2005.

Knetter and Slaughter (2001) observe that dispersion in relative prices across markets may differ based on whether markets are located in developed or developing economies. Specifically, the authors find that relative prices tend to be more similar across markets in developed economies, while there is evidence of both convergence and divergence in relative prices across their sample of developing economies. A number of studies have also found evidence of much greater volatility in product-level real exchange rates when compared to volatility in nominal exchange rates, which suggests that markets are segmented internationally (Broda & Weinstein, 2008; Burstein & Jaimovich, 2012; Gopinath et al., 2010).

There is widespread evidence documenting failures in the LOP in studies that use disaggregated product price data (see for instance Isard, 1977; Richardson, 1978; Giovanni, 1988; Froot et al., 1995; Parsley & Wei, 2002; Crucini et al., 2005b; Engel et al., 2005; Bergin & Glick, 2007; Crucini & Telmer, 2012; Cavallo et al., 2013). This is reflected in findings of wide dispersion in product prices within and between countries. These results are consistent across large cross-country studies (Crucini et al., 2005a; Anderson et al., 2014); studies using online prices (Cavallo et al., 2013); and studies focused on specific regions or groups of countries (for instance, Bradford & Lawrence (2004) for Organization for Economic Co-operation and Development (OECD) countries, and Moon (2013) for ten East Asian countries).

Price dispersion also seems to be present in the case of ex-factory prices. Using producer price data for goods across OECD countries, Bradford and Lawrence (2004) find considerable differences in producer prices across national markets, even in the presence of low tariffs. In adjacent countries in Europe and North America, the authors find that producer prices for comparable goods tend to differ by as much as 20%; while in the case of countries located on different continents, the price differential can range between 30% and 50%.

Some studies have noted that the extent of price dispersion across markets may depend on the nature of the product in question, suggesting that it is important to account for product heterogeneity. For example, the production composition of goods and their tradability may influence price dispersion. Crucini et al. (2005b) demonstrate empirically that the size of LOP deviations is larger for less tradable goods and for goods that use more non-tradable inputs in production. The authors find that price dispersion is around 10% higher among non-tradable goods across their sample of 122 cities in 79 countries. They also find that cross-country dispersion in prices is higher for services and for goods such as alcohol and tobacco which are typically subjected to additional taxes.

A key finding in the existing literature is that dispersion in product prices is greater between markets located in different countries than across markets *within* the same country. This has been demonstrated in several studies focused on markets in the United States and Canada (Gopinath et al., 2010; Burstein & Jaimovich, 2012). Other studies using much larger samples of countries typically arrive at the same conclusion. For instance, using product price data covering 142 cities, Anderson et al. (2012) show that mean price dispersion across city pairs located in the same country is notably smaller (at roughly 0.30) compared to mean absolute price deviations (0.56) between city pairs located in different countries. Similarly, using a sample spanning 123 cities in 79 countries, Crucini and Yilmazkuday (2014) show that price dispersion for traded goods within the United States (0.29) was much lower than the world average involving city pairs separated by an international border (0.68).

2.3.3 Studies that use disaggregated data on actual price levels to estimate trade costs and identify border effects

Given the overwhelming evidence in the literature of larger price deviations between countries compared to across locations within countries, a growing number of studies using disaggregated price data have sought to determine the sources of cross-country dispersion in price levels. Much of the focus in this regard has been on the role played by trade costs in driving a wedge between product prices in different markets.

There is considerable variation in these studies in terms of the nature of the disaggregated price data employed to measure price differentials and their determinants. This applies both in terms of the frequency of price data and the product coverage of the data. The former spans studies using city or district level average retail prices observed either monthly (Horváth et al., 2008; Versailles, 2012; Robinson, 2013; Aker et al., 2014; Brenton et al., 2014), quarterly (Broda & Weinstein, 2008; Schwarz, 2012), bi-annually (Wolszczak-Derlacz, 2008) or annually (Anderson & Smith, 2004; Crucini et al., 2005a; Engel et al., 2005; Bergin & Glick, 2007; Morshed, 2007; Morshed, 2011; Inanc & Zachariadis, 2012; Crucini & Yilmazkuday, 2014); supermarket prices (Mathä, 2003; Gopinath et al., 2011; Borraz et al., 2012); prices from a single multinational retailer (Landry, 2013); and online prices (Maier, 2010).

There is also variation in the literature with respect to product coverage. Some studies are confined to cross-border comparisons of the prices of a very small number of traded agricultural or food products (Morshed, 2011; Brenton et al., 2014; Robinson, 2013; Aker et al., 2014). In contrast, other studies employ datasets of prices of large numbers of traded goods (Anderson & Smith, 2004; Crucini et al., 2005a; Engel et al., 2005; Bergin & Glick, 2007; Wolszczak-Derlacz, 2008; Gopinath et al., 2011 Inanc & Zachariadis, 2012; Schwarz, 2012). Across these studies, there is variation in the extent of homogeneity in the products compared between locations.

Studies on the sources of cross-border dispersion in relative prices have covered different geographical regions, although the bulk of research in this area has maintained a narrow geographic focus, concentrating on industrialised countries in North America (Engel et al., 2005; Broda & Weinstein, 2008) and Europe (Mathä, 2003; Horváth et al., 2008; Wolszczak-Derlacz, 2008) rather than developing regions. That said, a few studies have employed large cross-country datasets covering countries spread across different continents (Parsley & Wei, 2002; Anderson & Smith, 2004; Crucini et al., 2005a; Bergin & Glick, 2007). While some of these studies include developing regions in their samples, there is a shortage of studies that focus exclusively on developing countries.

Most studies find that traditional trade costs (typically proxied by distance) contribute to dispersion in product-level retail prices across markets (Bradford & Lawrence, 2004; Edwards & Rankin, 2012; Zachariadis & Inanc, 2012; Anderson et al., 2013; Kano et al., 2013; Crucini & Yilmazkuday, 2014). However, most studies report relatively small distance elasticities (Mathä, 2003; Anderson & Smith, 2004; Crucini et al., 2005a; Engel et al., 2005; Bergin & Glick, 2007; Broda & Weinstein, 2008; Wolszczak-Derlacz, 2008). For instance, using data on the actual prices of nearly 150 products across 15 capital cities in the European Union (EU), Wolszczak-Derlacz (2008) reports that doubling the distance between city pairs raises the price difference by just 4.3% to 4.6%.

In a relatively recent development in the literature, the results documenting an economically small distance effect have been questioned. For instance, some studies argue – and demonstrate – that by not properly accounting for the spatial relationship between production and consumption these estimates are likely to be biased downwards due to misspecification bias arising from the exclusion of locations' distance from their production sites or the exclusion of production costs at the source of production (Anderson et al., 2013). In these cases, the failure to properly account for production and consumption locations can lead to underestimation of the role that transportation costs play as a determinant of price dispersion and LOP deviations across markets (Anderson et al., 2013; Kano et al., 2013).

In addition to isolating the effect of distance-related trade costs, the use of price level data makes it possible to analyse absolute deviations from the LOP and estimate the portion of the underlying difference in price levels across locations that can be attributed to the effect of a border (Anderson & Smith, 2004).[12] In this way, the contribution of the border to failures of the LOP can be analysed directly.

Many empirical studies using disaggregated price data have found that national borders contribute to price dispersion (Bradford & Lawrence, 2004; Crucini et al., 2005a; Bergin & Glick, 2007; Parsley & Wei, 2007; Crucini & Shintani, 2008; Crucini et al., 2008; Gopinath et al., 2010; Aker et al., 2014; Crucini & Yilmazkuday, 2014). In fact, some studies show that the role of borders is relatively more important than distance in explaining price deviations across locations (see, for instance, Anderson & Smith, 2004; and Horváth et al., 2008). It is generally accepted that borders can impose additional costs or barriers to trade that generate cross-border product price differentials and may segment markets.

There is, however, considerable variation in the magnitude of the border effects reported in the literature. Focusing on the United States and Canada, Engel et al. (2005) and Broda and Weinstein (2008) estimate that the international border is responsible for generating a 7%-7.3% wedge in cross-country prices. Mathä (2003) reports a smaller border effect on absolute price differences when looking at European countries. He finds that crossing a border raises the absolute percentage price deviation between Luxembourg and its surrounding regions by 4.2 percentage points, on average.

[12] With price indices, it is only possible to analyse deviations from the LOP in the short run (Versailles, 2012).

Some cross-country studies based on bigger samples of countries report larger border effects on international price differentials. For example, Anderson and Smith (2004) examine the impact of borders across a sample of 31 countries including several in North America, Europe and the Middle East, as well as China and Australia, and find individual border effects that range in magnitude between 18 and 33 percentage points. Disaggregating the impact of borders by region, however, they report smaller overall border effects within North America and the EU.

The border effects reported above represent measures of systematic differences in prices that are 'unexplained' by other observable factors. Even after controlling for a variety of different factors,[13] the majority of studies in the literature still find a positive border effect on dispersion in international price levels.

2.3.4 Price dispersion, border effects and product market integration in Africa

To date, very little research has investigated product market integration in Africa using actual price level data. Some notable exceptions are studies by Versailles (2012) for East Africa, Brenton et al. (2014) for Central and East Africa, Nchake (2013) for the SACU region, and Aker et al. (2014) looking at West Africa. Aker et al. (2014) use disaggregated monthly price data for two commodities (millet and cowpea) and find increased price dispersion for both commodities between markets in Niger and Nigeria.[14] Employing a regression discontinuity design, they estimate a border effect on price dispersion that varies in magnitude between 17% and 26% in the case of millet and is marginally higher for cowpeas.

For East Africa, Versailles (2012) tests the LOP across countries using disaggregated monthly product price level data for 24 goods in 39 cities in four EAC Member States: Burundi, Kenya, Rwanda and Uganda. His estimate of average price dispersion across city pairs within the same country (0.243) is notably lower than the equivalent estimate of between-country price deviations (0.443) for the 2004 to 2008 period. The author estimates that country borders in the region cause prices to deviate from the LOP benchmark by more than 13%.

---

[13] These factors include nominal exchange rate volatility (Wolszczak-Derlacz, 2008; Brenton et al., 2014); relative wages and other factor prices (Anderson & Smith, 2004; Anderson et al., 2010); non-tariff barriers (Versailles, 2012); differences in retail sales taxes (Engel et al., 2005); dispersion in relative incomes (Crucini et al., 2005a); variation in institutional quality (Schwartz, 2012); and differences in population size (Engel et al., 2005; Inanc & Zachariadis, 2012), language (Wolszczak-Derlacz, 2008) or ethnic composition (Robinson, 2013; Aker et al., 2014) across locations.

[14] They also find that ethnicity has important effects on price dispersion – with ethnic differences generating internal barriers to trade within Niger; and common ethnicities facilitating market integration in the case of cross-border markets between Niger and Nigeria.

Using a sample of only three agricultural commodities, Brenton et al. (2014) report mean within- and between-country price differentials of 0.20 and 0.33, respectively, across town pairs in 13 Central and East African countries. They find a lower border effect in Central and Eastern Africa compared to Versailles (2012), albeit based on a much smaller sample of products. Specifically, after accounting for the distance between towns and road quality, they estimate that crossing a national border causes the price differential between towns to increase by 7.7% on average.

Edwards and Rankin (2012) assess product market integration in Africa using disaggregated retail prices for more than 200 products across 13 African cities in 14 countries from 1990 to 2008. Their sample only includes annual price observations and covers a relatively dispersed set of countries, while not including within-country variation in prices. They find evidence of increased product market integration on the continent, but note that much of the movement towards greater integration was concentrated in North Africa during the early 1990s.

Focusing on Southern Africa, Nchake (2013) uses highly disaggregated monthly retail price data at the district/city level in three SACU countries: South Africa, Botswana and Lesotho. Her estimates of mean price differences between the SACU countries (0.43) are larger than those within the same country (0.225). She finds a larger border effect on relative prices in the case of the South Africa-Botswana border – 23% over the 2006 to 2008 period, compared to 14.6% for the South Africa-Lesotho border. She also finds that the border effect declined between 2004 and 2006 and 2006 and 2008, suggesting that markets in the SACU region have become more integrated over time.

Importantly, the existing studies on Africa are either limited in their product coverage (their samples are mostly dominated by agricultural or food products) and/or country and regional focus. In terms of the latter, with the exception of Nchake (2013), no studies focus on product price integration and border effects in the Southern African region. Moreover, the study by Nchake (2013) is limited to three SACU countries (Botswana, Lesotho and South Africa), meaning that a broader study of product price integration and border effects in the SADC region has yet to be undertaken. This chapter extends the literature by using highly disaggregated retail price data to examine product price integration within and between SADC countries. In doing so, the chapter provides empirical evidence on whether the key stylised facts in the burgeoning literature on price dispersion and market integration hold in the SADC context.

2.4. Data Description

To measure dispersion in product prices across the SADC region, this chapter utilises highly disaggregated retail price data collected at the district level in four SADC countries: Botswana, Malawi, South Africa and Zambia. The raw data is constructed from retail level data used in the computation of each country's CPI; and provides monthly observations for prices of a range of narrowly defined products, with the prices reported as average prices in individual districts. Importantly, for each country, the raw prices include value added tax (VAT). Consequently, to aid comparability across countries, the prices used in the empirical analysis are recalculated net of VAT.[15]

The fact that the disaggregated data exploited in this paper represents retail prices (rather than producer or wholesale prices) is advantageous. According to Hillberry and Hummels (2003), analyses of price dispersion that rely on business-to-business data may underestimate price differences within countries. In contrast, retail price data allows for tighter predictions on absolute and relative price movements in the case of consumer products (Edwards & Rankin, 2012).

Furthermore, the organisation of the data along narrowly defined product descriptions means that it is possible to compare prices for similar products across markets either within or between countries and, thereby, avoid the possible aggregation bias associated with price indices (Ceglowski, 2003). To obtain a dataset of comparable products across all four countries that are as close to homogenous as possible, a comprehensive system of new product and unit codes was developed, which was then used to map unique product and unit codes to each of the original product and unit combinations in the individual country raw price datasets. The mapped datasets were then merged, and the resulting combined cross-country dataset was collapsed to include only observations for common products and units across all four countries.

The final dataset constructed in this fashion includes monthly price observations for 24 narrowly defined products spanning the period from January 2001 to December 2011. Table 1 outlines the time, product and district coverage of the sample by country. There is significant heterogeneity in the time span of the data across countries. The common period in which price data is available for

[15] There is some variation in VAT rates across and within countries during the sample period. Across countries, VAT rates ranged from 10% in Botswana and 14% in South Africa to highs of 17.5% in Zambia, 18% in Uganda and 20% in Malawi and Tanzania in particular sub-periods.

all four SADC countries extends from September 2006 to December 2009. Consequently, all empirical analyses that follow in this chapter are confined to the 2006 to 2009 period.

The sampled products cover a diverse range of categories: foods; clothing and textiles; machinery, equipment and electronics; and other products. All 24 products are tradable, and each product is traded to some degree between the SADC countries; suggesting that the use of the sampled products is appropriate to gauge the level of product market integration in the region. Importantly, each of the products is broadly comparable across districts both within and between countries: the products share identical or very similar basic descriptions and the same unit of measurement in all districts (see Table A.1 in Appendix A). Moreover, the process followed by the relevant statistical agencies in collecting the microeconomic price data that underlies the CPI in each country is similar, with the standard international Classification of Individual Consumption by Purpose aiding the comparison of prices for well-defined products between countries.

Table 1: Time, district and product coverage, by country

Country	Timeframe	Number of Districts	Original Number of Products
Botswana	September 2006 – December 2009	46	384
Malawi	January 2002 – December 2011	4	162
South Africa	December 2001 – December 2010	25	951
Zambia	January 2001 – November 2011	39	354

Notes: In the case of certain countries, the product coverage in the raw datasets varies by district.

Nevertheless, while the sampled products are highly similar across countries, they may not be perfectly homogenous. In many cases, the original product descriptions provided in the raw country price data are devoid of detail on brand or other product-specific characteristics. As a result, potentially important product-level differences in quality or brand name may remain across districts, and could contribute to any observed dispersion in product prices across districts. This is, however, likely to be less of a concern within countries.

To account for the influence that within-product heterogeneity across districts in different countries may have on the results, Rauch's (1999) highly disaggregated product classification scheme is used to distinguish between highly homogenous and less homogenous products within the sample (see Table A.1 in Appendix A).[16] Using this classification, the robustness of the

[16] Rauch (1999) distinguishes between homogenous and differentiated products on the basis of the degree of product substitutability. Products that fall within the homogenous group are generally considered to be substitutable (even if

regression results presented later is tested using the sub-set of only highly homogenous products and controls for district-specific characteristics. The highly homogenous group consists of agricultural products (fresh fruits and vegetables) and paraffin, all of which are unlikely to differ much across locations, and are generally highly substitutable. In comparison, there is likely to be greater scope for prices to diverge on the basis of quality or other characteristics in the case of the less homogenous products.

Aside from the broad product coverage, the wide district coverage in the sample can be exploited to consider variation in prices within countries. However, there is heterogeneity in the number of districts for which price data is available within each country. The district coverage ranges from 46 districts in Botswana, 39 in Zambia and 25 in South Africa to just 4 districts in Malawi.[17]

The membership of the four countries in multilateral trade arrangements spans three different regional trade groupings: SADC, SACU and COMESA. At least two of the four countries are members of each of these regional trade groupings. Specifically, in addition to common membership within SADC, Botswana and South Africa share membership in SACU; and Malawi and Zambia are both members of COMESA. This makes it possible to explore the influence of membership in different regional trade and monetary arrangements on product price integration in the SADC region.

Before proceeding with the empirical analysis, in order to ready the raw price dataset a systematic procedure to remove outliers is implemented. Following Borraz et al. (2012), for each product, monthly dollar price observations either three times greater than the median monthly product price or less than one third of the median monthly product price are deleted. The deleted monthly price observations account for 5.6% of the original raw monthly price database.

in some cases they have certain unique attributes); while differentiated products are likely to have "many characteristics that vary across suppliers and may even be specifically tailored to the end-user's needs" (Besedeš and Prusa, 2006: 343). Rauch argues that the possession of a reference price is what distinguishes homogenous from differentiated products. He divides homogenous products further into those for which the reference prices are quoted on organised exchanges and those whose reference prices are quoted only in trade publications. This chapter does not distinguish between these two types of homogenous products, instead considering them simply as part of a highly homogenous product category (with those corresponding to Rauch's differentiated classification referred to as less homogenous products).

[17] For Malawi, separate price series for high, medium and low income areas in each of the four districts are available in the raw data. To arrive at a single price for each district, the average price by product and month across the three income series was used.

The remaining monthly price observations for each district in the sample are then used to compute an annual average price for each product i in district j in year t. The primary motivation for collapsing the monthly price observations to annual averages is that it is important to measure price integration across locations using stable long-run product prices so that any observed dispersion in prices is not simply a product of month-to-month fluctuations in a particular location. By smoothing out these fluctuations, the annual averages are more likely to approximate the stable long-run product prices in each district. Similarly, using the annual averages dampens the influence of seasonal factors or temporary promotions on product prices in particular districts.

2.5 Conceptual Framework and Empirical Methodology

The absolute price difference reflects the relative cost of an identical product between two locations at a particular point in time. Empirically, the relative price of product i between districts j and k at time t can be calculated as the difference in the log prices of that product between the two districts:

$$RP_{i,jk,t} = p_{i,j,t} - p_{i,k,t} \tag{2}$$

If the absolute version of the LOP holds, the value of $RP_{i,jk,t}$ in equation (2) should be equal to zero. A deviation from the LOP will be reflected in an absolute value of $RP_{i,jk,t}$ that differs from zero.

The simple relative price measure in equation (2) forms the conceptual basis for the subsequent analysis of price dispersion within and between countries. Individual *within*-country relative prices are calculated as the difference between the price of product i in district j and the price of that product in another district k located within the same country:

$$RPW_{i,jk,c,t} = p_{i,j,c,t} - p_{i,k,c,t} \tag{3}$$

where $p_{i,j,c,t}$ is the log price of product i in district j in country c at time t; and $p_{i,k,c,t}$ is the equivalent product price in district k in country c.

In turn, individual *between*-country relative prices are calculated as the difference between the price of product i in district j in country c and the price of that product in another district k located in a different country d (where $c \neq d$):

$$RPB_{i,jk,t} = p_{i,j,c,t} - p_{i,k,d,t} \tag{4}$$

Using these basic measures of product price deviations across districts (either within or between countries), a number of different aspects of price dispersion and product market integration in the SADC are investigated in the following section.

2.6 Empirical Analysis and Results

To provide a preliminary perspective on country level variation in prices for individual products, Table A.2 in Appendix A compares average prices (converted into United States dollars (US$)) in each country in 2008 for all products included in the matched sample.[18] In the case of certain products – in both the highly homogenous and less homogenous categories – there is significant variation in common currency prices across countries. For example, the average price of a refrigerator in 2008 ranged from US$471.59 in Malawi to US$290.60 in Zambia. In contrast, for some products, variation in prices between countries is low. For instance, the average price for 50ml of shoe polish in 2008 was US$0.67 in South Africa, US$0.78 in Zambia, US$0.79 in Botswana and US$0.81 in Malawi.

In general, however, product prices are dispersed across the four SADC countries, and the size of the dispersion varies depending on the product in question. While it is important not to draw any serious conclusions from a simple cross-sectional comparison of price levels, it is nevertheless an interesting point of departure.

2.6.1 Country level comparison of price dispersion within and between SADC countries

To provide an initial, cross-country perspective on variation in product prices within and between countries and over time, a country level comparison of average price dispersion across all products for the period from 2006 to 2009 is presented in Figure 1. To evaluate the mean price dispersion *within* each country, the absolute values of the individual within-country relative prices for all products are averaged across all bilateral district pairs *jk* within country *c* at time *t*: $MADW_{c,t} = E_{i,jk}\left[\left|RPW_{i,jk,c,t}\right|\right]$. Similarly, in order to obtain a single between-country price dispersion measure for each country, the absolute values of the individual between-country relative prices for all

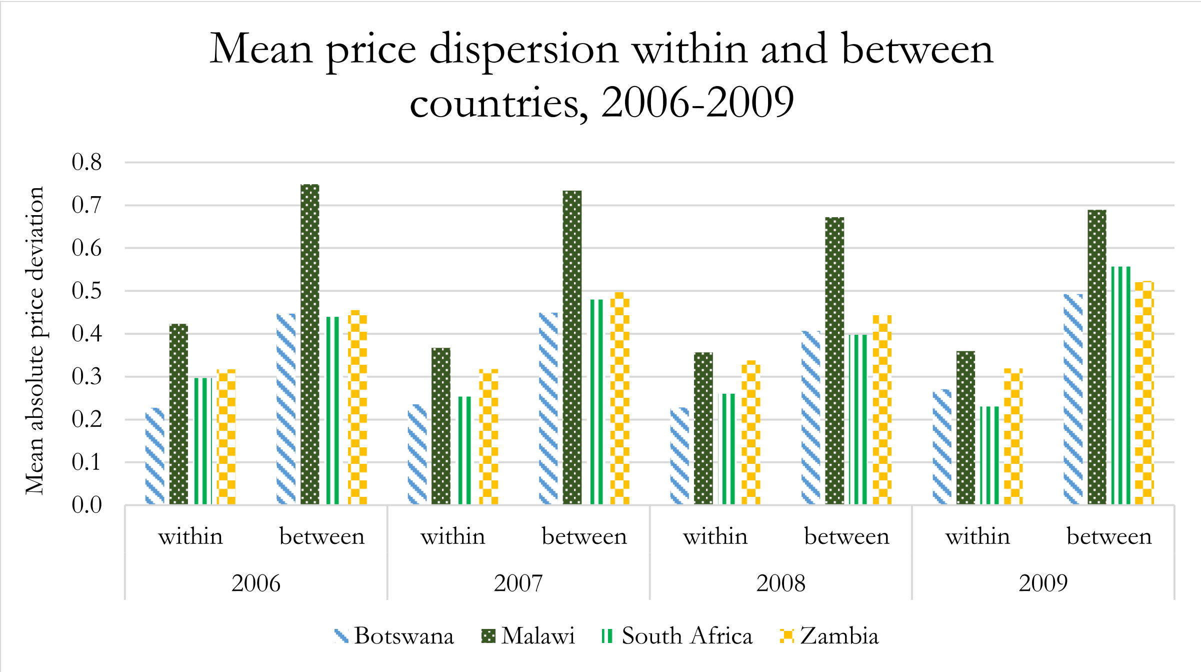
[18] The year 2008 is chosen as it provides the most recent time period for which the largest possible sample of common products across all countries is available.

products are averaged across all bilateral district pairs *jk* involving country *c* and districts in the other countries: $MADB_{c,t} = E_{i,jk}[|RPB_{i,jk,t}|]$.

Price dispersion within countries

The mean values of the within-country absolute price deviations differ across the four countries. In 2009, mean absolute price dispersion across districts within each country was lowest in South Africa (at 0.23) and highest in Malawi (at 0.36). Across all four SADC countries, product prices differed between districts within countries by 34% on average in 2009.[19]

Figure 1: Mean absolute price deviations within and between countries, 2006-2009



These values are somewhat higher than those obtained in other multi-country studies using disaggregated price level data. Also in Africa, Versailles (2012) reports a value of 0.24 for average price dispersion across city-pairs within four EAC countries, implying that prices in the region differ within countries by around 27.5%. Using monthly prices for maize, rice and sorghum, Brenton et al. (2014) find average price differences of around 22% between towns within Central and East Africa. Outside of Africa, Anderson et al. (2012) use a disaggregated dataset spanning almost 200 products and services across 142 cities internationally and find mean price dispersion over city pairs within the same country to be around 0.30.

[19] Calculated as: exp(0.29)-1.

The average values of within-country mean absolute price deviations for each of the four countries remained fairly stable between 2006 and 2009 (see Figure 1 and Table 2). This suggests that, internally, product prices within the African countries have not, on average, become more integrated over this period.

Price dispersion within versus between countries

Average dispersion in prices between the SADC countries is substantially higher than across districts *within* countries (see Table 2). This finding is consistent with the empirical literature (see recent studies by Gopinath et al., 2010; Burstein and Jaimovich, 2012; Anderson et al., 2014; Brenton et al. 2014; Crucini & Yilmazkuday, 2014). For the SADC countries in 2009, average between country price deviations across all products ranged from 0.49 in Botswana to 0.69 in Malawi. Absolute price dispersion between all countries averaged 0.53 over the 2006 to 2009 period – with the value increasing marginally from 0.52 in 2006 to 0.57 in 2009. Thus, in 2009, between-country price deviations averaged 77%.[20]

These values are in line with the general evidence in the literature of widespread deviations from the LOP between countries. While the estimated mean price deviations between SADC countries are higher than those reported in studies using similar measures of price dispersion between developed countries in North America (Engel et al., 2005; Gopinath et al., 2010) and within Europe (Wolszczak-Derlacz, 2008; Moon, 2013),[21] they are quite similar to those reported in other studies of price dispersion in developing regions (but larger than other estimates for Africa). In Africa, Versailles (2012) reports between-country price dispersion of 0.44 across 24 products in four EAC countries. Focusing on 13 African countries, Edwards and Rankin (2012) estimate absolute average log price deviations from the LOP (relative to the average log SSA price) of 0.24 between 2005 and 2008. Brenton et al. (2014) report average between-country price differentials of 0.33 across a number of East and Central African countries. For East Asia, Moon (2013) computes similar mean price differentials to those found in this chapter – ranging between 0.59

[20] Calculated as: exp(0.57)-1.
[21] For example, estimates of price dispersion between the United States and Canada range from 7% reported by Engel et al. (2005) to 22.2% found by Gopinath et al. (2010); while Wolszczak-Derlacz (2008) reports mean price dispersion (measured as the mean coefficient of variation) of around 34% across 15 capital cities in Europe.

and 0.61 – across ten countries in the region over the period from 2006 to 2009.[22] Using a much larger sample of 142 cities in countries spread across multiple continents (including developing countries), Anderson et al. (2012) estimate a mean absolute price differential between countries (over all goods and location pairs) of 0.56 (and 0.44 when restricted to OECD countries).

Table 2: Mean absolute price deviations within and between SADC countries, 2006-2009

	Mean absolute price dispersion (across all products)			
	2006	**2007**	**2008**	**2009**
Within-country district pairs	0.32	0.29	0.30	0.29
Between-country district pairs	0.52	0.54	0.48	0.57

For three of the four countries in the sample, the magnitude of mean between-country absolute price deviations has actually increased over time. Over the 2006 to 2009 period, the size of mean absolute price differences increased between Botswana (from 0.45 in 2006 to 0.49 in 2009), South Africa (from 0.44 in 2006 to 0.56 in 2009) and Zambia (from 0.46 in 2006 to 0.52 in 2009) and the other SADC countries. By comparison, mean absolute price dispersion between Malawi and the other SADC countries declined (from 0.75 in 2006 to 0.69 in 2009) over the same period, albeit from a notably higher base. The evidence of large and persistent LOP deviations between countries appears to suggest that product prices in the SADC region have not become more integrated over the sample period.

2.6.2 Comparison of price dispersion across products

The price deviation measures presented above are computed as averages across all products. While instructive in providing an aggregate, country level picture of price dispersion within and between the four SADC countries, they may mask important differences in price dispersion at the product level. For instance, in theory LOP deviations should be larger for less tradable goods and for goods that use more non-tradable inputs in production (Crucini et al., 2005b). Mindful of this, price deviation measures within and between countries for each product i are also computed. First, to compare *within*-country, product-level dispersion in prices, the absolute values of the individual within-country relative prices for product i, $\left|RPW_{i,jk,c,t}\right|$, are averaged across all bilateral district

[22] He attributes the comparatively high level of price dispersion in the East Asian region to the presence of large exchange rate volatilities, wide disparities in intra-regional incomes and a lack of emphasis on regionalisation during the sample period.

pairs *jk* that are both located in any one of the four SADC countries (denoted by *c*): $MADW_{i,t}$= $E_i[|RPW_{i,jk,c,t}|]$. This yields one observation, $MADW_{i,t}$, at time *t* for each of the 24 products. Similarly, the individual between-country relative prices for product *i* at time *t* are averaged across all cross-country bilateral district pairs *jk*: $MADB_{i,t}$= $E_i[|RPB_{i,jk,t}|]$.

The product-specific values of the within- and between-country mean absolute price dispersion measures are presented in Table 3. For the majority of products and in almost all years, price dispersion within countries is notably lower than dispersion in prices between countries. In 2009, this was the case for 23 of the 24 products (bath towel being the only exception).

Focusing on price dispersion *between* countries and looking at the variation across products, the values in Table 3 reveal considerable heterogeneity at the product level, even within the smaller sub-sets of products that are common to a particular group. In 2009, for instance, average absolute deviations from the LOP between countries ranged from 25% for shoe polish[23] and 28% for paraffin to 120% for 1kg of cabbage and 136% for a brassiere. Interestingly, the between-country price deviations for 2009 are identical, on average, for the highly homogenous and less homogenous products groups (0.53 in 2009). This is driven, in part, by the relatively high between-country price deviations for the food products in the highly homogenous group (which averaged 0.54 in 2009); and the comparatively lower between-country price deviations for the machinery, equipment and electronics (averaging 0.45 in 2009) and other products (average of 0.23 in 2009) in the less homogenous products group.

One possible explanation for the relatively high level of price dispersion for food products in the region is that, with the exception of rice, all of the food products in the highly homogenous group are perishable fresh fruits and vegetables.[24] These products are less tradable given that they are typically subject to fast deterioration and high transport costs. Furthermore, they may be subject to NTBs that inhibit arbitrage between countries.[25]

[23] Calculated as: exp(0.22)-1.

[24] In comparable work on Africa, Versailles (2012) reports larger deviations from the LOP both within and between countries for fruits and vegetables in comparison to staple foods, other food items and other products.

[25] Examples of NTBs potentially affecting fruits and vegetables are phyto-sanitary standards or problems related to the issuance of certificates of origin to traders for perishable products.

Table 3: Product-level price dispersion within and between countries, 2006-2009

	2006		2007		2008		2009	
	within	between	within	between	within	between	within	between
Highly homogenous								
Food								
bananas	0.36	0.44	0.30	0.51	0.34	0.42	0.45	0.58
cabbage	0.27	0.61	0.26	0.77	0.27	0.57	0.25	0.76
onions	0.32	0.54	0.24	0.37	0.32	0.42	0.33	0.37
oranges	0.37	0.67	0.41	0.50	0.45	0.58	0.60	0.79
pineapples	0.16	0.54	0.23	0.43	0.25	0.33	0.31	0.48
potatoes	0.29	0.43	0.27	0.50	0.31	0.37	0.30	0.55
rice	0.14	0.28	0.13	0.34	0.13	0.28	0.16	0.43
tomatoes	0.28	0.53	0.22	0.68	0.23	0.38	0.24	0.56
Other products								
paraffin	0.12	0.18	0.11	0.18	0.10	0.18	0.20	0.25
Less homogenous								
Food								
margarine	0.12	0.32	0.16	0.28	0.20	0.36	0.23	0.30
Clothing and Textiles								
bath towel	0.29	0.31	0.38	0.36	0.27	0.30	0.39	0.34
blanket	0.20	0.84	0.29	0.88	0.18	0.80	0.16	0.60
boy's shirt	0.17	0.48	0.19	0.42	0.20	0.56	0.17	0.56
brassiere	0.47	0.91	0.41	0.97	0.45	0.77	0.42	0.86
girl's dress	0.27	0.45	0.32	0.45	0.30	0.49	0.29	0.64
ladies dress	0.40	0.50	0.37	0.48	0.35	0.45	0.31	0.43
men's shirt	0.41	0.61	0.40	0.53	0.36	0.55	0.37	0.70
men's suit	0.32	0.46	0.37	0.51	0.25	0.40	0.36	0.60
men's trousers	0.35	0.62	0.31	0.58	0.33	0.49	0.39	0.83
Machinery, equipment and electronics								
colour television	0.12	0.27	0.14	0.26	0.14	0.33	0.21	0.40
electric iron	0.20	0.49	0.19	0.47	0.22	0.47	0.18	0.45
electric kettle	0.20	0.36	0.21	0.36	0.17	0.30	0.18	0.41
refrigerator	0.15	0.31	0.16	0.49	0.18	0.37	0.18	0.56
Other products								
shoe polish	0.11	0.12	0.10	0.15	0.10	0.13	0.09	0.22

There is considerable heterogeneity across products in the trends in product price integration over time. Average between-country price deviations declined from 2006 to 2009 for just seven of the 24 products in the sample and increased for the remaining 17 products. Most of the products (five out of seven) for which between-country price dispersion declined over this period are part of the less homogenous products group.

2.6.3 The importance of distance, borders and membership in regional trade groupings for product price integration

The stylized facts in the theoretical and empirical literature on price dispersion and product market integration indicate that traditional trade costs (typically proxied by distance) contribute to dispersion in product-level retail prices across locations. They also suggest that the presence of national borders generates dispersion in the prices of similar goods across locations in different countries, even after accounting for the distance between locations and other relevant factors (Crucini et al., 2005a; Versailles, 2012; Aker et al., 2014; Brenton et al., 2014). Nevertheless, even in the presence of cross-country price dispersion, markets situated in adjacent countries, and those separated by smaller distances, are likely to experience less dispersion in product prices over more geographically disparate locations. There is also a body of evidence in the empirical literature to suggest that common membership in regional trade and monetary agreements may reduce product price dispersion across locations in different countries (Parsley & Wei, 2002; Foad, 2004; Bergin & Glick, 2007; Versailles, 2012). This section presents a number of simple empirical estimates to test whether these key stylized features hold in the SADC context.

Figure 2, which compares kernel density estimates of mean absolute log price differences (pooled across all products) within and between countries, provides visual evidence of the additional dispersion in product prices generated by crossing a national border in the SADC region. There are clear differences between the within- and between-country price distributions, with the between-country densities notably larger than the mean deviations within countries. Furthermore, the separate distributions for contiguous and non-contiguous countries show that mean absolute deviations from the LOP are larger between countries that do not share a common border.

To provide an indication of the magnitude of these differences for between-country district pairs in each year, the data on the absolute values of between-country price deviations for all cross-country bilateral district pairs, $|RPB_{i,jk,t}|$, are pooled and mean absolute price deviations for specific groups of bilateral pairs are compared. The results are presented in Figure 3, which compares this measure for cross-country district pairs in either contiguous or non-contiguous countries, as well as against the average across all between-country district pairs, for the period from 2006 to 2009. On average (across all products), mean absolute price deviations for bilateral district pairs in contiguous countries are lower compared to between-country district pairs in countries that are not adjacent to each other. In 2009, the mean absolute price deviation between district pairs in non-contiguous countries was 0.66 compared to 0.49 for those in pairs of countries that share a common border.

Figure 2: Kernel density estimates of mean absolute price dispersion (across all products) within and between countries, 2006-2009

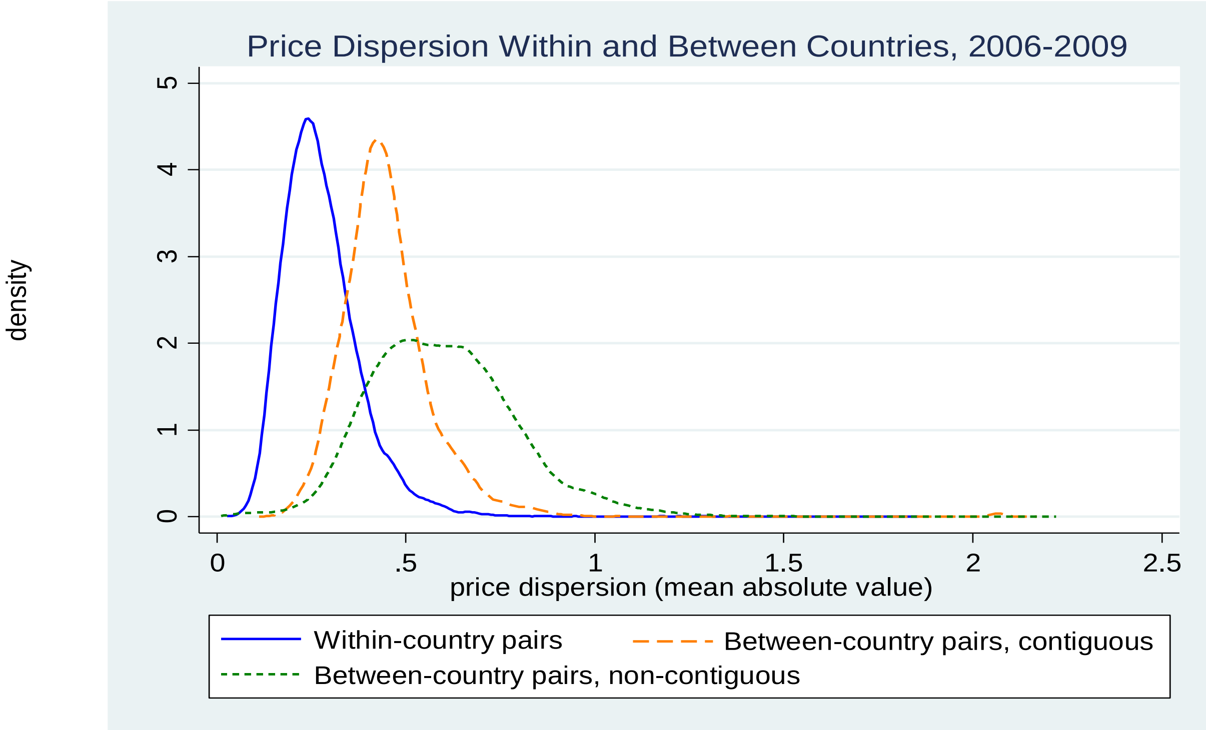
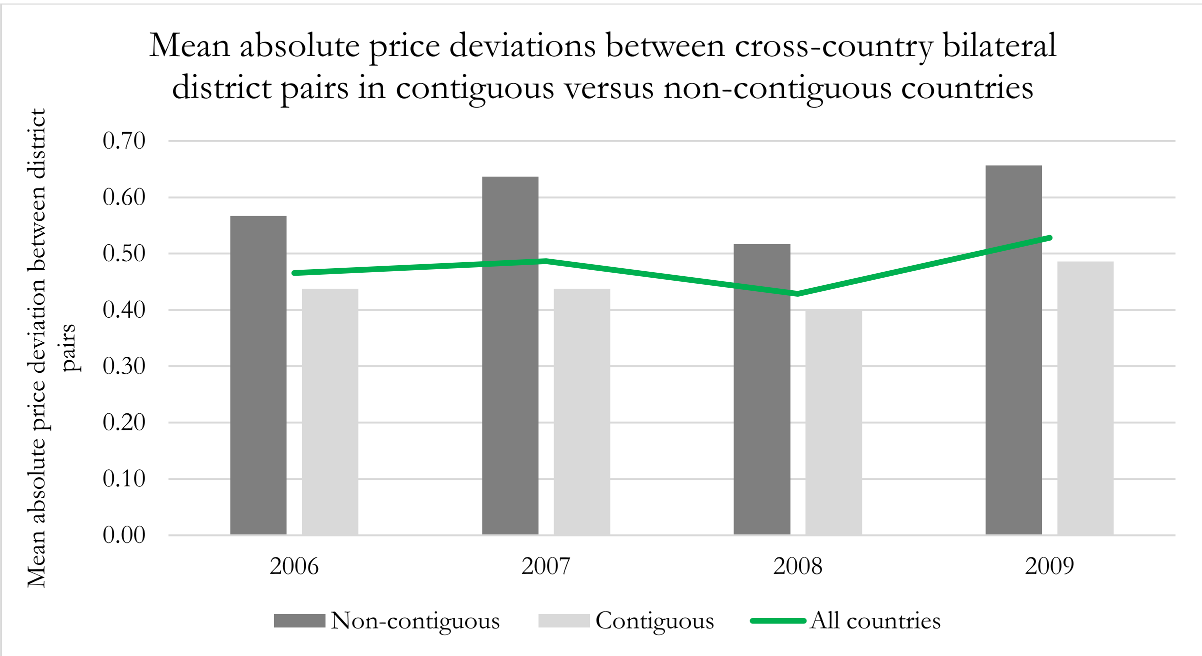


Figure 3: Comparison of mean absolute price deviations for cross-country district pairs in contiguous versus non-contiguous country pairs, 2006-2009



Turning to the effect of membership in additional regional trade and monetary agreements, Table 4 presents a comparison of the mean absolute price deviations for district pairs in the SADC countries that share common membership in SACU or COMESA against those in countries that are not both members of the respective groupings. The results indicate that absolute price dispersion (averaged over all products) is lower, on average, between cross-country district pairs located in SADC countries that share common membership in the SACU formation (Botswana and South Africa). However, the same does not hold in the case of countries that are members of both SADC and COMESA (Malawi and Zambia).

Table 4: Mean absolute price deviations for cross-country district pairs in countries with or without common membership in regional trade agreements, 2006-2009

	SACU		COMESA	
	Not both SACU	Both SACU	Not both COMESA	Both COMESA
2006	0.49	0.40	0.46	0.73
2007	0.53	0.39	0.48	0.61
2008	0.47	0.34	0.42	0.64
2009	0.54	0.49	0.53	0.61

Note: The 'Not both SACU' and 'Not both COMESA' columns are calculated as mean absolute price deviations for cross-country district pairs involving countries that do not share a common membership in these regional trade groupings.

Estimating the importance of distance, borders and membership in regional trade groupings for product price integration in the SADC region

The evidence presented thus far points to greater dispersion in product prices between countries than within countries. The role played by transaction costs of trade in impeding trade and hindering arbitrage activities between countries is well documented in the empirical literature; and costs associated with crossing national borders, in particular, have been shown to be a significant source of unexplained price disparities between countries (Engel & Rogers, 1996; Versailles, 2012; Aker et al., 2014; Brenton et al., 2014). These relationships are now tested more formally through simple econometric estimations to isolate the association between distance, national borders and common membership in regional trade and monetary agreements on product price deviations between SADC countries.

The aim in this chapter is to conduct a descriptive analysis of the relevance of these three factors for product price integration and, thus, it is not the intention in this chapter to precisely estimate the border effects. The latter would require accounting for sample selection bias and variation in mark-ups across locations (see Borraz et al. 2012; Atkin & Donaldson, 2014), which is beyond the scope of this chapter. Instead, in Chapter 3, a variety of different selection biases are identified that affect estimates of distance- and border-related trade costs using the standard approach in the price-based literature on product market integration (which is employed in this chapter). An extension of the quantile regression approach first introduced by Borraz et al. (2012) is then employed in Chapter 4 in order to obtain more precise estimates of the effect of borders on dispersion in relative prices in the SADC region.

For the purposes of this chapter, the relationships are estimated over the period from 2006 to 2009, pooling data on district-pair relative prices across all products and all district pair combinations in the sample. Selecting the absolute value of the relative price of product *i* between districts *j* and *k* at time *t* as the dependent variable, the basic regression model is specified as:

$$|RP_{i,jk,t}| = \alpha + \beta_1 \ln(dist_{jk}) + \beta_2 border_{jk} + \beta_3 SACU_{jk} + \beta_4 COMESA_{jk} + \delta_i + \gamma_t + \sum_{m=1}^{n} \gamma_m^i D_m + \varepsilon_{jk,t} \quad (5)$$

where $dist_{jk}$ is the log of the distance between districts *j* and *k*;[26] $border_{jk}$ is a dummy variable equal to one if districts *j* and *k* are separated by a national border and zero otherwise; $SACU_{jk}$ and $COMESA_{jk}$ are dummy variables equal to one if both districts in cross-country district pair *jk* are located in countries that share common membership in SACU or COMESA, and zero otherwise; δ_i and γ_t are product and time fixed effects, respectively; D_m is a dummy variable for each district included in the regression; and $\varepsilon_{jk,t}$ is the regression error term. The standard errors reported in the tables below are clustered by district pair to account for the possibility that the regression errors are not independent within the district-pair dimension (Versailles, 2012).

The initial estimation results are reported in columns (1) and (2) in Table 5. The results conform to expectations and corroborate both the earlier findings and the stylized facts in the literature. In both specifications, distance and the border dummy are highly significant. Absolute price deviations between district pairs in the SADC region increase the further apart the districts are from each other, and are higher in the case of districts separated by a national border. Specifically, a distance of 100km between districts causes prices to deviate from the LOP by 16.7%.[27] Crossing a national border increases absolute price dispersion between districts in the SADC region by more than 15% on average. Furthermore, relative to the rest of SADC, absolute price deviations between districts located in different countries are on average around 7.7% smaller when those countries

[26] Distance is measured as the shortest road distance between district pairs. These distances are calculated from Google Maps using the longitude and latitude coordinates of the main city in each district (or the mid-point in districts not defined by a particular city).
[27] Calculated as: 0.0364*ln(100).

share common membership in the SACU (Botswana and South Africa); and more than 6% larger on average when both districts in the pair are located in COMESA countries (Malawi and Zambia).

Table 5: Pooled regressions (all products), 2006-2009

	Basic regressions		**Accounting for country-heterogeneity effect**		**Accounting for country-heterogeneity effect (highly homogenous products only)**	
	(1)	(2)	(3)	(4)	(5)	(6)
log distance	0.0364***	0.0219***	0.0380***	0.0210***	0.0633***	0.0290***
	(0.00193)	(0.00182)	(0.00197)	(0.00180)	(0.00271)	(0.00217)
border	0.153***	0.191***	0.118***	0.134***	0.0844***	0.119***
	(0.00278)	(0.00328)	(0.00445)	(0.00395)	(0.00578)	(0.00472)
SACU		-0.0767***		-0.0929***		-0.187***
		(0.00326)		(0.00356)		(0.00470)
COMESA		0.0626***		0.0717***		0.119***
		(0.0143)		(0.0149)		(0.0206)
constant	0.0589***	0.165***	0.106***	0.269***	-0.112**	0.214***
	(0.0115)	(0.0200)	(0.0230)	(0.0125)	(0.0456)	(0.0158)
Observations	412,972	412,972	412,972	412,972	174,193	174,193
R-squared	0.254	0.259	0.255	0.261	0.257	0.281
Distance-equivalent of the border effect	66.9	6,133.1	22.3	590.5	3.8	60.5

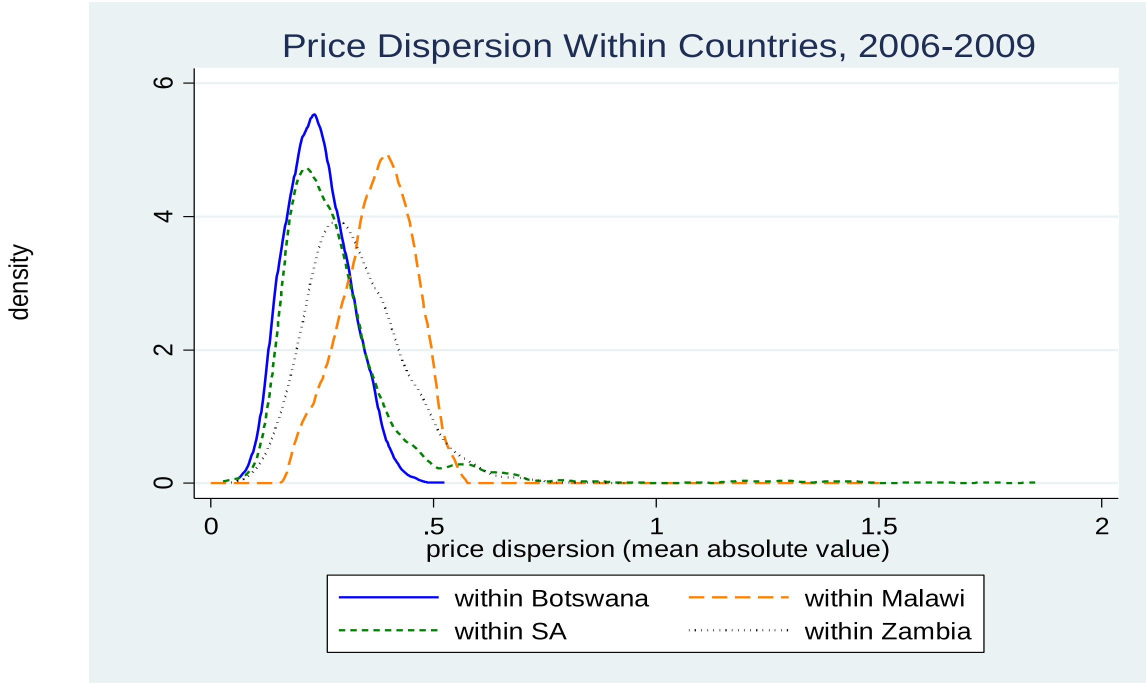
Notes: All regressions are estimated with product and district fixed effects. Standard errors are clustered by district pair and reported in parentheses. The rest of SADC dummy is the omitted category (covering Botswana-Malawi, Botswana-Zambia, Malawi-South Africa and South Africa-Zambia country pairs). In the regressions in columns (3)-(6), country-specific dummies are included for 'Botswana-Botswana', 'Malawi-Malawi' and 'South Africa-South Africa' district pairs (with the 'Zambia-Zambia' dummy omitted). Significance at the 10 percent, 5 percent and 1 percent levels is denoted by *, ** and *** respectively.

Importantly, however, the regressions in columns (1) and (2) compute the international border effect by comparing price dispersion across districts in different SADC countries against a baseline of average price variability across districts *within* the four countries. Consequently, the coefficients on the border dummy reported in columns (1) and (2) measure the average increase in relative prices when moving from the average within-country district pair in the sample to a between-country district pair. Gorodnichenko and Tesar (2009) note, however, that the average price variability within countries is not necessarily a good benchmark from which to gauge the effect of borders on international price differentials. This is because it does not account for potential differences across countries in the underlying variability in prices *within* countries.

If there is cross-country heterogeneity in the distribution of prices within countries, then the standard border effect regressions "cannot separate the border frictions from the effect of trading with a country that has a different distribution of prices." (Gorodnichenko & Tesar, 2009: 220) Consequently, in the presence of this cross-country heterogeneity estimates of the average border effect are likely to be biased *upwards* because they conflate the effect of differences in the internal distribution of prices across countries with the effect of the border; and, thereby, overstate the true effect of national borders on price dispersion.

To investigate whether this is likely to be a problem in the data, Figure 4 compares the kernel density distributions of mean absolute log price differences (across all products) within each of the four SADC countries over the 2006 to 2009 period. The kernel density estimates show that the within-country relative price distributions differ across the four SADC countries, with mean deviations from the LOP noticeably larger within Zambia and Malawi compared to within Botswana and South Africa. These differences suggest that it will be important to account for country-heterogeneity in the regression estimates.

Figure 4: Kernel density estimates of mean absolute price dispersion (across all products) within countries, 2006-2009



In order to do so, dummies are added for district pairs that are both in the same country to the regression specification in columns (3) and (4). One of the country dummies is omitted since, due to perfect collinearity, it is not possible to include a border dummy and dummies for all four SADC countries. In selecting which dummy to omit, the suggestion of Gorodnichenko and Tesar (2009) is followed and the dummy is omitted for the country with the smallest difference between within-country and between-country district-pair price dispersion (Zambia in the data). This means that the estimated border effect will be as small as possible (Versailles, 2012).

As expected, the estimated coefficient on the border dummy (and the distance-equivalent of the border effect) falls once accounting for the influence of country heterogeneity. Relative to average price dispersion within Zambia (the benchmark country), crossing a national border in the SADC region adds, on average, 11.8% to relative prices between countries. Measured in distance-equivalent terms,[28] an additional 22.3km in distance between districts in the SADC region would generate the same degree of price dispersion as that generated by crossing a national border.[29] By comparison, 100km in distance between districts causes prices to deviate from the LOP by 17.5%.

The estimated average border effect for the SADC region is higher than equivalent estimates based on actual price level data for developed regions. For example, Engel et al. (2005) and Broda and Weinstein (2008) estimate that the United States-Canada border is responsible for generating a differential of around 7% in cross-country prices. Nevertheless, the results presented here fall in the middle of the range of existing estimates of border effects in different regions in Africa. The estimated price differences associated with crossing a national border range from 7% in Central and East Africa (Brenton et al., 2014) and 13.6% among EAC countries (Versailles, 2012) to as high as 23% in the case of the South Africa-Botswana border (Nchake, 2013) and between 17% and 26% for the Niger-Nigeria border (Aker et al., 2014).

Robustness tests

One concern is that there may be unobserved heterogeneity in the products compared across districts in the sample, which would be reflected in differences in the final retail prices of products

[28] In the empirical literature, the distance-equivalent of the border effect is typically calculated as: $\exp(\beta_2/\beta_1)$. One shortcoming of this way of measuring the distance-equivalent is that it is 'unitless' and does not vary with a change in the unit of measurement of distance (Parsley & Wei, 2001). Nevertheless, it is standard in the literature to express the distance-equivalent of the border effect in terms of the unit of measurement of distance in the sample.

[29] We also consider whether the effect on relative prices of crossing a shared border differs to that of crossing multiple borders by separating the border dummy in equation (5) into distinct dummy variables for between-country district pairs in contiguous and non-contiguous countries. When this is done, the estimated coefficient on the non-contiguous border dummy is markedly larger across all variants of the core specification.

sold in different districts. To the extent that these differences may be magnified across districts located in different countries, then they may be picked up by the border dummy, resulting in an upward bias on the border coefficient. With this concern in mind, the robustness of the regression results is tested using only the sub-sample of highly homogenous products.

While the magnitudes of the estimated coefficients change when the sample is restricted to highly homogenous products, the general conclusions remain as before. As anticipated, the estimated border coefficient declines when the sub-sample of only highly homogenous products is used (columns (5) and (6) in Table 5), suggesting that some of the observed price dispersion between countries previously attributed to the presence of the border may, indeed, be due to unobserved heterogeneity within products across locations. Relative to average price dispersion within Zambia (the benchmark country), crossing a national border in the SADC region adds, on average, 8.4% to the relative prices of homogenous products between countries. Now, just 3.8km in additional distance between districts in the SADC region would generate the same degree of price dispersion as that generated by crossing a national border.

The larger distance coefficient when the sub-sample of highly homogenous products is used is likely due to the dominance of fresh fruit and vegetables in the homogenous products group. Per unit transportation costs are typically higher for perishable foods (which may require refrigeration and special packaging), and this should be reflected in a higher average distance coefficient (which is used as a proxy for transport costs) for a sample dominated by these products. Finally, the result that relative to the rest of SADC, absolute price deviations between districts located in different countries are smaller (larger) on average when those countries share common membership in SACU (COMESA) is unchanged.

As a further robustness check, the sensitivity of the core regression results presented in Table 5 is tested to the exclusion of all South African districts from the sample (see Table A.3 in Appendix A). There are several reasons why products in South Africa may differ in important ways from those in the three other SADC countries. The South African economy is considerably larger than its SADC counterparts and products in South Africa tend to be more sophisticated than those produced in the three other countries. Moreover, South Africa's seasons differ from those in the other countries in the region, which are generally more sub-tropical, and this may have important implications in the case of the food products. While the sign and significance of the log distance and border variables remain unchanged, the coefficients on the distance variable are uniformly smaller when estimated using the sample excluding South Africa. In contrast, the coefficients on the border dummy are mostly larger when South Africa is excluded. As a result of these twin effects, the distance-equivalent of the border effect generally increases markedly when the South African districts are excluded from the sample.

2.6.4 Price convergence over time

It is also instructive to investigate the degree of absolute price convergence over time within the sample of SADC countries. A decline in absolute price dispersion across the sample period would suggest that product prices (and hence product markets) are becoming more integrated in the region; as would be anticipated given the reduction in tariffs under the SADC Protocol on Trade.

The focus here is on between-country price dispersion in order to investigate whether product prices across countries in the SADC region converged between 2006 and 2009. This involves regressing the absolute value of the relative price of product *i* between districts *j* and *k* at time *t* on a time trend. Additional regressions that include interactions between the time trend and dummies for common membership in the SACU and COMESA formations are also estimated. The full specification is as follows:

$$|RP_{i,jk,t}| = \alpha + \beta_1 trend + \beta_2 trend * SACU_{jk} + \beta_3 trend * COMESA_{jk} + \delta_i + \varepsilon_{jk,t} \qquad (6)$$

where $trend$ is the time trend variable; $SACU_{jk}$ and $COMESA_{jk}$ are dummy variables equal to one if the districts in district pair *jk* are located in two countries that share common membership in SACU or COMESA, respectively and zero otherwise; and δ_i are product fixed effects. The regressions are estimated with country pair fixed effects so that the estimates explain the within country pair variation in absolute price dispersion over time.

Overall, the results reported in column (1) of Table 6 indicate that product prices became *less* integrated in the SADC region between 2006 and 2009. The coefficient on the time trend variable is positive and significant, indicating that absolute price dispersion between SADC countries *increased* over the period. However, as shown in column (2) there are important differences across the trade groupings. The coefficients on the interaction terms in column (2) indicate that price dispersion *increased* between districts located in the SACU countries (Botswana and South Africa); while, in contrast, prices became *more integrated* between the COMESA countries (Malawi and

Zambia).[30] Price gaps between SACU and COMESA countries also increased, as reflected by the significant positive coefficient on the trend variable. Importantly, these trend results contrast with the levels results presented in Table 5 (where, relative to the rest of SADC, mean absolute price dispersion was lower between the SACU pairs and higher between the COMESA pairs).

Table 6: Trend in mean absolute price dispersion for between-country district pairs, 2006-2009

	(1)	(2)
trend	0.0132***	0.0139***
	(0.000580)	(0.000709)
trend*SACU		0.00247**
		(0.00122)
trend*COMESA		-0.0403***
		(0.00414)
Observations	261,978	261,978
R-squared	0.255	0.255

Notes: All regressions include a constant and are estimated with product and country pair fixed effects. Robust standard errors are clustered by district pair and reported in parentheses. Significance at the 10 percent, 5 percent and 1 percent levels is denoted by *, ** and *** respectively.

The overall finding of a lack of price convergence in the SADC region over time is consistent with the earlier evidence that the magnitude of between-country price deviations has remained fairly similar over the sample period (see sections 2.6.1 and 2.6.2). However, the absence of price convergence in the region (except between the COMESA countries) is somewhat at odds with the evidence in other disaggregated price-based studies of product market integration in Africa.[31] In comparable work on East Africa that covers a similar time period, Versailles (2012) finds reduced deviations from the LOP over the period between 2004 and 2008 (which coincides with the advent of the customs union in the region in 2005) between two members of the EAC Customs Union (Kenya and Uganda).

Looking more broadly at retail price convergence in Africa, Edwards and Rankin (2012) find evidence of improved product market integration – reflected in declining average volatility of real

[30] These results are robust to the exclusion of the less homogenous products from the sample. The only difference when the less homogenous products are excluded is that there is evidence of price convergence between the SACU countries (Botswana and South Africa). The results are also robust to the exclusion of South Africa from the analysis (see Table A.4 in Appendix A).

[31] It is, however, consistent, with some comparable studies that focus on other regions over a similar time period. Moon (2013), for example, finds little evidence of convergence in the prices of tradable products in East Asia between 1990 and 2011. This is despite the dramatic rise in the number of free trade agreements (FTAs) in East Asia since 2000 – the number of FTAs in the region increased more than three times over between 2000 and 2010 (Kawai and Wignaraja, 2010).

exchange rates between two periods (1985-1996 and 1997-2008) – within several regional trade and monetary formations including SADC, the EAC and COMESA, as well as the *Communauté Financiére Africaine* (CFA) Franc zone and the Economic Community of West African States. More generally, they find evidence of convergence over time in both absolute and relative prices within their sample of SSA and North African countries. However, the authors note that much of the convergence occurred during the 1990s and was concentrated in North African cities; with the degree of absolute price convergence lower within SSA.

2.7 Conclusion

This chapter addresses the shortage of studies in the international literature on product price integration in developing regions. Focusing on Africa, and the SADC region in particular, highly disaggregated retail price data collected at the district level in four SADC countries is used to provide a descriptive analysis of the extent to which product prices are integrated within and between countries.

The price-based results presented in this chapter point to a degree of market fragmentation in the region. In line with the balance of evidence in the international literature, large and persistent deviations from the LOP are found both within and between the SADC countries. On average over all products, there is dispersion in product prices across districts *within* each country. It is also shown that mean within-country absolute price deviations have remained fairly stable over time; suggesting that, internally, product prices within each of the SADC countries have not, on average, become more integrated.

Also consistent with the international literature, dispersion in product prices is found to be substantially greater *between* districts located in different SADC countries in comparison to across districts *within* the same country. Within the four countries, product prices differed between districts by 34% on average in 2009; while between-country price deviations averaged 77% in that year. In general, while the price dispersion estimates for the four SADC countries are larger than estimates reported in the literature for North American and European countries, they are quite similar to those computed in comparable work in the empirical literature on developing regions (albeit larger than other estimates for different regions in Africa).

These broad findings are replicated when the sampled products are considered individually. Across products, there is evidence of considerable heterogeneity in price dispersion, even within smaller sub-groups of highly homogenous and less homogenous products. In 2009, for instance, average absolute deviations from the LOP between countries ranged from 25% for shoe polish and 28% for paraffin to 120% for 1kg of cabbage and 136% for a brassiere.

The chapter shows that product prices in the region remain dispersed. When all four countries are considered together, there is no clear evidence to suggest that product prices have become more integrated between 2006 and 2009 (with the exception of increased price integration between the COMESA countries), despite the liberalization of tariffs under the SADC Protocol on Trade. This implies that trade liberalization may not be sufficient on its own to generate greater product market integration within the region.

The chapter also provided an initial exploration of the relationships between price dispersion and distance, the presence of national borders separating districts and shared membership in regional trade and monetary agreements in the SADC. Geographic proximity (measured in terms of contiguous versus non-contiguous countries) and shared membership in the SACU are found to reduce absolute price dispersion between district pairs located in different countries. In contrast, absolute price dispersion between district pairs in the SADC region increases the further apart the districts are from each other, and is higher in the case of districts separated by a national border. The results from the preferred regression specification indicate that 100km in distance between districts causes prices to deviate from the LOP by 17.5%, while crossing a national border in the SADC region adds 11.8% to relative prices between countries. These findings are consistent with evidence in the international literature that distance and borders contribute to price dispersion across countries, while shared membership in regional monetary agreements may reduce product price dispersion across locations in different countries.

Importantly, the focus in this chapter has been on the relevance of distance and national borders for product price integration. Chapter 3 highlights a range of possible factors that may bias estimates of trade costs inferred from standard approaches in the literature (as is done in this chapter), and tests these hypothesised biases using the actual product price data introduced in this chapter. Chapter 3 does so by extending the alternative quantile regression approach introduced by Borraz et al. (2012) to analyse cross-country distance and border effects.

3. Product Market Integration and Border Effects in SADC: Methodological considerations for disaggregated price-based analysis

3.1 Introduction

A key focus in the price-based literature on product market integration is on the role played by transaction costs and border effects in generating price differentials between locations and, potentially, segmenting markets. The econometric analysis in Chapter 2, which is based on the standard methodology in the empirical literature, indicates that greater distances between districts and crossing a national border both generate dispersion in relative prices between countries in the SADC region.

In recent years, however, studies using spatial price differentials to estimate transaction costs and border effects using the standard methodology have come in for criticism on several fronts. Scholars have identified a number of possible factors that may bias estimates of distance- and border-related trade costs when the standard methodology is used. Recognition of these shortcomings has inspired recent methodological advances designed to allow for more accurate estimation of trade costs using disaggregated price level data. These have included new approaches to estimating the inequality constraint that is central to the relationship between price gaps and trade costs (see, for instance, Borraz et al., 2012; Inanc & Zachariadis, 2012; Atkin & Donaldson, 2014), as well as methods to deal with a variety of other biases generated by factors such as cross-country heterogeneity in the distribution of prices within countries (Gorodnichenko & Tesar, 2009), unobserved heterogeneity across markets (Gopinath et al., 2011; Aker et al., 2014; Brenton et al., 2014), heterogeneity in the products compared across locations, and the presence of imperfect competition and variable mark-ups across locations (Atkin & Donaldson, 2014).

This chapter contributes to the literature by presenting a critical evaluation of the standard empirical methodology and certain recent methodological advances used to estimate transaction costs and border effects. The analysis seeks to address the following specific objectives:

- To identify and explain possible sources of bias on estimates of distance- and border-related trade costs using pooled spatially defined data on product price differentials for bilateral market pairs.

- To test the magnitude and direction of these biases on estimates of distance- and border-related trade costs using actual product price data for the SADC region.

- To determine an approach that can be used in empirical studies to estimate distance and border coefficients in as unbiased a manner as possible.

The remainder of the chapter is structured as follows. Section 3.2 discusses the inequality constraint problem that is fundamental to the relationship between price gaps, trade costs and product market integration. This is followed in section 3.3 by a systematic identification and discussion of potential sources of sample seletion bias on estimates of transaction costs and border effects. This section then empirically assesses the impact of these sample selection effects on estimates of distance and border-related trade costs for the SADC region. The robustness of the core empirical results is tested in section 3.4. Section 3.5 concludes.

3.2 The Inequality Constraint Problem

In the presence of trade costs, the theoretical foundation linking prices, trade and the integration of markets, centres on a simple inequality constraint. Representing the ad valorem equivalent of all transaction costs involved in trading an identical product *i* between two markets *j* and *k* as $\tau_{i,jk}$, and the log prices of the product in each market as $p_{i,j}$ and $p_{i,k}$ respectively, the inequality can be expressed as:

$$|p_{i,j} - p_{i,k}| \leq \tau_{i,jk} \quad (1)$$

This inequality implies that the difference in prices between two markets is bounded by a no-arbitrage condition (Borraz et al., 2012). The condition states that arbitrage will not be profitable if the magnitude of the difference in (log) prices between markets *j* and *k* is less than or equal to the trade cost, $\tau_{i,jk}$.[32] In the event that the price differential is initially sufficiently large so as to exceed the cost of trading a product between two markets, the difference in prices will trigger arbitrage – a buyer will purchase the product in the cheaper location and incur all of the transaction costs involved in getting that product to another location in order to sell it there at a higher price and make a profit. By engaging in trade across locations, entrepreneurs will arbitrage away the price

[32] Whereas the strict version of the LOP will only hold in cases when there are no transport costs or trade barriers, a weak form of the LOP holds when the difference in prices for an identical good between two locations is no larger than the magnitude of the trade cost (Schulze & Wolf, 2009).

difference until it is once again no larger than the magnitude of the trade cost, and the opportunities for arbitrage are eliminated.

The potential for arbitrage implies that the price setting behaviour of firms in different locations is influenced by the magnitude of the cost of trade between locations, $\tau_{i,jk}$. In practice, by influencing market structures, costs or barriers to trade between markets have important implications for price setting behaviour. For instance, the presence of trade costs creates opportunities for firms to engage in price discrimination, and influences the degree of market power enjoyed by firms in specific locations (which impacts on whether or not they apply mark-ups – as well as the size of these mark-ups – when setting prices). Trade costs also increase the cost associated with sourcing goods from elsewhere, thereby potentially raising retail prices in consumption locations. Finally, trade costs may influence the optimal location of retail outlets.

The price-based empirical research in this area has generally focused on using equation (1) to identify the size of the trade cost, $\tau_{i,jk}$, between markets. If the no-arbitrage condition is binding, the observed difference in the price of an identical product *i* sold in markets *j* and *k*, $|p_{i,j} - p_{i,k}|$, provides a measure of the magnitude of the cost of trade between the two markets.

The main problem in estimating this relationship using price data is that it is almost always not possible to identify whether the no arbitrage constraint is binding given the available data. Consequently, most studies using spatial price differentials to estimate the impact of transaction costs and border effects on product price dispersion assume that the inequality constraint in equation (1) always holds with equality. In other words, they assume that for all possible bilateral market pair combinations: $|p_{i,k} - p_{i,j}| = \tau_{i,jk}$; and, thus, infer that spatial price differentials directly identify trade costs. On this basis, they typically estimate the following regression equation using Ordinary Least Squares (OLS) and pooling data to include all possible market pair combinations:

$$|p_{i,j} - p_{i,k}| = \alpha + \beta D_{j,k} + \delta B_{j,k} + \gamma X_{j,k,t} + \varepsilon_{j,k,t} \quad (2)$$

where $|p_{i,j} - p_{i,k}|$ is the absolute value of the log price difference for product *i* between markets *j* and *k*; $D_{j,k}$ is the log of the distance between the two markets; $B_{j,k}$ is a dummy variable equal to one if the two markets are separated by a political border; $X_{j,k}$ is a vector of additional controls; and $\varepsilon_{j,k}$ is the regression error.

Borraz et al. (2012: 4) point out, however, that in most instances trade costs between any two markets cannot be accurately estimated using a simple OLS regression of equation (2) "because prices in the two locations are an optimal choice subject to a constraint that may not be binding." In some cases, transaction costs between two markets may be so high that firms' optimal prices in each market lead to a price gap that falls within the bounds of the no-arbitrage constraint. In this case, the observed gap in prices will be *less* than the actual cost of trade between the two markets. Hence, the inclusion of this price gap observation in the regression sample will bias the estimate of the true effect of distance- and border-related trade costs *downwards*. If this sample selection bias is present, then cross-border price differentials will only provide a lower bound estimate of the true trade cost (Inanc & Zachariadis, 2012).

Since the no-arbitrage condition imposes only an inequality; instead of specifying the model as in equation (2), as the majority of studies in the empirical literature do, it should instead be expressed as:

$$|p_{i,j} - p_{i,k}| \leq \alpha + \beta D_{j,k} + \delta B_{j,k} + \gamma X_{j,k,t} + \varepsilon_{j,k,t} \tag{3}$$

Using OLS to estimate equation (3) will only produce unbiased estimates in the event that the price differential is exactly equal to the trade cost and the inequality constraint holds with equality.

3.2.1 Dealing with the inequality constraint

Production-consumption pairs

One approach used in the empirical literature to deal with this sample selection bias and ensure that the equality constraint is always binding is to restrict the sample only to production and consumption market pairs (Anderson et al., 2010; Inanc & Zachariadis, 2012; Anderson et al., 2013; Atkin & Donaldson, 2014). Proponents of the production-consumption pairs approach contend that the binding determinant of the price gap is the distance between the producer and the market. This is because for any product *i,* the inequality constraint holds with equality if that product is actually traded between markets *j* and *k*. Thus, by estimating trade costs using only spatial price data for production and consumption market pairs, the sample selection bias is attenuated because irrelevant pairs for which trade in a particular good between two markets does not occur will be excluded.

Studies that apply the production-consumption pair approach generally employ some variant of the following basic specification used by Anderson et al. (2013):

$$ln\left({}^{P_{it}}\!/_{P_{jt}}\right) = \beta_1 \ln(distance_{ij}) + \beta_2 X_{ijt} + \sum_c \delta_c city_c + \sum_t k_t year_t + \varepsilon_{ijt} \qquad (4)$$

where P_{jt} is the retail price in destination city j and P_{it} is the retail price in production location city *i*; $distance_{ij}$ is the distance from the production location city *i* to the destination city *j*; X_{ijt} is a vector of other explanatory variables (such as a border dummy or measures of wage or rent differentials across locations); and $city$ and $year$ represent city and year fixed effects, respectively.

The production-consumption pair model has been employed in several studies. To date, however, most have focused on estimating distance-related trade costs rather than border effects; and have generally focused on *intranational* trade costs.[33] These studies generally confirm that the failure to take into account the spatial relationship between the production and consumption of a good by considering the distance from the location in which the good is produced to where it is consumed, biases estimates of distance-related trade costs downwards (Anderson et al., 2010; Inanc & Zachariadis, 2012; Anderson et al., 2013; Kano et al., 2013; Atkin & Donaldson, 2014). They cite this as a possible explanation for the small distance effects on price dispersion reported in non-spatially informed studies such as Engel and Rogers (1996), Anderson and Smith (2004), Engel et al. (2005), and Bergin and Glick (2007).

One practical challenge related to estimating a production-consumption pair model is the relatively demanding data requirements: it is necessary to know the exact production and destination locations of individual products. In most cases, this information cannot be ascertain from existing price data.

Furthermore, given that the approach focuses only on a sub-set of production-consumption market pairs, it is possible that cases in which the equality constraint is binding may be excluded. It is the *potential* for trade that constrains prices, meaning that trade in products does not necessarily have to occur for the equality constraint to be binding. In this respect, the equality constraint can be binding even between two consumption locations. For instance, unless consumption districts *j* and *k* are segmented, the potential to trade product *i* between the two districts will mean that the price in district *j* will be affected by the price in district *k*, and the size of the price gap between

[33] One exception is Inanc and Zachariadis (2012) who study the impact of trade costs in generating international price differences and segmenting markets across a number of European countries.

the two districts will be bound by the magnitude of the trade cost (and by retail mark-ups in settings of imperfect competition). In this case, excluding these relevant consumption-consumption district pairs from the regression will result in a *downward* bias on the distance coefficient.[34] Similarly, the exclusion of relevant cross-country consumption-consumption pairs from the regressions will bias the estimates of the border effect *downwards*. Hence, the production-consumption pairs approach is also susceptible to sample selection bias.

Quantile regressions

Borraz et al. (2012) propose an alternative approach to deal with the sample selection bias problem. They acknowledge that the only price deviation observations that are not affected by the bias are those at the boundary of the inequality constraint – where the price gap is exactly equal to the trade cost. The authors argue that using the maximum value of the observed absolute price gap for a given distance will provide a better estimate of the true cost of trade between two locations. Using store-level product price data for Uruguay, they construct distance-border-bins in which each bin is defined by a distance range and whether or not there is a city border between stores; and compute the mean, median, 80th, 85th, 90th, 95th, 97.5th, 99th, 99.5th and 99.9th percentiles as well as the maximum observed price differential within each bin. They then estimate a sequence of quantile regressions using these percentiles, together with the maximum observed price gap, as the dependent variable. Their basic regression specification is as follows:

$$Q\left(\left|p_{j,t} - p_{k,t}\right|_n, \theta\right) = \alpha + \beta D_n + \gamma B_n + \delta B_n * D_n + \gamma Firm_n + \epsilon_n \qquad (5)$$

where Q estimates the quantile θ of the absolute price difference for all store pairs *j* and *k* that have distances that belong to bin *n*; D_n measures distances between stores that belong to bin *n*; B_n is a dummy variable equal to 1 if supermarkets are in different cities; and $Firm$ is a dummy variable equal to 1 if the price difference in that bin came from the same supermarket chain.

Borraz et al. (2012) argue that estimates based on the higher percentiles of the distribution of price differences between stores will be less affected by the sample selection bias. They contend that the maximum observed price gap is expected to provide the best estimate of the lower bound of the trade cost. These assertions are backed by their empirical results. Using product-level price data for 202 products sold in 333 supermarkets across 47 cities within Uruguay, the authors show that

[34] Atkin and Donaldson (2014) attempt to deal with this issue indirectly by estimating location specific mark-ups and adjusting their price gap observations to account for the spatial variation in mark-ups.

the distance-equivalent impact of city borders on market segmentation is over estimated when the standard approach based on the average deviation of prices across cities is used. In fact, they find that the distance-equivalent effect of the border becomes small and insignificant when the upper quantiles of the price distribution within each bin are used.

3.3 Testing Potential Biases on Estimates of Transaction Costs and Border Effects

In this section, the quantile regression methodology (rather than the production-consumption pair approach) is used in order to systematically test the influence of different sources of bias on estimates of transaction costs and border effects. The focus is on examining the effects of different types of sample selection bias, referring in this case to how the nature of the sample of bilateral price pairs used in spatial price regressions may bias estimates of the true distance and border coefficients.

The empirical analysis draws on the same bilateral district-pair product price dataset initially employed in Chapter 2 (see section 2.4 for details on the data). The sample used in the empirical analysis is again restricted to the 2006 to 2009 period for which data is available for all four SADC countries. To date, no study has employed the quantile regression approach to estimate cross-country border effects. In their application of the approach, Borraz et al. (2012) consider only *intranational* city border effects within Uruguay. This chapter thus extends the empirical literature by applying the quantile regression approach to study national border effects on *international* price dispersion in the SADC region.

To implement the quantile regression methodology, separate distance-border-bins are defined. A total of *N* bins are constructed, with each bin, b_n, defined by a distance, D_n, and whether or not the districts are separated by a national border. The distance bins are constructed by first identifying the maximum (3,951 km) and minimum (4.8 km) road distances between districts in the sample. The log values of these distances are then used to construct discretely spaced bins with increments of: $\frac{1}{N}\left[log(3{,}951) - \log(4.8)\right]$.

Using these bins, the maximum absolute price difference, together with the mean, median, 80th, 90th, 95th and 99th percentiles of the distribution of absolute price gaps within each bin are computed. In effect, this results in a series of bins each containing one observation of the maximum or q[th] percentile of all absolute price deviations between districts *within* countries that are located within specific distance ranges of each other; and, for the same distance ranges, separate bins each containing one observation of the maximum or q[th] percentile of all absolute price deviations between districts located in different countries. These price gap observations are used as the dependent variable to estimate the following basic specification adapted from Borraz et al. (2012):

$$Q\left(\left|p_{j,t} - p_{k,t}\right|_n, \theta\right) = \beta_0 + \beta_1 D_n + \beta_2 B_n + \gamma_t + \gamma_i + \epsilon_n \quad (6)$$

where $Q\left(\left|p_{j,t} - p_{i,k,t}\right|_n, \theta\right)$ represents the quantile θ of the absolute value of the log price difference at time *t* for all district pairs *j* and *k* separated by distances that fall within bin b_n;[35] D_n is the log road distance between district pairs that are included in bin b_n; B_n is a dummy variable equal to 1 if districts *j* and *k* are separated by a national border; γ_t and γ_i are time and product fixed effects,[36] respectively; and ϵ_n is the regression error term.

In order to undertake the cross-country analysis, the quantile regression approach is extended here by constructing separate within-country distance bins by country and between-country distance bins by country pair. The motivation for doing so in the within-country case is that pooling all bilateral price gap observations within different countries for a specific distance range into a single bin would mean that the maximum price gap in each bin would be selected from the country that has the largest bilateral price deviation across domestic markets.[37] In this case, the average distance coefficient estimated from the quantile regression may be biased *upwards*. A similar motivation is behind the separation of the between-country distance bins by country pair.[38]

In the analysis that follows, results are reported for quantile regressions using either the mean, 95[th] percentile or maximum of the price gap distribution in each bin as the dependent variable. The focus is on the preferred specification that uses the maximum price gap in each bin (and all

[35] The q[th] percentile of the price gap is selected from each bin for inclusion in the regression. All other observations in each bin are dropped.
[36] The product fixed effects are included to account for the product-specific effect influencing relative prices.
[37] Price gaps may be systematically higher within one country if, for example, there are higher transport costs in that particular country (or within a region in that country).
[38] Table B.1 in Appendix B compares the distance and border coefficients estimated from the standard Borraz et al. (2012) application of the quantile regression approach with those estimated using separate within-country distance bins for each country and separate between-country distance bins for each country pair. The results corroborate the arguments made for separating within country bins by country and between-country bins by country pair. When the maximum price gap within each bin is used as the dependent variable, the distance coefficient is *smaller* and the coefficient on the border dummy is *larger* when estimated using the specification with bins separated in this manner.

discussions interpreting the coefficients are based on this specification), with the results for the mean and 95th percentile included to provide additional checks on the main conclusions.

3.3.1 Testing for sample selection bias in applications of the standard methodology

As a starting point, the assertion that standard estimates of distance and border effects in the literature are biased downwards due to sample selection bias is tested using the data for the SADC region. This is done by comparing the coefficients obtained from a simple pooled OLS regression of the price gaps from all district-pair combinations in the sample on (log) distance and a border dummy (based on the specification in equation (2)) with those computed from the quantile regression specification in equation (6) using 500 distance-border-bins. The estimation results are compared in Table 7.

Table 7: Comparison of standard pooled OLS regression and quantile regression estimates

	Pooled OLS regression	Quantile regressions		
		mean	p95	max
	(1)	(2)	(3)	(4)
log distance	0.0447***	0.0371***	0.0727***	0.141***
	(0.00201)	(0.00355)	(0.00656)	(0.00963)
border	0.160***	0.211***	0.338***	0.312***
	(0.00322)	(0.00712)	(0.0144)	(0.023)
Observations	412,972	4,754	4,754	4,754
R-squared	0.229	0.464	0.451	0.381
Distance-equivalent of the border effect	35.9	295.1	104.5	9.1

Notes: For the quantile regressions, the dependent variable is the mean, maximum or 95th percentile of the absolute value of the log price difference between districts *j* and *k* in year *t* for bin *n* (using 500 distance bins). All regressions are estimated with a constant and year and product fixed effects. Robust standard errors (reported in parentheses) are clustered by district pair for the pooled OLS regression and by distance bin for the quantile regressions. Significance at the 10 percent, 5 percent and 1 percent levels is denoted by *, ** and *** respectively.

The quantile regression results (in columns (2) through (4)) corroborate the Borraz et al. (2012) finding that the pooled OLS estimates (in column (1)) are biased downwards. The quantile regression estimates for the distance variable are larger than the pooled OLS estimate from the 95th percentile onwards. Based on the estimates from the quantile regressions using the maximum price gap, a distance of 100km between districts causes prices in the two districts to deviate by 64.9%[39] (versus 20.6% when estimated using pooled OLS). For all percentiles, the coefficients on

[39] Following Versailles (2012), this is calculated as: $0.141 * \ln(100)$.

the border dummy are larger when the quantile regressions are used. When estimated from the quantile regressions with the maximum price gap, crossing an international border raises the average difference in relative prices by 31.2% (versus 16% when estimated using pooled OLS). For both the distance and border estimates, the downward bias is more prominent the lower the quantile estimated.

In contrast, the distance-equivalent of the border effect computed from the pooled OLS regression (35.9km) is larger than that for the quantile regressions when the maximum price gap is used as the dependent variable (9.1km). This suggests that the pooled OLS regression overestimates the distance-equivalent of the border effect. This is also the case for the lower quantiles in the quantile regressions: the distance-equivalent declines monotonically as the higher percentiles of the price gap distribution are used. Indeed, when moving from the mean to the maximum price gap, the distance equivalent declines from 295.1km to just 9.1km. This result is similar to that found by Borraz et al. (2012: 5) who show that as the higher percentiles of the distribution of price gaps between Uruguayan cities are used, the distance-equivalent of the border effect "falls (almost) monotonically towards zero."[40]

The quantile regression results do not appear to be overly sensitive to the selection of bin width. To illustrate this, Table B.2 in Appendix B presents the estimated coefficients from quantile regressions using 1,000 distance-border bins. The results are very similar to the quantile regression estimates reported in Table 7 (which use 500 bins), with only marginal differences in the size of the distance and border coefficients.[41] At least for the regression using the maximum price gap the computed distance-equivalent of the border effect is also similar. Consequently, 500 distance bins are used in all of the quantile regressions that follow.

The analysis now moves on to analyse the impact of additional sample selection biases that are not dealt with in the Borraz et al. (2012) application of the quantile regression approach. The focus in

[40] Borraz et al. (2012) calculate the distance equivalent in a different way. For a specific distance, they first compute the degree of price dispersion between stores located in different cities using the coefficients estimated from the quantile regression. For example, for a distance of 10km, price dispersion, PD, between stores located in different cities is calculated as: $PD = \beta_0 + \beta_D * 0.1 + \beta_B - \beta_{D*B} * 0.1$, where β_0, β_D, β_B and β_{D*B} are the estimated coefficients for the constant, distance, border dummy and distance-border interaction variables, respectively. They then solve for the equivalent distance, X, that would be required in order to generate the same degree of price dispersion between two stores located in the same city as follows: $X = \frac{PD - \beta_0}{\beta_D}$.

[41] In general, the estimated distance coefficients are larger when 1,000 bins are used (except for the quantile regression using the maximum price gap in each bin). In turn, the coefficients on the border dummy tend to be marginally smaller when 1,000 bins are used (the opposite is true when the 99th percentile and the maximum price gap are used).

on testing the sensitivity of the transaction cost and border effects estimates to product or distance samples of bilateral price gaps. The effect of omitted variable bias arising in the presence of unobserved heterogeneity within narrowly defined products is also investigated.

3.3.2 Product selection bias

The standard empirical approaches in the literature that pool bilateral price gap observations for different types of products to estimate *average* distance and border effects on product price dispersion suffer from an additional selection problem. This is because by imposing a common coefficient on all product-level betas they implicitly assume that distance and border effects on relative prices are identical across all types of products. This is not necessarily the case. For a given distance, trade costs may differ from product-to-product. For example, trade in perishable goods is more likely to be localised than trade in durable goods, which can typically be transported more easily across larger distances. Hence, the cost per unit of distance associated with trading perishable goods is likely to be higher than that for durable goods.

The standard applications of the quantile regression approach introduced by Borraz et al. (2012) as well as the extension introduced above are particularly susceptible to this sample selection problem. This is because they pool absolute price differences for different types of products in the same distance bins. If distance or crossing a border affects the type of products traded across markets, then pooling market-pair absolute price differences for different types of products in the same distance bins, and selecting only one price gap observation from each bin, may bias estimates of the cost of trade. For example, when using the maximum observed price gap observation as the dependent variable, the product for which the price difference is the *largest* in each bin will be included in the regression. Thus, in each bin the selected price gap observation will reflect the product for which distance has the greatest impact on relative prices and, thereby, bias the average distance coefficient *upwards*.

Similarly, the maximum price gap observations in the bins for cross-country market pairs are likely to be drawn from products for which the effect of the border on relative prices is the largest. Thus, the average border coefficient estimated from quantile regressions using distance-border-bins that pool all products is expected to be biased *upwards* as well. Furthermore, it is unlikely that the maximum price gap in the between-country bins will be drawn from the same product as that in the within-country bins.

Testing for product selection bias in applications of the quantile regression methodology

To deal with the product sample selection bias associated with the Borraz et al. (2012) application of the quantile regression method, this sub-section introduces a further extension by constructing separate bins for each product in the sample. This is done so that the distance bins each contain one bilateral price gap observation for each product, with the aim of ensuring that the distance and border estimates are closer to the average effect for all products. The modified quantile regression specification with product-specific distance bins, $b_{i,n}$, is as follows:

$$Q\left(\left|p_{i,j,t} - p_{i,k,t}\right|_n, \theta\right) = \beta_0 + \beta_1 D_n + \beta_2 B_n + \gamma_t + \gamma_i + \epsilon_n \quad (7)$$

In Table 8, columns (1) to (3) replicate columns (2) to (4) of Table 7 and are compared to estimates obtained using product-specific bins as per equation (7). The distance coefficient is smaller when estimated with product-specific bins (however, this is not consistent when using the 95th percentile and the mean). The increase in absolute price dispersion generated by a distance of 100km between districts falls from 64.9% to 43% when the bins are constructed by product. The smaller estimate when distance-border-product-bins are used is consistent with the product selection bias hypothesis. The result confirms the intuition that when price gaps for all products are pooled within the same bin (as done in Borraz et al. (2012)), drawing the maximum observed price gap from each bin means selecting the product for which distance has the greatest impact on relative prices. This generates an upward bias on the distance coefficient.

The coefficients on the border dummy are smaller when estimated using quantile regressions with product-specific bins. Crossing a national border in the SADC region adds 22.7% to the absolute price gap between districts (compared to 31.2% when estimated by pooling price gap observations for different products in the same bin). Again, this result is consistent with the product selection bias hypothesis, and shows that the border effect is overstated when all cross-border product price gap observations for a given distance range are pooled into a single bin. In this respect, the results suggest that the construction of distance-border-bins by product allows for more accurate estimation of the border effect when using the quantile regression approach.

Table 8: Quantile regression estimations with and without accounting for product selection bias

	Without accounting for product selection bias			Accounting for product selection bias		
	(1)	(2)	(3)	(4)	(5)	(6)
	mean	p95	max	mean	p95	max
log distance	0.0371***	0.0727***	0.141***	0.0406***	0.0919***	0.0933***
	(0.00355)	(0.00656)	(0.00963)	(0.00278)	(0.00490)	(0.00498)
border	0.211***	0.338***	0.312***	0.195***	0.224***	0.227***
	(0.00712)	(0.0144)	(0.0230)	(0.00526)	(0.0129)	(0.0131)
Observations	4,754	4,754	4,754	92,631	92,631	92,631
R-squared	0.464	0.451	0.381	0.298	0.368	0.370
Distance-equivalent of the border effect	295.1	104.5	9.1	121.9	11.4	11.4

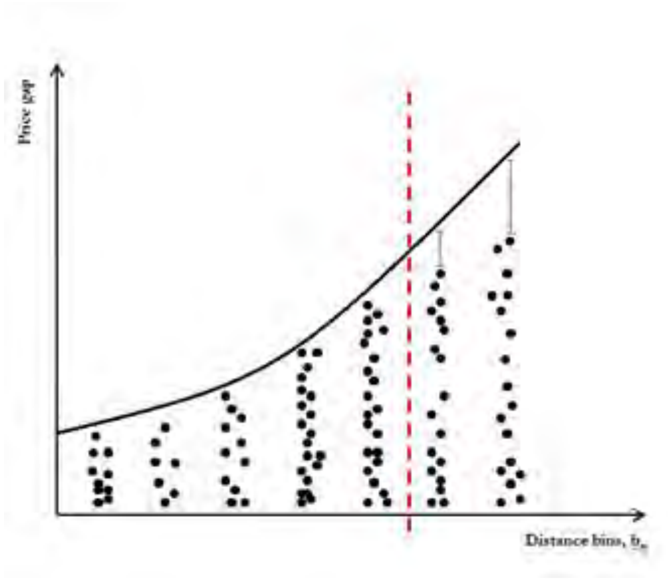
Notes: The dependent variable is the mean, 95th percentile or the maximum of the absolute value of the log price difference between districts *j* and *k* in year *t* for bin *n* (using 500 distance bins). All regressions are estimated with a constant and year and product fixed effects. Robust standard errors (reported in parentheses) are clustered by distance bin. Significance at the 10 percent, 5 percent and 1 percent levels is denoted by *, ** and *** respectively.

3.3.3 Distance and market pair selection bias

The standard price-based empirical approaches used to infer trade costs may suffer from a further sample selection bias related to the magnitude of distances separating market pairs. In the absence of a sample in which the equality constraint holds for all market pairs, the sample selection problem in section 3.3.1 may be present for market pairs separated by any distance. However, as the distance between markets increases, the transaction costs associated with trade are likely to rise. Hence, at greater distances there are more likely to be instances in which price gaps between markets are less than trade costs and the spatial arbitrage conditions are not binding. The inclusion of these observations in the market pair regressions will bias the average distance and border estimates *downwards*.

This potential distance sample selection bias problem applies in the case of the quantile regression methodology as well. The intuition behind the potential downward bias on quantile regression estimates of distance-related trade costs is illustrated in Figure 5. The thick black line represents the true upper bound of the no-arbitrage constraint, while the dotted line denotes the distance beyond which the spatial arbitrage conditions are no longer binding. For the distance bins beyond this line, even the maximum observed price gap (represented by the highest black dot) is well below the price difference implied by the no-arbitrage constraint. In these cases, the gap between the highest black dot and the upper bound of the no-arbitrage constraint is indicative of the size of the downward bias when the maximum observed price difference is used in the quantile regressions.

Figure 5: Distance sample selection bias in quantile regressions



Testing for distance sample selection bias in applications of the quantile regression methodology

Borraz et al. (2012) do not test the sensitivity of their distance and border estimates to the use of different distance samples. Consequently, they do not explicitly account for the possibility that the spatial arbitrage conditions may not be binding for Uruguayan city pairs beyond a certain distance apart. In this section, the sensitivity of the transaction cost and border effect estimates obtained from the modified quantile regressions is tested to different cut-off points for the maximum distance separating district pairs. This is done by enlarging the sample in consecutive regressions by adding district pairs that are separated by progressively higher maximum distances.[42]

The estimation results from quantile regressions using the maximum price gap in each bin are presented in Table 9.[43] For each successive quantile regression, the maximum distance separating district pairs in the sample is increased in increments of 100km. In the final regression in column

[42] The approach of employing specific distance cut-offs has been applied elsewhere in the literature, although the motivation for doing so has mostly been to deal with the possible effect of unobserved heterogeneity across markets – which is expected to be greater between more distant markets (Nchake, 2013; Robinson, 2013; Aker et al., 2014). Aker et al. (2014), for example, restrict the sample in their market-pair analyses to market pairs that are separated by no more than 250km.

[43] The results are qualitatively the same when the 95th percentile of the observed price gap in each bin is used as the dependent variable in the quantile regressions.

(10) in Panel B the distance between district pairs is restricted to 1,798km – the maximum distance that separates within-country district pairs in the sample. This ensures that both within-country and between-country district pairs are included in the regression sample.

Table 9: Quantile regressions using sub-samples of district pairs with distance restrictions

Panel A					
	(1) 800km	(2) 900km	(3) 1000km	(4) 1100km	(5) 1200km
log distance	0.0982***	0.1000***	0.100***	0.0954***	0.0896***
	(0.00507)	(0.00472)	(0.00448)	(0.00463)	(0.00465)
border	0.105***	0.115***	0.131***	0.148***	0.163***
	(0.0100)	(0.0101)	(0.0107)	(0.0118)	(0.0123)
Observations	48,996	53,034	56,633	60,003	63,256
R-squared	0.277	0.284	0.294	0.297	0.298
Distance-equivalent of the border effect	2.9	3.2	3.7	4.7	6.2
Panel B					
	(6) 1300km	(7) 1400km	(8) 1500km	(9) 1600km	(10) 1798km
log distance	0.0872***	0.0860***	0.0854***	0.0848***	0.0863***
	(0.00458)	(0.00466)	(0.00462)	(0.00457)	(0.00476)
border	0.172***	0.187***	0.197***	0.204***	0.215***
	(0.0123)	(0.0131)	(0.0133)	(0.0133)	(0.0132)
Observations	66,261	69,284	72,027	74,135	78,376
R-squared	0.301	0.309	0.316	0.321	0.335
Distance-equivalent of the border effect	7.2	8.8	10.0	11.1	12.1

Notes: The dependent variable is the maximum of the absolute value of the log price difference between districts j and k in year t for bin n (using 500 distance bins). All regressions are estimated with a constant and year and product fixed effects. Robust standard errors (reported in parentheses) are clustered by distance bin. Significance at the 10 percent, 5 percent and 1 percent levels is denoted by *, ** and *** respectively.

Looking first at the trend in the distance coefficients in Table 9, as the maximum distance between district pairs included in the quantile regression increases (from 1,000km onwards), the magnitude of the estimated distance coefficient declines monotonically (with the exception from a small increase when the maximum distance is raised from 1,600km to 1,798km). As a result, the effect in raising absolute deviations from the LOP generated by a distance of 100km between district pairs drops from 46.1% to 39.7% as the maximum distance between districts is increased from 1,000km to 1,798km. The implication is that the inclusion in the regression of price gap observations for district pairs separated by larger distances increases the likelihood of including price gaps for district pairs drawn from distance bins where the observed price gaps are less than the actual transaction costs of trade; leading to selection bias and a lower estimated distance coefficient. With each successive increase in the maximum allowed distance between district pairs, the magnitude of the selection bias grows.

In contrast, the coefficients on the border dummy, together with the distance-equivalent of the border effect, *increase* monotonically as the distance restriction between district pairs is progressively relaxed. The impact of a national border on the average difference in absolute prices increases from 10.5% to 21.5% as the maximum distance between districts included in the regression sample is increased from 800km to 1,798km. In turn, the distance-equivalent of the border effect rises from 2.9km to 12.1km as the distance restriction is relaxed.

One possible explanation for the rise in the border coefficient is that as the distance bins increase there is more likely to be unobserved product heterogeneity between cross-country district pairs, which may be picked up by the border dummy (the issue of unobserved product heterogeneity is explored in detail in section 3.3.4 below). Similarly, there are more likely to be differences in unobserved district characteristics at greater distances (Brenton et al., 2014); which, in the absence of specific controls for these characteristics, may be picked up by the border dummy. Finally, the border dummy may be picking up the influence of remote (non-contiguous) border effects at larger distances.

It is also possible that the changes in the border effect estimates presented in Table 9 are driven by a composition effect. This could arise due to the fact that the sample of country borders included in the regressions changes as the distance restriction is relaxed. To eliminate this possibility, the quantile regressions are re-run with a sub-set of country borders that remains stable as the maximum distance between district pairs is extended.[44] The results from quantile regressions estimated using the stable sub-set of countries are presented in Table B.3 in Appendix B. The results remain qualitatively the same as before, suggesting that the sample bias on the border effect coefficient at greater distances is not due to any country pair composition effect.

[44] The stable sub-set of country borders is Botswana-South Africa, Botswana-Zambia and Malawi-Zambia. Observations for between-country district pairs involving these countries are included in the quantile regressions together with the within-county district pairs for each country.

3.3.4 Omitted variable bias due to unobserved heterogeneity within narrowly defined products

Variation in product characteristics – such as quality or brand name – across locations may be reflected in differences in the final retail prices of products sold in different locations. If this is the case, then the observed price gap for a specific product across two locations will not only be driven by the cost of trade between the two locations but also by actual differences in the characteristics of the product sold in each location.

Product heterogeneity across locations may itself arise due to the effect of trade costs. According to the theorem first proposed by Alchian and Allan in 1964, the presence of a fixed per-unit transport cost, for example, that is applicable to substitutable high-quality and low-quality variants of a particular product will generate differences in the quality of products available at their source and destination. The effect of the per-unit transport cost is to reduce the relative price of the high-quality variant in the consumption destination, which tends to produce a substitution effect in favour of consuming the high-quality product at the destination. The net result is that only the high-quality products are shipped out of the source location,[45] resulting in variation in product quality between the source and consumption locations.

By not including an appropriate measure of variation in product quality or characteristics across locations, the price-based regressions in the literature may attribute the portion of the price gap that is actually generated by unobserved product heterogeneity to the role of distance or border-related trade costs; resulting in omitted variable bias on the estimated distance and border coefficients. For example, quality differences within narrowly defined products are likely to rise with distance, while quality is also positively correlated with price gaps. Consequently, the estimated distance coefficient may pick up the quality effect in addition to the effect of the trade cost.

A comparison of perfectly homogenous products across locations is required to fully overcome these problems. Recognising this, most studies use disaggregated prices for a set of very narrowly defined goods and services that is common across all locations. Within this group, some studies use barcode-level data in order to precisely identify identical products sold in different stores or locations (see, for instance, Broda & Weinstein, 2008; Gopinath et al., 2011; Atkin & Donaldson,

[45] The classic example of this is the so-called 'Washington apples effect', which results in only the good quality apples being shipped out of Washington.

2014). However, unobserved heterogeneity may be present even in comparisons of products with identical barcodes if, for example, products for sale at a particular location come with additional bundled services that are not available at another sale location.

Testing for bias stemming from the presence of unobserved within-product heterogeneity across locations

While the 24 narrowly defined products used in this chapter are highly similar across the countries included in the sample, they are not perfectly homogenous. In many instances, a lack of detail on brand names and descriptions of product-specific characteristics in the raw country price data means that there may be unobserved differences in the quality of individual products across countries (or even districts within countries). These differences may generate deviations in product prices between districts in different countries for ostensibly common products over and above the differences in prices that are driven by the cost of trade between the two districts.

As an initial test of the potential influence of within-product heterogeneity across locations, the quantile regression estimates obtained from the full sample of 24 nearly homogenous products (presented initially in Table 8) are compared to estimates from separate quantile regressions run for two sub-sets of the full sample: the first with only highly homogenous products; and the second with less homogenous products (see Table 10).[46] The highly homogenous group consists of agricultural products (fresh fruits and vegetables) and paraffin, all of which are unlikely to differ much across districts, and are generally highly substitutable. In comparison, there is likely to be greater scope for within-product variety across districts in the case of the less homogenous products. Hence, it is anticipated that the coefficients on the distance and border variables will be smaller when estimated using the sub-set of highly homogenous products.

As expected, the border coefficients are smaller when estimated using the sub-set of highly homogenous products compared to those obtained from the full sample and the sub-set of less homogenous products (except in the case of the regressions using the mean price gap). Similarly, the values for the distance-equivalent of the border effect are largest when the sub-set of less homogenous products is used.

[46] These products are distinguished using Rauch's (1999) highly disaggregated product classification scheme (see Table A.1 in Appendix A).

Table 10: Comparison of quantile regression estimates – full sample versus highly homogenous products

	Full sample			Highly homogenous products			Excluding highly homogenous products		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	mean	p95	max	mean	p95	max	mean	p95	max
log distance	0.0406***	0.0919***	0.0933***	0.0567***	0.119***	0.122***	0.0306***	0.0747***	0.0753***
	(0.00278)	(0.00490)	(0.00498)	(0.00406)	(0.00627)	(0.00642)	(0.00296)	(0.00484)	(0.00487)
border	0.195***	0.224***	0.227***	0.206***	0.201***	0.207***	0.188***	0.237***	0.239***
	(0.00526)	(0.0129)	(0.0131)	(0.00772)	(0.0153)	(0.0157)	(0.00535)	(0.0125)	(0.0126)
Observations	92,631	92,631	92,631	35,808	35,808	35,808	56,823	56,823	56,823
R-squared	0.298	0.368	0.370	0.291	0.369	0.373	0.309	0.372	0.372
Distance equivalent of the border effect	121.9	11.4	11.4	37.8	5.4	5.5	465.8	23.9	23.9

Notes: The dependent variable is the 95th percentile or the maximum of the absolute value of the log price difference between districts *j* and *k* in year *t* for bin *n* (using 500 distance bins). All regressions are estimated with a constant and year and product fixed effects. Robust standard errors (reported in parentheses) are clustered by distance bin. Significance at the 10 percent, 5 percent and 1 percent levels is denoted by *, ** and *** respectively.

However, contrary to expectations the distance coefficients are larger for the highly homogenous products. The larger distance estimates may be due to the dominance of fresh fruit and vegetables in the highly homogenous products sub-sample (seven of the nine homogenous products are perishable fruits or vegetables). Per unit transportation costs are typically higher for perishable foods, which would likely be reflected in a higher average distance coefficient for a sample dominated by these products. Indeed, the empirical evidence in Table B.4 in Appendix B – which reports the distance and border estimates obtained from quantile regressions using the highly homogenous products sub-set but with the fruit and vegetables excluded – suggests that this is the case.[47] The estimated distance coefficients are notably smaller when the fruit and vegetable products are excluded from the highly homogenous products sample.

These results provide some preliminary evidence to suggest that unobserved within-product heterogeneity across districts may be an important factor within the sample of 24 nearly homogenous products.

[47] The homogenous products sample used in these regressions now only includes two products: rice and paraffin.

Nearly homogenous versus perfectly homogenous products

In order to provide a more rigorous examination of the extent of the potential bias in the estimates arising due to the presence of within-product heterogeneity across districts, the same quantile regression methodology is applied to an alternative dataset of *perfectly* homogenous products that share exactly the same brand and unit across districts.[48] Within this alternative dataset, only Botswana and Zambia share more than two or three identical products – prices for eight perfectly homogenous products are observed in both countries. The analysis in this section is thus restricted to districts in Botswana and Zambia. The perfectly homogenous products included in the dataset for these two countries are: baking powder (Royal, 100g); biscuits (Eet-Sum-Mor, 200g); coffee (Ricoffy, 250g); floor polish (Cobra (white), 400ml); margarine (Butter Cup, 250g); pilchards (Lucky Star, 155g); petroleum jelly (Vaseline Blue Seal, 50g); and shoe polish (Kiwi, 50ml).

The original sample of 24 nearly homogenous products is now also restricted to price gap observations drawn from Botswana and Zambia. This allows for a comparison of the Botswana-Zambia border coefficients obtained from estimations using the dataset of perfectly homogenous products with those estimated from regressions based on the original sample of nearly homogenous products. In this way, the sensitivity of the estimated border effect is tested to the possibility of unobserved within-product heterogeneity across countries (which will not be present for the perfectly homogenous products).

The estimation results are compared in Table 11. In both cases, the quantile regressions are estimated using the preferred specification that for the product sample selection bias. As expected, the coefficients on both the log distance and Botswana-Zambia border dummy variables are consistently larger when the sample of 24 nearly homogenous products is used. The deviation in absolute prices generated by a distance of 100km between districts either within Botswana or Zambia or between the two countries falls from 45.6% to 37.3% when the relationship is estimated using perfectly homogenous products. Similarly, the impact of crossing the Botswana-Zambia border on the price gap falls from 19.1% to 12.8%. The upward bias on the estimated coefficients is larger for the border estimate in all specifications. As a result, the distance-equivalent of the

[48] As in the case of the nearly homogenous products dataset, the prices of the perfectly homogenous products are recalculated net of VAT and converted into common US$ prices. The same procedure for removing outliers (explained earlier) is implemented, resulting in the deletion of prices for 0.7% of the whole perfectly homogenous products dataset. Thereafter, the monthly common currency prices are collapsed to obtain average annual prices for each district and product combination.

border effect is uniformly smaller when calculated from the estimates obtained using the sample of perfectly homogenous products.

The results indicate that some of the dispersion in relative prices across districts that would have been attributed to distance-related trade costs and the border effect may actually be due to unobserved differences in product quality or within-product variety between districts in the SADC sample. This product heterogeneity generates an omitted variable bias that is not accounted for the in the main regressions.

Table 11: Comparison of quantile regression estimates – nearly homogenous versus perfectly homogenous products for Botswana-Zambia district pairs only

	Full sample of 24 nearly homogenous products			Perfectly homogenous products only		
	(1) mean	(2) p95	(3) max	(4) mean	(5) p95	(6) max
log distance	0.0208***	0.0970***	0.0992***	0.0197***	0.0771***	0.0811***
	(0.00252)	(0.00636)	(0.00654)	(0.00217)	(0.00538)	(0.00570)
border	0.138***	0.174***	0.191***	0.0916***	0.109***	0.128***
	(0.00472)	(0.0214)	(0.0226)	(0.00411)	(0.0129)	(0.0141)
Observations	48,703	48,703	48,703	14,978	14,978	14,978
R-squared	0.297	0.390	0.397	0.265	0.344	0.367
Distance-equivalent of the border effect	761.0	6.0	6.9	104.6	4.1	4.8

Notes: The dependent variable is the mean, 95th percentile or maximum of the absolute value of the log price difference between districts *j* and *k* in year *t* for bin *n* (using 500 distance bins). All regressions are estimated with a constant and year and product fixed effects. Robust standard errors (reported in parentheses) are clustered by distance bin. Significance at the 10 percent, 5 percent and 1 percent levels is denoted by *, ** and *** respectively.

3.4 Robustness Tests

In this section, the robustness of the key results documented above is tested in two different ways. First, the robustness of each of the main conclusions is tested after accounting for the potential influence of cross-country heterogeneity in the distribution of prices within countries as emphasised in Gorodnichenko and Tesar (2009) (this issue is discussed in detail in Chapter 2). Thereafter, the results from the preferred quantile regression specification are compared to those from an alternative estimation procedure in which trade costs are estimated using only spatial price data for production and consumption district pairs.

3.4.1 Accounting for cross-country heterogeneity in internal price distributions

Adopting the approach suggested by Gorodnichenko and Tesar (2009) to account for differences across the four SADC countries in the underlying variability in prices *within* each of the countries, the quantile regressions used to examine the various biases discussed above are re-run with the inclusion of dummies for district pairs that are both in the same country. As in Chapter 2, the dummy for Zambian district pairs is omitted from the quantile regressions since it is the country with the smallest difference between within-country and between-country district-pair price dispersion. The estimation results from these quantile regressions are presented in Tables B.5 to B.9 in Appendix B. Importantly, the main conclusions reached above – with respect to the four sample selection biases, as well as the omitted variable bias arising due to the presence of unobserved heterogeneity within narrowly defined products across districts – remain robust to the inclusion of dummies for within-country district pairs in the quantile regressions.

3.4.2 Evaluating the accuracy of the quantile regression methodology against an alternative production-consumption pair approach

In this section, the results from the estimation of the modified quantile regression approached developed in this chapter are compared with those obtained using a production-consumption pair specification. One practical difficulty in estimating the production-consumption pair model is that it is not possible to directly identify the actual production locations of each of the 24 nearly homogenous products in the SADC sample. To get around this, trade data is used to identify products where the bulk of imports into the remaining three SADC countries are sourced from South Africa.[49] The threshold in this respect is whether at least 50% of each country's total imports of a particular product from the world were sourced from South Africa in 2008. Using this criterion, the sample is reduced to nine products: bath towel, boy's shirt, cabbage, electric kettle, girl's dress, men's shirt, onions, potatoes and refrigerator. This sub-sample includes three products from the highly homogenous products group and six products that fall into the less homogenous category.

One further assumption is made regarding the production location. Based on the fact that the Gauteng province is South Africa's economic engine, it is assumed that all nine products originate from Pretoria; and the observed prices of each of the products in Pretoria are taken as the reference

[49] Even if the product for which the price is observed is not actually imported from South Africa, but is rather a substitute for the imported product, the potential to import will constrain prices in the consumption location.

price for the product district.[50] The relative prices for production-consumption district pairs for product *i* are then calculated as the absolute log price deviations from the Pretoria price for all districts in Botswana, Malawi and Zambia where product *i* is sold (the consumption districts).

The production-consumption pair regression is estimated using the following basic specification:

$$|p_{i,p,t} - p_{i,c,t}| = \beta_0 + \beta_1 D_{pc} + \beta_2 B_{pc} + \gamma_t + \gamma_i + \epsilon_n \quad (8)$$

where $p_{i,p,t}$ is the retail price of good *i* in the production district *p* (Pretoria in the regressions) at time *t* and $p_{i,c,t}$ is the corresponding retail price in the consumption district *c*; D_{pc} is the log distance from the production district to the consumption district; B_{pc} is a dummy variable equal to one if the production and consumption districts are separated by a national border; γ_t and γ_i are year and product fixed effects, respectively; and ϵ_n is the regression error term.

Table 12 compares the regression results from the pooled OLS and quantile regressions (using the maximum price gap in each distance bin) with those from the production-consumption pairs approach. Focusing first on the distance estimates of the pooled OLS (column (1)) and the production-consumption pairs (column (3)) regressions, the distance coefficient is larger in the latter case. This finding confirms that trade costs are mis-measured when data on price gaps for all possible bilateral district pair combinations are included in the regression. The result corroborates the findings in other multi-country studies that use disaggregated product price data and report larger estimates of distance-related trade costs when estimated in a spatially informed manner (Anderson et al., 2010; Inanc & Zachariadis, 2012; Anderson et al., 2013; Atkin & Donaldson, 2014).

Notably, however, the distance coefficient is larger when estimated using the quantile regression approach compared to the production-consumption pairs estimate. When estimated from the quantile regressions, a distance of 100km between districts adds 40% to deviations from the LOP, versus 30.2% when estimated using the production-consumption pairs approach. A simple z test

[50] A consistent price series over the sample period for the products in the more obvious candidate, Johannesburg, is not available. Nevertheless, using the prices for Pretoria instead is unlikely to be problematic as Pretoria is located in close proximity (around 50km) to Johannesburg and product prices in Pretoria are likely to closely match those in Johannesburg.

of the equality of the coefficients suggests, however, that the difference is *not* statistically significant.[51]

Table 12: Comparison of estimates from pooled OLS, quantile and production-consumption pair regressions

	Pooled OLS	Quantile regression	Production-consumption pairs
	(1)	(2)	(3)
log distance	0.0328***	0.0868***	0.0655***
	(0.00234)	(0.00508)	(0.0154)
border	0.177***	0.259***	0.193***
	(0.00369)	(0.0134)	(0.0208)
Observations	160,159	35,342	2,979
R-squared	0.154	0.299	0.219
Distance-equivalent of the border effect	220.6	19.8	19.0

Notes: For the quantile regressions, the dependent variable is the maximum of the absolute value of the log price difference between districts *j* and *k* in year *t* for bin *n* (using 500 distance bins). All regressions are estimated with a constant and year and product fixed effects. Robust standard errors (reported in parentheses) are clustered by district pair for the pooled OLS regression and production-consumption pair regressions and by distance bin for the quantile regressions. Significance at the 10 percent, 5 percent and 1 percent levels is denoted by *, ** and *** respectively.

Compared to the production-consumption pair estimation, the coefficient on the border dummy is larger when estimated from the quantile regressions, and the difference is statistically significant. When estimated from the quantile regression and accounting for distance, crossing a national border in the SADC region causes prices to deviate from the LOP by 25.9% (versus 19.3% in the case of the production-consumption pairs regression). The distance-equivalents of the border effect are, however, almost identical when calculated from the two different approaches.[52]

One possible explanation for the smaller estimates obtained when using the production-consumption pairs approach is that the sample of pairs represents only a *subset* of bilateral district pairs for which the equality constraint holds. By focusing only on this narrow subset of possibilities, the production-consumption pairs approach potentially ignores other relevant district

[51] The following z test formula, proposed by Paternoster et al. (1998), is used to evaluate whether the difference between the two regression coefficients, b_1 and b_2, is statistically significant: $z = \frac{b_1 - b_2}{\sqrt{SEb_1^2 + SEb_2^2}}$. The null hypothesis that $b_1 = b_2$ is rejected if the value of z is greater than 2.

[52] The finding that the distance and border effects are larger when estimated using quantile regressions compared to a production-consumption pairs approach is robust to the use of a more stringent product selection criteria in which at least 80% of each country's total imports of a particular product from the world must have been sourced from South Africa in 2008 in order to be included in the sample (see Table B.10 in Appendix B). However, when this more stringent product selection criteria is used, the distance coefficient is clearly larger when computed from the quantile regression.

pair combinations for which the arbitrage condition is also binding; and is thus also susceptible to sample selection bias.

3.5 Conclusion

This chapter presents a critical evaluation of the standard empirical methodology (and certain recent methodological advances) designed to estimate transaction costs and border effects using spatial price data. This evaluation forms the basis for a systematic identification of a number of sources of sample selection bias on empirical estimates of distance and border coefficients in the existing literature. The chapter draws on the same novel dataset of disaggregated retail prices initially employed in Chapter 2, and builds on the quantile regression approach initially introduced by Borraz et al. (2012) to test the sensitivity of the distance and border coefficients to the various sample selection biases.

Four key sample selection effects are shown to bias estimates of the impact of distance- and border-related trade costs on price dispersion in the SADC region. First, it is shown that the standard pooled OLS estimates (reported in much of the existing literature) suffer from a sample selection problem which biases the estimated distance and border coefficients *downwards* relative to the true cost of trade. This is because the standard approach of pooling data on price gaps for all possible market pair combinations is prone to the inclusion of price gap observations drawn from market pairs for which the equality constraint linking prices, trade and arbitrage is not binding.

The chapter then highlights the impact of two additional selection biases that are not dealt with in the Borraz et al. (2012) application of the quantile regression approach. It shows that by pooling absolute price differences for different types of products in the same distance bins and selecting one observation from each bin for inclusion in the regression, the standard application of the quantile regression approach involves selecting products for which distance and borders have the greatest impact on relative prices. This product sample selection bias results in larger average distance and border coefficients.

The chapter also shows that the standard approach of pooling relative prices for all possible bilateral market pair combinations regardless of the distance between them does not account for the possibility that transaction costs are more likely to exceed price gaps at larger distances. This raises the likelihood of including market pairs in the sample for which the equality constraint is not binding and where the observed price gaps are less than the actual transaction costs of trade. The quantile regression approach is also susceptible to this distance selection bias. By progressively relaxing restrictions on the maximum distance separating district pairs in the sample to test the sensitivity of the modified quantile regressions to different distance cut-offs, the chapter demonstrates empirically that the inclusion of these observations in the quantile regressions results in a sample selection bias which lowers the estimated distance coefficient. In contrast, the border estimates presented in this chapter increase in magnitude as the distance between district pairs increases, which may be due to higher levels of unobserved product heterogeneity, greater variation in unobserved district characteristics or the influence of remote (non-contiguous) border effects at larger distances.

Finally, it is shown that by not accounting for variation in within-product quality across districts, the price-based regressions attribute the portion of the price gap that is actually generated by unobserved product heterogeneity to the role of distance and border-related trade costs. This results in omitted variable bias which raises the estimated distance and border coefficients.

Even after controlling for the various sample selection effects and omitted variable bias, distance and national borders play a substantial role in generating product price dispersion between SADC countries. After accounting for distance and dealing with potential product selection effects, crossing a national border adds an additional 22.7%, on average, to relative prices between districts in the region. In distance-equivalent terms, an additional 11.4km in distance between districts in the region would generate the same degree of price dispersion as that generated by crossing a national border. When estimated using a sub-sample of highly homogenous products – which are unlikely to differ much across districts and are generally highly substitutable – the border effect falls marginally to 20.7% (with a distance-equivalent of 5.5km). The border effect falls further to 13.1% (distance-equivalent of 3.7km) when estimated using the full sample of products but restricting the maximum distance separating market pairs to 1,000km in an attempt to exclude segmented market pairs at greater distances. These figures are comparable with other estimates in the literature of border effects in different regions in Africa, but generally larger than those estimated for developed countries. On balance, the findings presented in this chapter suggest that barriers to trade stemming from the presence of national borders hamper greater product market integration within the SADC region.

The following chapter applies the extended and refined quantile regression methodology derived in this chapter in order to compute precise estimates of the magnitude of individual border effects in the SADC region. The approach is then applied to examine trends in border effects in the region over time.

4. Quantifying the Impact of Borders on Product Market Integration in the Southern African Development Community

4.1 Introduction

Despite the proliferation of regional integration agreements and concerted efforts to liberalise tariffs on intra-regional trade in goods in Africa over the past two decades, product markets on the continent remain fragmented (World Bank, 2012). This is reflected in comparatively low levels of intra-African trade. Intra-regional trade accounts for less than 12% of total African trade,[53] the lowest share across all of the world's major regions (Mbekeani, 2013). The fact that the vast majority of Africa's exports are destined for countries outside of the continent indicates that Africa's integration with the rest of the world has outpaced internal integration of national markets.

This is the case even within the continent's various RECs. Although the values of intra-REC exports and imports have generally expanded in recent years (Hartzenberg, 2011), they remain small when measured as a share of total trade. This is clearly evident within the SADC, where the expansion of exports to the rest of the world has been accompanied by relatively static intra-regional trade, a pattern reflecting a growing de-regionalisation of SADC trade (World Bank, 2011).

One possible contributory factor to the fragmentation of markets on the continent is the presence of transaction costs of trade, which have been found to be particularly high in Africa (Amjadi & Yeats, 1995; Teravaninthorn & Raballand, 2009; World Bank, 2012). Moreover, as tariffs on intra-regional trade in SADC and elsewhere on the continent have undergone liberalization, the scope for extending trade preferences has become more limited, with the result that the impact of high transaction costs and other restrictions on cross-border trade has become increasingly apparent. As the World Bank (2012: 20) attests, the "current thinking on regional integration in Africa has moved beyond removing tariffs to regulatory issues that raise trade costs and prevent goods, services, people, and capital from moving freely across borders within Africa."

In broad terms, trade costs encompass any cost – aside from the actual cost of producing a product – that is incurred in getting that product to an end-user (Portugal-Perez & Wilson, 2008). These costs can generally be grouped into four distinct categories: transport costs, border-related costs,

[53] This share is even lower within certain Regional Economic Communities (RECs) on the continent. In the case of COMESA, for instance, goods imported from within the region account for just 5% of total imports (World Bank, 2012).

behind-the-border costs, and a variety of compliance costs. High trade costs can adversely affect the competitiveness of producers and the welfare of consumers by raising the cost of imported inputs as well as the price of final goods. In doing so, high trade costs may "prevent the full realization of potential gains from trade and can wither the poverty reduction effect of export opportunities for African countries" (Portugal-Perez & Wilson, 2008: 3). Reducing these costs can serve as a catalyst for greater market integration, boost cross-border trade and support export growth.

The limited available evidence suggests that trade costs are high within SADC (Mengistae, 2012). This is particularly true for cross-border trade in the region, where costs are inflated by additional production and administrative costs associated with complying with different production and labelling standards, customs procedures and onerous rules of origin requirements; and transportation costs arising from inefficient trade administration procedures and the resulting delays at border posts (Charalambides, 2013).

These costs are exacerbated by the large number of different regional trade agreements to which Southern African countries are party, which results in numerous instances of countries with multiple and overlapping memberships in different regional integration groupings. Variation in standards, market entry requirements, trade rules and requirements across the many regional trade agreements raises the cost of exporting to different markets in the region (Keane et al., 2010; World Bank, 2012). The high costs associated with cross-border trade are likely to serve as a constraint to intra-regional trade and the integration of product markets in Southern Africa.

The central aim of this chapter is to precisely quantify the impact of border-related trade costs on the integration of product markets in the SADC region. This is achieved by using the highly disaggregated product price level data introduced in the previous chapters and the modified quantile regression approach developed in Chapter 3 to precisely estimate the effect of distance and political boundaries on product price dispersion in the region.

By examining the effect of national borders on product price dispersion in the SADC region, this chapter advances the international literature on product market integration. Empirical studies on border effects in developing regions are rare, particularly in Africa. Most studies on Africa and other developing regions have examined the influence of different types of trade costs on cross-border trade using data on the volume of trade flows between countries. However, as discussed in detail in Chapters 2 and 3, using disaggregated product price data – rather than information on trade flows – represents a more effective way to evaluate the integration of product markets (Mengistae, 2012).

As explained in the Chapter 2, most of the existing studies on Africa that employ the preferred method of using disaggregated price level data to estimate border effects are focused on Central, East or West Africa (Versailles, 2012; Nchake, 2013; Aker et al., 2014; Brenton et al., 2014), with little attention given to product market integration in Southern Africa (one notable exception is Nchake (2013), but her analysis is confined to three SACU countries), and the SADC in particular. Furthermore, the results reported in these studies are obtained using the standard empirical approach in the literature (or variants thereof) which is subject to a range of sample selection biases that may affect the accuracy of their estimates (see Chapter 3).

The analysis in this chapter extends the literature on the impact of national borders on product market integration to the SADC region. Moreover, by dealing with several potential sample selection biases that affect standard empirical analyses of the impact of borders on product price integration using disaggregated product price data, this chapter provides more accurate estimates of border effects in Africa. This is achieved by employing the quantile regression approach introduced by Borraz et al. (2012) and extended in the previous chapter, to precisely quantify the effect of distance and national borders on product market integration in the SADC region.

The remainder of the chapter is structured as follows. The next section describes the product price data and empirical specification used in this chapter. Thereafter, an analysis of the impact of borders on product market integration in the SADC is presented, which focuses on precisely estimating the role played by borders in generating product price dispersion between countries in the region. This is followed by a check on the robustness of the core results using an alternative production-consumption pair estimation technique.

4.2 Data and Empirical Specification

The empirical analysis in this chapter draws on the product price dataset initially employed in Chapters 2 and 3. This dataset contains actual monthly retail prices for a diverse set of 24 products (including various food, clothing and textiles, equipment and electronics, and other products) collected at the district-level in Botswana, Malawi, South Africa and Zambia. As before, the monthly prices are collapsed to annual averages for each district for the period from 2006 to 2009 and used to calculate annual product-level absolute price deviations for all bilateral district-pair combinations in the sample.

The analysis in Chapter 3 highlighted a range of sample selection effects that can potentially bias estimates of the impact of distance and border-related trade costs on product price dispersion across locations when disaggregated price level data is used. The empirical methodology employed in this chapter is designed to deal explicitly with these sample selection effects in order to provide trade cost estimates that are as accurate as possible.

For this purpose, the quantile regression methodology (described and extended in Chapter 3) is used again in this chapter to deal with the sample selection effect highlighted by Borraz et al. (2012) and Atkin and Donalson (2014) among others. As in Chapter 3, a total of 500 distance bins are constructed, with separate within-country distance bins constructed by country and between-country distance bins constructed by country pair. These bins are separated further by product to deal with the product sample selection effect documented in Chapter 3. Finally, adopting the same approach used in Chapter 2 (and the robustness test in Chapter 3) to address the country heterogeneity effect identified by Gorodnichenko and Tesar (2009), dummies for district pairs that are both in the same country are included in the quantile regressions in order to account for differences across the four SADC countries in the underlying variability in prices *within* each of the countries. The core specification used to estimate the impact of national borders on product market integration in the SADC region is as follows:

$$Q\left(\left|p_{i,j,t}-p_{i,k,t}\right|_{n}, max\right)=\beta_0+\beta_1 D_{jk,n}+\beta_2 B_{jk,n}+\gamma_t+\gamma_i+\sum_J \delta_{J-J} W_{jk,n}+\epsilon_n \qquad (1)$$

where $Q\left(\left|p_{i,j,t}-p_{i,k,t}\right|_{n}, max\right)$ represents the maximum of the absolute value of the log price difference for product *i* at time *t* for all district pairs *j* and *k* separated by distances that fall within bin b_n;[54] $D_{jk,n}$ is the log road distance between district pairs that are included in bin b_n; $B_{jk,n}$ is a dummy variable equal to 1 if districts *j* and *k* are separated by a national border; γ_t and γ_i are time and product fixed effects, respectively; $W_{jk,n}$ is a dummy variable equal to one if districts *j* and *k* are in the same country and zero otherwise (the dummy for South African district pairs is omitted);[55] and ϵ_n is the regression error term.

[54] The maximum price gap is selected from each bin for inclusion in the regression. All other observations in each bin are dropped.
[55] Given the relative size and significance of the South African economy within the region, it makes intuitive sense to use the level of domestic market integration within South Africa as a benchmark from which to assess price dispersion

The key coefficients of interest are those on the $W_{jk,n}$ and $B_{jk,n}$ variables, which measure the degree of market integration within and between the SADC countries, respectively, after controlling for transaction costs of trade associated with the distance between districts. The inclusion of dummies for district pairs within countries in the quantile regression specification has important implications for the interpretation of the coefficient on the border dummy. Specifically, when the dummy for South African district pairs is omitted, the estimated coefficients on $W_{jk,n}$ and $B_{jk,n}$ measure the average level of market integration within and between the SADC countries with respect to the average level of domestic market integration within South Africa.

The multi-country nature of the data means that it is possible to estimate the impact of specific borders in the SADC region. Thus, in addition to estimating the average SADC border effect using equation (1), in an alternate specification the $B_{jk,n}$ variable is split into separate dummies for each of the Botswana-Malawi, Botswana-South Africa, Botswana-Zambia, Malawi-South Africa, Malawi-Zambia and South Africa-Zambia borders.

Importantly, while the specification in equation (1) deals with the general sample selection and product sample selection biases highlighted in Chapter 3, it does not address the distance sample selection bias. In Chapter 3, it was shown that the inclusion of price gap observations for district pairs separated by large distances (in which case the observed price gaps are more likely to less than the actual transaction costs of trade) biased the estimated distance coefficient downwards, while resulting in larger border coefficient estimates (and, thus, a larger computed distance-equivalent of the border effect). Given that the focus in this chapter is on an analysis of all individual border effects among the sampled SADC countries, the maximum distance between the district pairs in the sample is *not* restricted as this would exclude certain borders from the analysis. As a result, the distance coefficients reported in this chapter may be biased downwards, while the border coefficients may be *overestimated*.

One other potential source of bias highlighted in Chapter 3, which is not addressed in this chapter, is the issue of unobserved heterogeneity across locations within narrowly defined products. Unfortunately, a key limitation of the alternative sample of *perfectly* homogenous products introduced in Chapter 3 is that only the Botswana and Zambia samples have more than one or

within and between the other countries in the region. It is important to note, however, the difference from Chapters 1 and 2 (where the Zambia-Zambia dummy is omitted in the pooled OLS and quantile regressions).

two products that can be compared between countries. This means that a broader analysis of all six individual border effects between the four SADC countries using perfectly homogenous products is not feasible. Consequently, based on the evidence presented in Chapter 3, the distance and border coefficients reported in this chapter may be *overestimated* in the presence of unobserved within-product heterogeneity across districts due to omitted variable bias.

4.3 Border Effects and Market Integration in the SADC Region

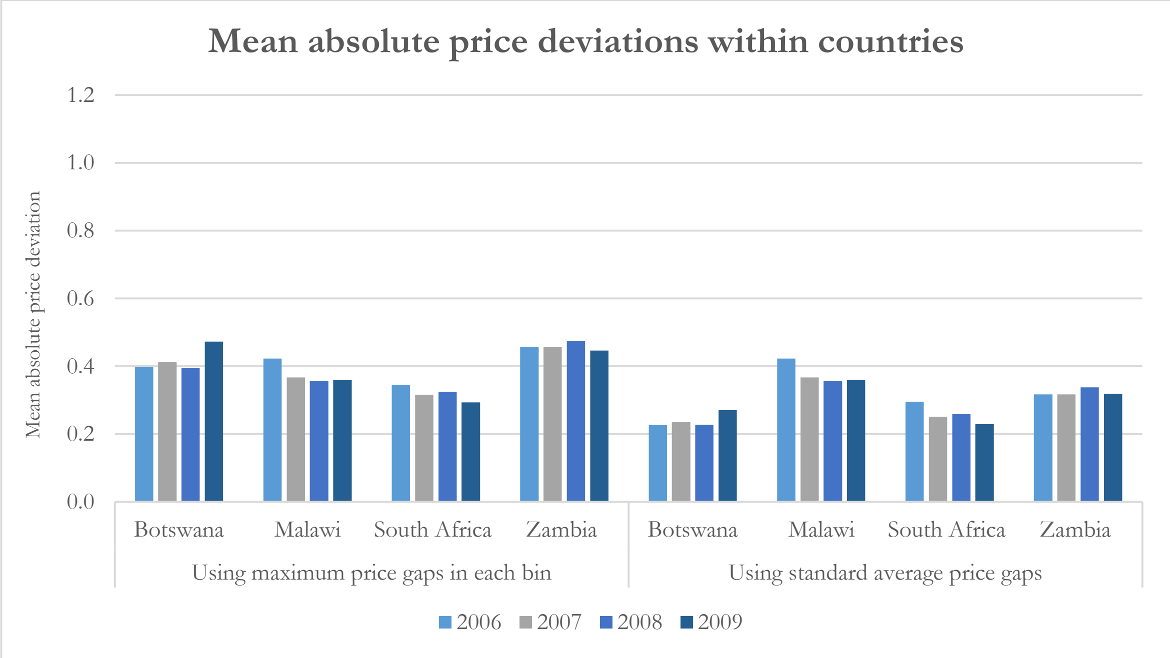
In equation (1), price deviations between districts (calculated from price level data) are used to infer information on distance- and border-related trade costs in the SADC region. To provide an initial picture of the magnitude of absolute price deviations across districts within and between the four SADC countries, and trends in these price deviations over time, panels A and B in Figure 6 plot the mean absolute price deviations across all products in each year for district-pair combinations within each country and between specific country pairs, respectively. In both panels, two different relative price measures are used. First, the maximum price gaps in each distance bin are used to calculate mean absolute price deviations (the bins are separated into distinct bins for district pairs within each country and between each country pair for a given distance range). These price gaps are the ones used in the quantile regressions in Chapter 3 as well as those that follow in this chapter. Second, mean absolute price deviations are computed in the same manner but using standard average price gaps. These price gaps correspond to those used in standard regressions in the literature.

Several interesting observations can be drawn from the relative prices outlined in Figure 6. First, both the within- and between-country relative prices are generally markedly larger when computed using the maximum price gaps within each bin.[56] Given that information about the magnitude of trade costs is inferred directly from spatial price gaps in price-based approaches to measuring product market integration, the substantially lower average price deviations is consistent with the assertion made by Borraz et al. (2012) that standard approaches are likely to underestimate the impact of trade costs on market integration by using price gaps that are lower than the actual cost of trade.

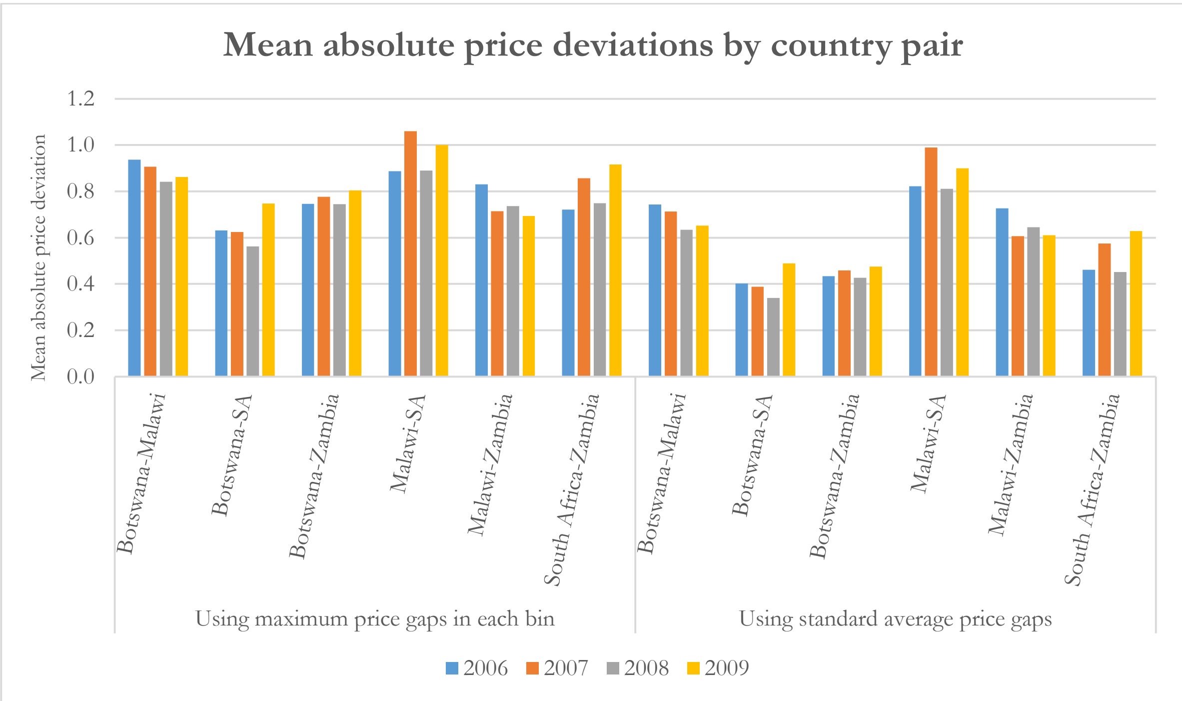
[56] The only exception is the within-country relative prices for Malawi, which are unchanged when the maximum price gap in each bin and the standard average price gaps are used.

Figure 6: Mean absolute price deviations within countries and between country pairs, 2006-2009

Panel A: Within-country relative prices



Panel B: Between-country relative prices



Focusing, then, on the average relative prices computed using the maximum price gaps, the second interesting feature relates to the trend in relative prices over time. Within the four SADC countries, internal relative prices declined, on average, in Malawi, South Africa and Zambia in 2009 compared to the initial value in 2006; and increased within Botswana over the same period. In turn, between-country relative prices declined, on average, between Botswana and Malawi, and Malawi and Zambia; while increasing between the four other country pairs (Botswana-South Africa, Botswana-Zambia, Malawi-South Africa and South Africa-Zambia). There is, nevertheless, significant variation in trends between specific years across the full period.

Finally, the between-country relative prices are noticeably larger than the comparable internal relative prices *within* the four SADC countries. This provides some cursory evidence that significant border effects on price dispersion may be present between countries in the region. In 2009, between-country relative prices were largest, on average, between Malawi and South Africa, followed by South Africa and Zambia and Botswana and Malawi. These three country pairs are all separated by comparatively larger distances and transport between them involves crossing multiple borders – pointing to possible incremental border effects on relative prices.

Econometric analysis is required to examine these issues more formally. To this end, in the remainder of this section the quantile regression specification outlined in equation (1) is used to compute precise estimates of the magnitude of average and specific border effects in the SADC. Thereafter, trends in both average and individual border effects in the region are examined over the 2006 to 2009 period.

4.3.1 The magnitude of border effects in the SADC region

Table 13 reports coefficient estimates of the effects of distance and border barriers on relative prices in the SADC region obtained using the model in equation (1). The regression in column (1) includes a single dummy to estimate the average border effect in the SADC region, while in column (2) separate dummy variables are included for each of the bilateral borders between country pairs in the sample, allowing for an analysis of specific border effects. In both regressions, the respective coefficients on the country dummies reflect the relative integration of domestic markets in Botswana, Malawi and Zambia with respect to market integration within South Africa (the omitted country dummy).

Table 13: Quantile regression estimation of average and individual border effects in the SADC region

	(1)	(2)
log distance	0.107***	0.0914***
	(0.00455)	(0.00470)
border	0.376***	
	(0.0101)	
Botswana-Malawi border		0.496***
		(0.0163)
Botswana-SA border		0.310***
		(0.0105)
Botswana-Zambia border		0.402***
		(0.0142)
Malawi-SA border		0.535***
		(0.0172)
Malawi-Zambia border		0.420***
		(0.0139)
SA-Zambia border		0.377***
		(0.0154)
Botswana	0.216***	0.205***
	(0.0118)	(0.0112)
Malawi	0.182***	0.177***
	(0.0229)	(0.0202)
Zambia	0.195***	0.192***
	(0.0101)	(0.00957)
constant	-0.257***	-0.162***
	(0.0300)	(0.0307)
Observations	92,631	92,631
R-squared	0.388	0.399

Notes: The dependent variable is the maximum of the absolute value of the log price difference between districts *j* and *k* in year *t* for bin *n* (using 500 distance bins). All regressions are estimated with year and product fixed effects as well as country-specific dummies for 'Botswana-Botswana', 'Malawi-Malawi' and 'Zambia-Zambia' district pairs (with the 'South Africa-South Africa' dummy omitted). Robust standard errors (reported in parentheses) are clustered by district pair. Significance at the 10 percent, 5 percent and 1 percent levels is denoted by *, ** and *** respectively.

All variables included in the regressions in both specifications have the expected signs and are significant at the 1 percent level. Focusing initially on the results presented in column (1), the large positive coefficient on the log distance variable indicates that greater road distances between districts have a substantial effect on product price differences. On average, a distance of 100km between districts in the SADC region causes prices in the two districts to deviate by 49.3%.[57]

The magnitudes of the coefficients on the country dummies (equal to one if district pairs are both in the same country) in Table 13 are indicative of the level of impediment to market integration

[57] Calculated as: 0.107*ln(100).

within Botswana, Malawi and Zambia relative to domestic market integration in South Africa. Interestingly, the estimates show that markets in each of these countries are less integrated than South Africa – internal price dispersion within these economies is substantially higher than within South Africa. This is consistent with the relatively poorer infrastructure in these countries. These results echo the evidence presented in Chapter 2, which revealed comparatively large mean absolute price deviations within the SADC countries. Among the four countries, domestic markets are marginally more integrated in Malawi compared to Zambia and Botswana, respectively (when measured relative to South Africa);[58] although the differences are comparatively small.

The border effect captures the additional effect of crossing the border relative to this internal price dispersion. After controlling for distance, crossing a national border generates larger absolute price differences between districts. Relative to average price dispersion within South Africa (the benchmark country), crossing a national border in the SADC region adds, on average, 45.6% to price differences between countries.[59] This effect is large compared to the average border effects reported in similar studies focused on different regions in Africa (see, for instance, Versailles, 2012; Nchake, 2013; Aker et al., 2014; and Brenton et al., 2014). However, as discussed in Chapter 3, the estimates reported elsewhere in the literature are likely to be biased downwards due to sample selection bias. Measured in distance-equivalent terms, an additional 33.6km in distance between districts in the SADC would generate the same degree of price dispersion as that generated by crossing an average border in the region.[60]

The large average border effect in the region does, however, mask substantial heterogeneity in the effects of specific borders on relative prices between SADC countries. This is evident in the variation in the coefficients on the individual border dummies estimated in column (2). These coefficients measure the level of impediment to market integration generated by crossing a specific border relative to the level of market integration with South Africa. The estimates can be used to compute specific average border effects for each SADC country pair in the sample. Following Brenton et al. (2014), the average border effect for each country pair is calculated as the difference between the specific border coefficient estimated using equation (1) and the average value of the coefficients on the within-country dummies for countries on each side of the border (which, as explained above, reflect the level of impediment to market integration within each country relative

[58] This result is consistent with Brenton et al. (2014) who find that domestic impediments to market integration (which they measure relative to market integration in Djibouti) are marginally lower in Malawi in comparison to Zambia.
[59] Calculated as: exp(0.376)-1.
[60] Calculated as: exp(0.376/0.107).

to market integration within South Africa). In turn, lower and upper bounds of the specific border effects are computed by taking the difference between the relevant border coefficient estimate and the coefficient on the dummy for each country located on a particular side of the border.[61]

The computed average and upper and lower bound border effects, as well as their distance-equivalents, are compared across different regional trade agreements as well as between contiguous and non-contiguous countries in Table 14. To aid the visualisation of differences across these dimensions, the average border effects are reproduced graphically in Figure 7.

Table 14: Border effects, regional trade agreements and contiguity

Country pair	Border coefficient	Average border effect	Lower bound	Upper bound	Distance-equivalent	Contiguous
SACU						
Botswana-SA	0.310	0.208	0.105	0.310	29.7	Yes
COMESA						
Malawi-Zambia	0.420	0.236	0.228	0.243	99.0	Yes
REST OF SADC						
Botswana-Malawi	0.496	0.305	0.291	0.319	227.4	No
Botswana-Zambia	0.402	0.204	0.197	0.210	81.3	Yes
Malawi-SA	0.535	0.447	0.358	0.535	348.4	No
SA-Zambia	0.377	0.281	0.185	0.377	61.9	No
Average SACU	0.310	0.208	0.105	0.310	29.7	
Average COMESA	0.420	0.236	0.228	0.243	99.0	
Average Rest of SADC	0.453	0.309	0.258	0.360	179.7	
Average contiguous	0.377	0.216	0.177	0.254	70.0	
Average non-contiguous	0.469	0.344	0.278	0.410	212.6	

Notes: The average border effect for each country pair is calculated as the difference between the estimated border coefficient and the average value of the coefficients on the within-country dummies for countries on each side of the border (for borders involving South Africa – the reference country, the coefficient on the within-country dummy for the other country in the country pair is simply divided by two). The lower and upper bounds of the border effects are computed by taking the difference between the relevant border coefficient estimate and the coefficient on the dummy for each country located on a particular side of the border. The distance-equivalent of the border effect is calculated using the actual estimated border coefficient.

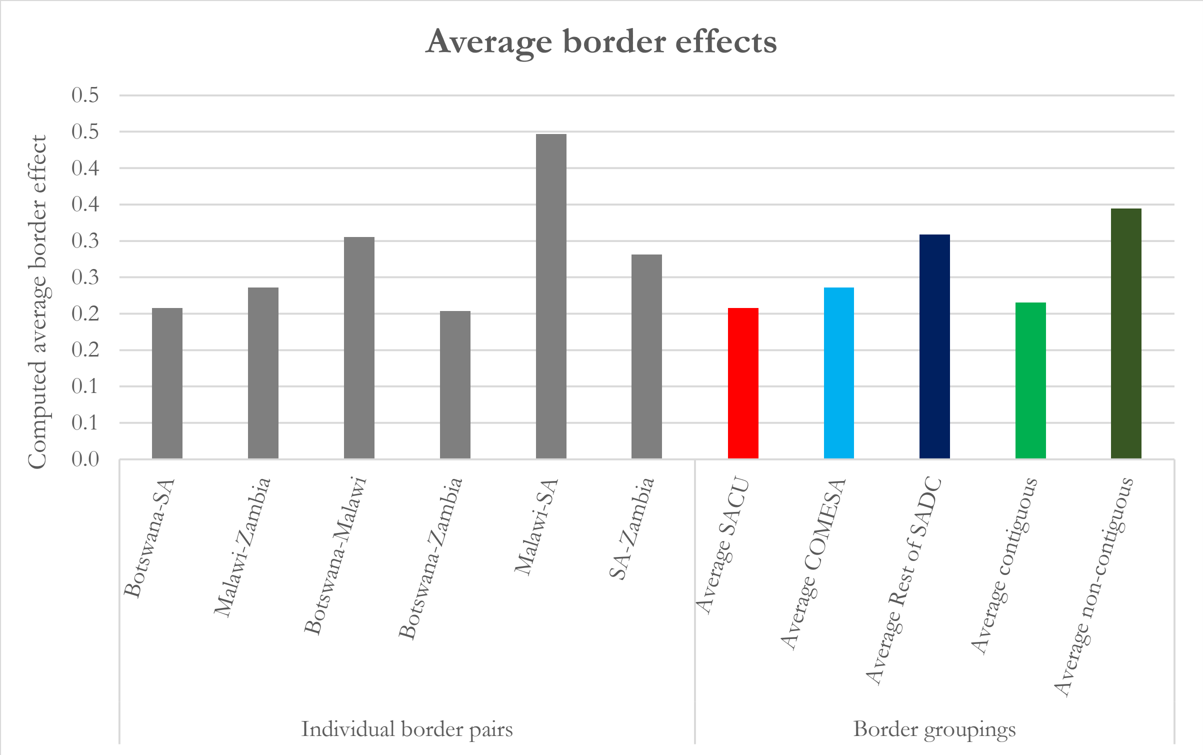
There is significant variation in border effects on price dispersion across the individual country pairs. The border effect is smallest for Botswana and Zambia, where crossing the border increases price dispersion between the two countries by 22.6% relative to the average level of price dispersion within the two countries.[62] The border effects on price dispersion are consecutively

[61] For example, the lower bound of the Malawi-Zambia border effect is calculated by subtracting the coefficient on the Zambia dummy (the country with higher internal price dispersion) from the estimated Malawi-Zambia border coefficient: 0.42-0.192. In turn, the upper bound is calculated by subtracting the coefficient on the Malawi dummy (the country with lower internal price dispersion) from the estimated Malawi-Zambia border coefficient: 0.42-0.177.

[62] Calculated as: exp(0.204)-1.

larger for the Botswana-South Africa (23.1%), Malawi-Zambia (26.6%), South Africa-Zambia (32.4%), Botswana-Malawi (35.7%) and Malawi-South Africa (56.4%) borders, respectively.

Figure 7: Average border effects in the SADC region



Looking at the various country groupings, border effects between countries are relatively lower for the SADC countries that are also both members of either the SACU or COMESA. This is in line with studies elsewhere in the literature that have found price dispersion and/or border effects to be lower between countries that participate in a regional trade agreement or currency union (Parsley & Wei, 2002; Bergin & Glick, 2007; Versailles, 2012). The average border effect is lowest between Botswana and Zambia, followed by the SACU countries (Botswana and South Africa), the COMESA countries (Malawi and Zambia) and the average for the remaining SADC country pairs in the sample, respectively. After controlling for distance, the border effect for the SACU countries is 23.1% (this is very close to the 23% estimated by Nchake (2013) for the South Africa-Botswana border). The increase in price dispersion generated by the border effect for the COMESA countries is 26.6%. In comparison, the average effect of crossing a border between the rest of the SADC countries corresponds to a price difference of 36.2%. The distance-equivalent of the border effect is 29.7km for the SACU countries, compared to 99km for COMESA (crossing the Malawi-Zambia border) and an average of 179.7km for the remaining SADC borders.

There are also clear differences in the magnitude of border effects for contiguous and non-contiguous SADC countries. On average, border effects for the contiguous country pairs (Botswana-South Africa, Malawi-Zambia and Botswana-Zambia) are notably lower than those for non-contiguous country pairs (Botswana-Malawi, Malawi-South Africa, South Africa-Zambia). After controlling for distance, the border effect on price dispersion for contiguous countries is 24.1%; compared to 41.1% for district pairs separated by more than one national border. The distance-equivalent of the border effect is around three times larger, on average, for the non-contiguous SADC countries. In a similar study on East Africa, Versailles (2012) also finds lower border effects between neighbouring countries compared to non-contiguous countries separated by more than one national border.

Border effects for perishable products

The border effects described above are estimated as averages across all of the different types of products included in the sample. There may, however, be some differences in product characteristics that influence the way in which distance and borders affect relative prices. This is particularly relevant in the case of perishable products, for which transportation may be time sensitive and require additional refrigeration or specialised packaging. To investigate this, a dummy variable is constructed that is equal to one for all of the perishable products in the sample (bananas, cabbage, onions, oranges, pineapples, potatoes and tomatoes) and zero otherwise. The log distance and border dummy variables are then interacted with the perishable products dummy, meaning that the specification introduced in equation (1) is modified as follows:

$$Q\left(\left|p_{i,j,t} - p_{i,k,t}\right|_n, max\right) = \beta_0 + \beta_1 \ln(dist_{jk,n}) + \beta_2 B_{jk,n} + \beta_3 perishable + \beta_4 D_{jk,n} * perishable + \beta_5 B_{jk,n} * perishable + \gamma_t + \gamma_i + \sum_J \delta_{J-J} W_{jk,n} + \epsilon_n \quad (2)$$

where $perishable$ is equal to one if a product is perishable and zero otherwise; and the remaining variables are defined as before. The key coefficients of interest are β_4 and β_5 on the interaction terms, which describe the marginal effects of distance and national borders on price differences between districts for perishable products. The results from the quantile regression estimation in equation (2) are presented in Table 15.

Table 15: Quantile regression estimation of average and individual border effects for perishable products

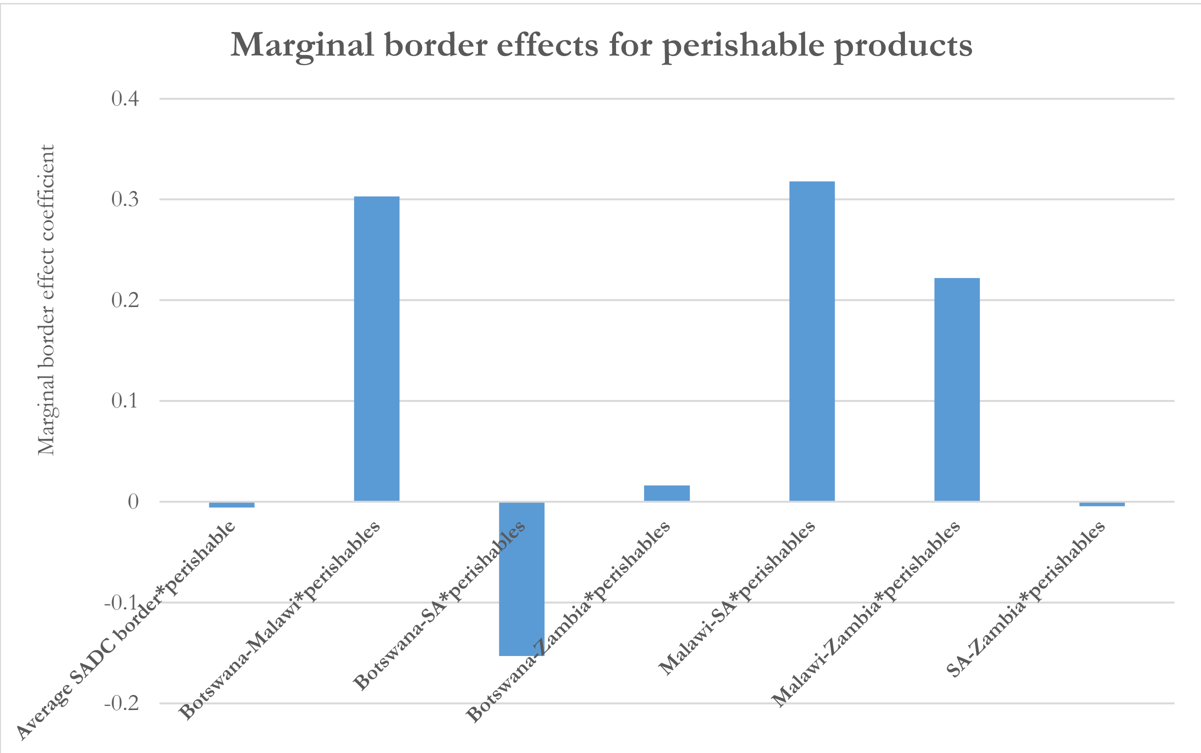
	(1)	(2)
log distance	0.0868***	0.0800***
	(0.00445)	(0.00484)
border	0.379***	
	(0.00966)	
perishable	-0.0857***	0.0696**
	(0.0313)	(0.0303)
log distance*perishable	0.0633***	0.0362***
	(0.00529)	(0.00511)
border*perishable	-0.00564	
	(0.0105)	
Botswana-Malawi border		0.411***
		(0.0170)
Botswana-SA border		0.355***
		(0.0106)
Botswana-Zambia border		0.397***
		(0.0135)
Malawi-SA border		0.444***
		(0.0197)
Malawi-Zambia border		0.345***
		(0.0158)
SA-Zambia border		0.378***
		(0.0153)
Botswana-Malawi*perishable		0.303***
		(0.0177)
Botswana-SA*perishable		-0.153***
		(0.00959)
Botswana-Zambia*perishable		0.0161
		(0.0127)
Malawi-SA*perishable		0.318***
		(0.0228)
Malawi-Zambia*perishable		0.222***
		(0.0243)
SA-Zambia*perishable		-0.00440
		(0.0172)
Botswana	0.215***	0.203***
	(0.0118)	(0.0112)
Malawi	0.180***	0.175***
	(0.0227)	(0.0201)
Zambia	0.198***	0.194***
	(0.0101)	(0.00956)
constant	-0.446***	-0.401***
	(0.0302)	(0.0320)
Observations	92,631	92,631
R-squared	0.393	0.419

Notes: The dependent variable is the maximum of the absolute value of the log price difference between districts *j* and *k* in year *t* for bin *n* (using 500 distance bins). All regressions are estimated with year and product fixed effects as well as country-specific dummies for 'Botswana-Botswana', 'Malawi-Malawi' and 'Zambia-Zambia' district pairs (with the 'South Africa-South Africa' dummy omitted). Robust standard errors (reported in parentheses) are clustered by district pair. Significance at the 10 percent, 5 percent and 1 percent levels is denoted by *, ** and *** respectively.

In columns (1) and (2), the positive and significant coefficients on the log distance-perishable interaction variable indicate that the marginal effect of distance on price dispersion between districts is greater in the case of the perishable products. This is anticipated given the likely higher costs of transportation associated with perishable products. Notably, however, the coefficient on the border-perishable variable in column (1) is insignificant, suggesting that the average border effect on price dispersion in the SADC region is no greater in the case of perishable products

In the individual border effects specification in column (2), however, the marginal effects of crossing the Botswana-Malawi, Malawi-South Africa and Malawi-Zambia borders on price dispersion are larger for perishable products (reproduced in Figure 8). In the case of the Botswana-Malawi and Malawi-South Africa border effects, this may be due to the fact that transportation between these countries involves multiple border crossings, suggesting that the marginal effect of crossing more than one border between these countries is accentuated in the case of perishable products. In contrast, however, the coefficient on the Botswana-South Africa-perishable interaction term is negative and statistically significant, pointing to a smaller marginal effect.

Figure 8: Plot of marginal border effects for perishable products

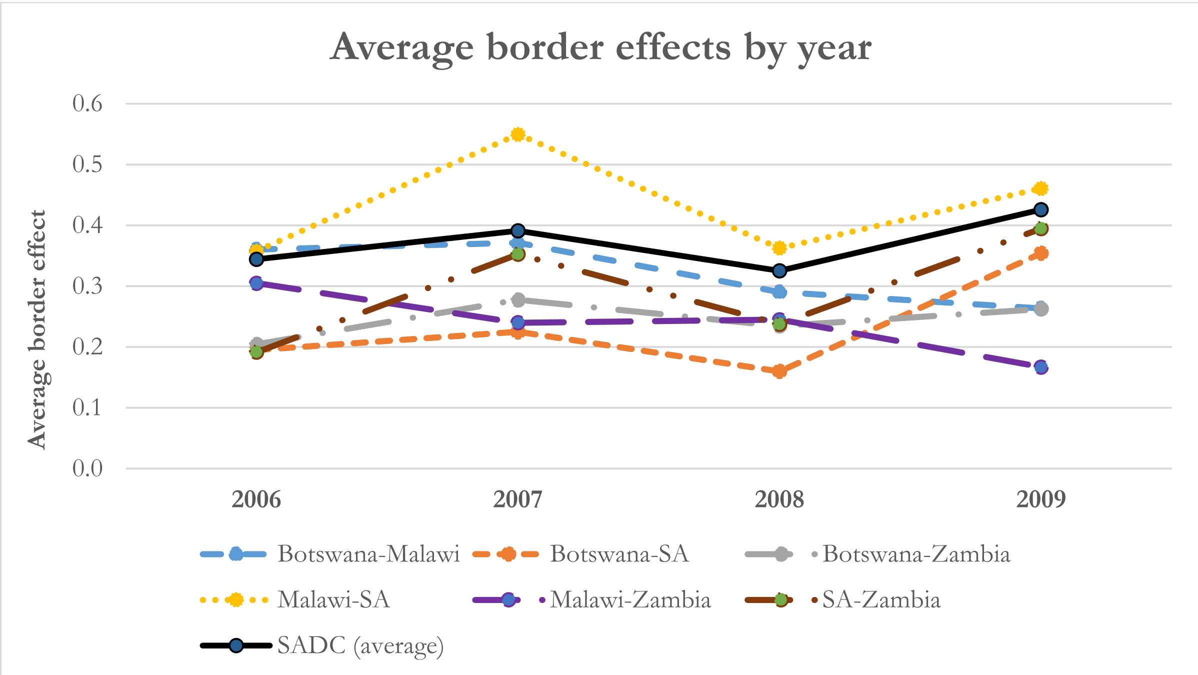


4.3.2 Trends in border effects over time in the SADC region

The evidence presented above, which is based on pooling data for the 2006-2009 period, indicates that average border effects on product price dispersion in the SADC region over the sample period as a whole are comparatively large. What is not clear from this initial analysis, however, is whether these effects have changed significantly over time. This is particularly interesting in the SADC context given that the years from 2006 to 2009 coincided with a period of marked acceleration in the phase down of tariffs on intra-SADC trade. This was especially true in the lead up to the formal launch of the SADC FTA on 17 August 2008.[63]

Changes across the four-year period in the magnitudes of the coefficients on the SADC border dummy as well as the specific border dummies for bilateral country pairs are examined by estimating equation (1) in separate regressions each with the sample restricted to observations for a single year. The regression results are presented in Tables C.1 and C.2 in Appendix C. Using the same procedure as before, the estimated border coefficients in Table C.2 are used to compute average border effects for each specific border and year. The resulting values are plotted in Figure 9, along with the average border effect for the SADC region as a whole.

Figure 9: Average border effects in the SADC region by year, 2006-2009



[63] By the time of the introduction of the SADC FTA in 2008 as part of the implementation of the SADC Protocol on Trade, customs duties had been eliminated on 85% of intra-SADC trade.

The estimated coefficient on the SADC border dummy increased marginally between 2006 and 2009. There were also increases in the computed average border effects for the Botswana-South Africa, Botswana-Zambia, Malawi-South Africa and South Africa-Zambia borders between 2006 and 2009; although these increases were also relatively small. In contrast, the magnitudes of the Botswana-Malawi and Malawi-Zambia border effects declined over this period. Overall, however, there was relatively little variation in the average and individual border effects in the SADC region between 2006 and 2009, despite the trade reform implemented during this period.

4.4 Robustness Test

The focus in this section is on testing the sensitivity of the border effects estimated above to the methodology employed. This is done by re-estimating the border effects using an alternative production-consumption pair approach that uses only spatial price data for production and consumption district pairs. This approach provides another way to estimate distance and border-related trade costs in the presence of inequality constraints (see Chapter 3) and, thus, presents an alternative to the quantile regression methodology.

The production-consumption pair regression is estimated using the following basic specification:

$$|p_{i,p,t} - p_{i,c,t}| = \beta_0 + \beta_1 D_{pc} + \beta_2 B_{pc} + \gamma_t + \gamma_i + \epsilon_n \quad (3)$$

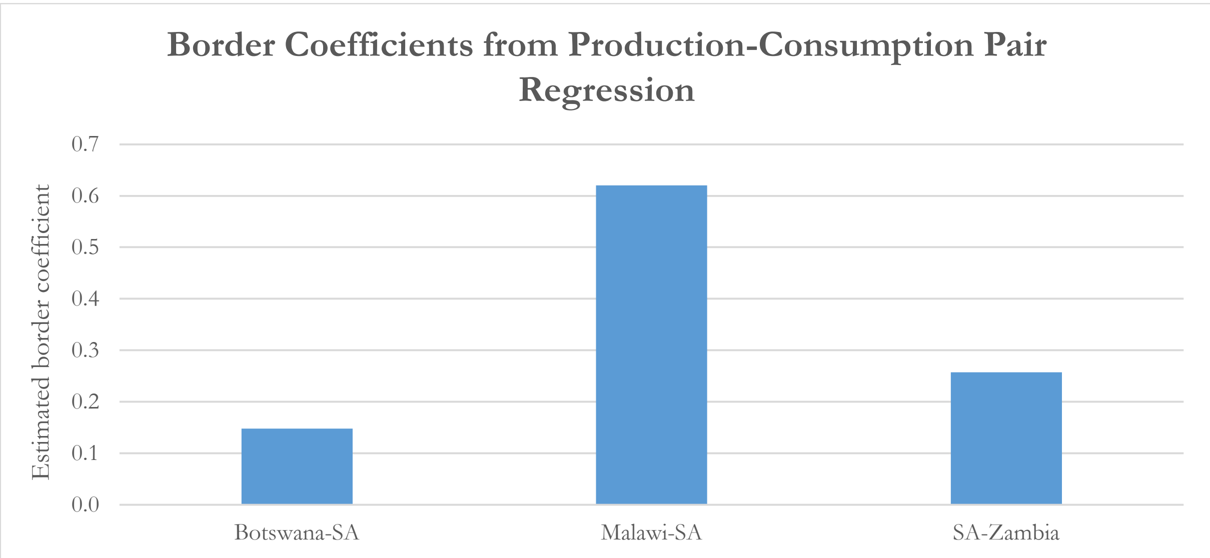
where $p_{i,p,t}$ is the retail price of good *i* in the production location *p* (Pretoria in the regressions) at time *t* and $p_{i,c,t}$ is the corresponding retail price in the consumption location *c*; D_{pc} is the log distance from the origin location to the consumption location; B_{pc} is a dummy variable equal to one if the origin and consumption locations are separated by a national border and zero otherwise; γ_t and γ_i are year and product fixed effects, respectively; and ϵ_n is the regression error term.

In contrast to Chapter 3, the production-consumption pair regression in equation (3) is estimated using the full sample of 24 products (rather than the subset of nine products for which more than 50% of total imports in Botswana, Malawi and Zambia are sourced from South Africa). It is assumed here that the 24 products sold in the consumption locations all originate from South Africa. The accuracy of the production-consumption pair results are conditional on this assumption being correct. Hence, the comparison of the border effects computed from the quantile regression and production-consumption pair models is only relevant in the event that the assumption holds. Nevertheless, even in cases where the sampled products do not actually originate from South Africa, the products for which the prices are observed in the dataset are likely to be close substitutes for the products coming from South Africa and compete directly with the South African products in the consuming countries. These forces should ensure convergence in retail prices between the South African products and their competing substitutes.

By restricting the sample to products originating from South Africa in the production-consumption pairs approach, it is only possible to examine border effects for country pairs involving South Africa. As a result, the comparison with the quantile regression estimates is confined to the effects of the Botswana-South Africa, Malawi-South Africa and South Africa-Zambia borders on relative prices.

The full regression outputs from the production-consumption pair regressions are presented in Table C.3 in Appendix C, and the border coefficients are reproduced in Figure 10. The coefficients on the border dummies estimated from the production-consumption pair and the average border effects computed for the quantile regressions (reported in Table 14) are directly comparable because, in both cases, the border effects are measured as the additional effect of crossing the border on price dispersion relative to internal price dispersion within South Africa (the omitted category).[64,65]

Figure 10: Estimated border coefficients from production-consumption pair regression



[64] In the case of the production-consumption pair regressions, observations for within-country district pairs are only included for South Africa since production of all products is assumed to take place in Pretoria.

[65] The production-consumption pair regressions were also estimated with the inclusion of a log distance squared variable to account for possible non-linearity in the relationship between price gaps and distance. The estimated border coefficients are generally quite similar when this variable is included, although the Botswana-South Africa border coefficient is marginally larger, while the opposite is true for the non-contiguous countries (Malawi-South Africa and South Africa-Zambia).

Based on the production-consumption pair estimates, crossing the Botswana-South Africa border increases price dispersion by 16%.[66] This is smaller than the Botswana-South Africa border effect of 23.1% estimated from the quantile regressions. The estimates from the two techniques for the South Africa-Zambia border effect on price dispersion are similar: the effect of the border in increasing price dispersion between the two countries is 29.3%[67] when estimated from the production-consumption pair specification compared to 32.4% when calculated from the quantile regression. This comparison suggests that the estimates of the Botswana-South Africa and South Africa-Zambia border effects are not particularly sensitive to the estimation technique.

In contrast, the level of the Malawi-South Africa border effect is very sensitive to changes in the estimation technique. The border effect is substantially larger when estimated using the production-consumption pair regression (85.9%)[68] compared to the border effect when calculated from the quantile regression (56.4%).

4.5 Conclusion

This chapter has sought to obtain precise estimates of the impact of border-related trade costs on product market integration in the SADC region. The existing price-based studies of border effects in Africa all employ variants of the standard empirical methodology and are, thus, subject to the sample selection biases discussed in Chapter 3. This chapter makes an important contribution to the literature by accounting for general sample selection and product sample selection effects using quantile regressions and, thereby, providing more accurate estimates of border effects in Africa. Importantly, however, the empirical methodology used in this chapter does not address the potential issue of distance sample selection bias or the issue of omitted variable bias in the presence of unobserved heterogeneity across locations within narrowly defined products. As a result, the border coefficients reported in this chapter may be *overestimated.*

The estimation results presented in this chapter reveal significant variation in border effects on price dispersion across the individual country pairs. The border effect is found to be smallest for Botswana and Zambia, where crossing the border increases price dispersion between the two countries by 22.6% relative to the average level of price dispersion within the two countries. The

[66] Calculated as: exp(0.148)-1.
[67] Calculated as: exp(0.257)-1.
[68] Calculated as: exp(0.62)-1.

border effects on price dispersion are consecutively larger for the Botswana-South Africa (23.1%), Malawi-Zambia (26.6%), South Africa-Zambia (32.4%), Botswana-Malawi (35.7%) and Malawi-South Africa (56.4%) borders, respectively.

By comparing average border effects across different country groupings, the chapter also provides important insights into the nature of border effects in the region. The quantile regression estimates show that border effects tend to be comparatively lower between SADC countries that share membership in additional regional trade agreements. The borders between the SACU and COMESA countries generate smaller increases in price dispersion compared to the average effect of crossing a border between the rest of the SADC countries. Furthermore, the distance-equivalent of the border effect is 29.7km for the SACU countries, compared to 99km for COMESA (crossing the Malawi-Zambia border) and an average of 179.7km for the remaining SADC borders.

When using the quantile regression approach, clear differences are also found in the magnitude of border effects for contiguous and non-contiguous SADC countries. The additional price dispersion generated by crossing the border between contiguous countries is 24.1%; compared to 41.1% for countries separated by more than one national border. This points to the presence of incremental border effects on relative prices as products are traded across multiple borders.

However, robustness results estimated from an alternative production-consumption pair model do suggest that the levels of the individual border effects in the region vary somewhat according to the estimation technique. Nevertheless, the ranking of the border effects across different regional trade agreements and for contiguous versus non-contiguous countries is robust across different specifications.

The chapter examines potential product heterogeneity by comparing border effects on relative prices for perishable and non-perishable products. While it is found that the average border effect on price dispersion in the SADC region is no greater in the case of perishable products, the marginal effects of crossing the Botswana-Malawi, Malawi-South Africa and Malawi-Zambia borders on price dispersion are larger for perishable products. The results for the Botswana-Malawi and Malawi-South Africa borders suggest that incremental border effects are more substantial for perishable products.

Another important insight that can be drawn from the results presented in this chapter is that there was generally little change in the magnitude of border effects in the SADC region between 2006 and 2009. In fact, the average border effect on cross-country price dispersion for the SADC region as well as the individual border effects generally increased by small margins (with the exception of the Malawi-Zambia border) between 2006 and 2009. The main insight that can be drawn from this is that border effects in the region generally did not fall, despite the liberalization of tariffs on intra-SADC trade over this period.

The following chapter seeks to examine the relationship between tariffs and product market integration in the SADC region. The main emphasis is on unpacking the tariff contribution to the border effect. This is achieved by focusing on how Zambian preferential tariffs affect cross-border product price dispersion between South Africa and Zambia.

5. Tariffs, Relative Prices and Border Effects in the Southern African Development Community

5.1 Introduction

For some time, the promotion of regionalism and market integration has formed a central thrust of the strategic efforts of African governments to drive economic and political unity on the continent. The desire to boost intra-regional trade represents a key objective of these efforts, with greater regional integration in trade expected to boost competitiveness and raise overall economic efficiency in Africa. Importantly within this context, tariff policies have been identified as a key instrument through which governments can influence trade and product market integration (Behar & Edwards, 2011). In SADC, since coming into effect in 2000 the SADC Protocol on Trade has served as the principal mechanism driving economic integration and tariff liberalization.

Despite the efforts to promote integration within SADC, the analysis presented in Chapters 2 through 4 shows that, even after accounting for distance-related trade costs, product prices remain dispersed in the region. In particular, comparatively large border effects are present between South Africa, Botswana, Malawi and Zambia – reflected in product price deviations across districts *between* these countries that far exceed the equivalent price differences between districts *within* the four countries. The presence of tariffs on intra-SADC trade may play some role in contributing to the large border effects observed in the data. In this regard, the links between trade policy and the influence of borders on international relative prices in the SADC region may be significant.

The main objective of this chapter is to examine the relationship between tariffs and product market integration in the SADC region. The emphasis is on investigating whether or not the presence of tariffs on intra-SADC trade accounts for some of the observed dispersion in product prices generated by crossing national borders. Owing to a lack of variation over time in Malawi's tariff phase down offers to other SADC countries since 2004,[69] the empirical analysis in this chapter is confined to three SADC countries: South Africa, Botswana and Zambia. Furthermore, given that trade between South Africa and Botswana is not subject to tariffs as both countries are members of the SACU, the focus is on how Zambian tariffs on South African imports affect prices in the Zambian market.

[69] Malawi's tariff liberalization with respect to SADC stalled between 2004 and 2010, with the country's tariffs on SADC imports remaining unchanged over this period.

Most disaggregated price-based studies in the existing literature have focused on the pass-through effects of tariffs onto domestic prices or on the relationship between tariffs and distance-related trade costs, and the majority have been confined to an analysis of these relationships *within* countries (Nicita, 2009; Cherkaoui, 2011; De Loecker et al., 2012; Bas & Strauss-Kahn, 2013). Few studies have examined the relationship between tariffs and border-related trade costs or *international* trade costs more generally, and no studies have looked directly at the potential contribution of tariffs in generating border effects on international relative prices. This chapter makes an important contribution in extending the literature in this direction by unpacking the tariff contribution to the border effect between South Africa and Zambia.

The empirical analysis in this chapter uses disaggregated price level data for narrowly defined products. This facilitates more accurate measurement of the border effect and its tariff component compared to studies based on aggregate price indices, which suffer from a number of compositional issues and aggregation problems that may bias the border effect estimates (see Chapters 2 and 3).

The analysis in this chapter is also differentiated from much of the existing research on tariffs in the product market integration literature in that it accounts for the spatial relationship between production and consumption. Existing methodologies in the international literature that are based on gravity models are likely to suffer from misspecification bias because they exclude relevant information on locations' distance from their production sites or on costs at the source of production (Anderson et al., 2013). This can lead to underestimation of the impact of transport costs on price deviations across markets (Anderson et al., 2013; Kano et al., 2013), in turn affecting the accuracy of the tariff and border effect estimates. This chapter exploits the fact that trade flows between South Africa and Zambia are dominated by South African exports to the Zambian market – a reality that reflects the broader patterns of intra-regional trade within Southern Africa – by assuming that there is a unidirectional flow of trade in products between South Africa (the production source) and Zambia (the consumption location). This, in turn, makes it possible to isolate the relationship between preferential tariffs applied to South African products imported into Zambia and the border effect between the two countries.

The preferential tariff data used in this chapter varies by product and time, thus lending itself to a disaggregated product-level analysis of the tariff contribution to the border effect. Many studies in the literature use average tariff rates across products (see, for instance, Bergin & Glick, 2007). These studies are unable to exploit variation in tariffs at the product level, despite the generally high level of tariff variation across products at the country level. They are also likely to suffer from an endogeneity problem arising from the difficulty of excluding product-specific effects influencing tariffs when using aggregate tariff data. For example, a high tariff may be imposed on a high priced product that is uncompetitive in the domestic market. In this case, it is not the effect of the tariff that is leading to a higher domestic price for that product, but rather the high price that results in a higher tariff. The analysis in this chapter is able to deal with the causality concerns that affect other studies by using data on prior agreed tariff phase downs.

Similarly, existing studies that look at the pass-through effects of tariffs on prices, as well as those that examine the relationship between tariffs and product market integration either directly or indirectly, tend to assume homogeneity in these effects across products. By using product-level data on the share of South African imports in Zambia's total imports, the analysis in this chapter accounts for the possibility that the effect of national borders and tariffs on cross-country relative prices may vary according to the intensity with which the products are traded across the border.

The tariff contribution to the South Africa-Zambia border effect is analysed in this chapter using data on retail product prices and tariff rates over the period from 2002 to 2009. This period covers the initial years when Zambia's preferential tariff rates for South African imports remained relatively high due to the back-loading of tariff phase down schedules, as well as the period of accelerated tariff reductions in the lead up to the establishment of the SADC FTA in 2008. The analysis thus covers a period in which there has been significant variation in the preferential tariff rates applied to South African imports into the Zambian market, while both Zambia's and South Africa's MFN tariffs have remained largely unchanged.

The SADC region is a particularly interesting setting in which to examine the relationship between tariffs and border effects. Despite the establishment of the FTA in the region, the complex and restrictive rules of origin and import sourcing requirements contained within the SADC Protocol on Trade hamper trade, compromise the potential benefits of preferential tariff reductions and raise transaction costs (Brenton et al., 2005). In this context, the effect of tariff liberalization is uncertain and may vary from country to country. This is, however, less likely to be an issue in the case of Zambian imports of South African products because South Africa is generally able to meet the rules of origin requirements to qualify for preferential tariffs owing to its industrial base. This provides further motivation for focusing on the tariff contribution to the South Africa-Zambia border effect. Importantly, however, the results cannot be extended to cases where rules of origin are binding and prevent firms from responding to lower bilateral tariffs.

The remainder of the chapter is structured as follows. Section 5.2 outlines the specifics of the tariff reform process that has occurred following the introduction of the SADC Protocol on Trade in 2000. This is followed in section 5.3 by a discussion of the theoretical relationship between tariffs, relative prices and border effects. Section 5.4 describes the key findings in the relevant empirical literature. Thereafter, section 5.5 outlines the main features of the product-level price and tariff data used in the empirical analysis; and section 5.6 describes the conceptual framework and empirical methodology underpinning the analysis. Section 5.7 presents empirical results on the relationship between tariffs, prices and the South Africa-Zambia border effect; and section 5.8 concludes.

5.2 Tariff Reform under the SADC Protocol on Trade

By acceding to the SADC Protocol on Trade, the 12 signatories to the Protocol were required to submit an instrument of implementation in the form of a tariff phase down schedule – with annual tariff phase downs to be implemented each year on the 1st of January from 2000 onwards. Under the terms of the Protocol and in recognition of asymmetries in the levels of development of individual countries, the participating SADC Member States submitted differentiated tariff phase down offers for South Africa and the remaining Member States. At the same time, the SACU grouping committed to eliminate import duties at a faster rate to that of the other Member States (Flatters, 2010). By countries committing to a schedule for tariff reform at the outset through a categorization of products devised at the SADC level, the liberalization of tariffs on intra-SADC trade occurred according to a prior agreed phasing down of tariffs. This is important for the empirical estimation of tariff effects because it makes it possible to exclude product or country specific effects influencing tariffs. This allays potential endogeneity concerns.

In practical terms, the application of the tariff reforms enshrined in the SADC Protocol on Trade was implemented in stages, with products placed in specific categories for the phasing down of tariffs. Certain products that already attracted low or zero tariffs were placed in Category A and earmarked for immediate liberalization. This meant that tariff lines on these products were to be reduced to zero as from the date of implementation of the Protocol in 2000.

In turn, tariffs on a second category of products (Category B) – which were identified as important sources of customs revenue – were to be liberalized gradually over a period of eight years. In recognition of the differences in levels of development across SADC countries, there was significant heterogeneity in the rates at which the participating SADC Member States committed to reducing their tariff lines on Category B products. For their part, the SACU Member States (who have a common external tariff) agreed to front-load their tariff reform, committing to reduce tariff lines on Category B products in equal annual instalments from 2000 to 2008. In turn, Mauritius and Zimbabwe agreed to reduce tariff lines on Category B products by equal instalments from year four to year eight of the implementation period. In contrast, the tariff phase down schedules for Category B products in Malawi, Mozambique, Tanzania and Zambia were back-loaded, with tariff lines on these products to be reduced in equal instalments from years six to eight.

Alongside these differences in the speed of tariff reform, the tariff reductions were also implemented asymmetrically among the SADC Member States. In this respect, the countries outside of the SACU made two different preferential tariff offers: a SACU offer (to South Africa, Botswana, Lesotho, Namibia and Swaziland) and a differentiated offer to the rest of SADC (Mudzonga, 2008). The SACU offer comprised a more gradual phasing down of tariffs, meaning that the rest of the SADC Member States liberalized tariffs more rapidly among themselves compared to their liberalization with respect to SACU (Mudzonga, 2008)

In the case of both the SACU and the differentiated preferential tariff offers, the schedules for tariff phase downs in the SADC Protocol on Trade were heavily back-loaded, with much of the reduction in tariffs occurring in the latter stages of the implementation period. Consequently, initially only modest phase downs were implemented between 2000 and 2004. The period thereafter, however, saw an acceleration of tariff reform in the lead up to the establishment of a SADC FTA in August 2008 – reflected in a significant increase in the percentage of duty free tariff lines in the region between 2005 and 2008 (with the most substantial phase downs occurring in 2007 and 2008). The introduction of the SADC FTA saw 85% of intra-SADC trade afforded duty free treatment.

Thereafter, tariffs on products regarded to be of economic importance to SADC Member States (Category C) were to be eliminated between 2008 and 2012. The designation of products in

Category C was limited to a maximum of 15% of each Member State's intra-SADC merchandise trade (Kalenga, 2009). A small set of products such as firearms and munitions were placed in Category E and excluded from tariff liberalization under the Protocol. Products in this category constitute only a small share of intra-SADC trade (Kalenga, 2009).

Within the region, compliance with the agreed tariff phase downs has generally been high, with the exception of a small number of countries that have either lagged in their implementation (Malawi) or have been granted derogations (Tanzania and Zimbabwe). As of 2011, the SACU Member States had completed their agreed tariff phase down obligations and Mozambique, Tanzania and Zambia submitted notifications for block tariff phase downs to the SADC. In contrast, budgetary considerations forced Malawi to delay the implementation of its tariff phase down commitments, with the country remaining at its 2004 phase down levels as of 2010. As of 2011, Malawi had only complied with 46% of the obligations under its tariff phase down offers to South Africa and the remaining SADC Member States.

Despite the significant progress that has been made in reducing barriers to trade between SADC Member States through the tariff reform process, there remains scope for further reform, particularly in the case of tariffs on intermediate inputs and rules of origin that constrain regional trade flows (Erasmus et al., 2004; Brenton et al., 2005; Behar & Edwards, 2011). The complex product-specific rules of origin and input sourcing requirements contained in the SADC Protocol on Trade are particularly problematic. In many cases, these rules of origin requirements only serve to shield existing industries in the region from greater intra-regional competition by affecting the ability of producers in the region to benefit from preferential tariff reductions and trade with other SADC countries (Brenton et al., 2005). In SADC countries where producers are unable to meet the rules of origin requirements to qualify for trade preferences, the effect of the tariff phase downs is likely to be very limited. These factors may continue to inhibit product market integration, despite the fact that the SADC FTA has largely eliminated tariffs on intra-SADC trade (Behar & Edwards, 2011).

5.3 Theoretical Framework: Tariffs, Relative Prices and Border Effects

Of relevance to price-based studies of product market integration is the relationship between tariffs, relative prices and border effects. In theory, tariff policies can influence product prices in a particular market through two channels. First, tariffs affect the price of traded goods directly. By imposing a tax at the border, tariffs raise the retail price of imported products in the domestic market (Nicita, 2009). As a result, tariffs have a direct impact in generating discontinuities in relative prices at the border (Rossi-Hansberg, 2005). Specifically, when tariff $\tau_{i,c,t}$ is applied on imports of product *i* into country *c* at time *t,* the price of that product in the importing country will be given by:

$$p_{i,c,t} = p^{*}_{i,p,t}\left(1 + \tau_{i,c,t}\right) \tag{1}$$

In this simple exposition, once accounting for the cost of transporting product *i* between the two countries, the price of the exported product in country *p* will differ from its price in the importing country *c* by the size of the tariff, $\tau_{i,c,t}$.

In practice, however, the tariff effect on prices in the importing country may vary for several reasons. First, the pass-through effects of tariffs on domestic retail prices *within* a country may vary significantly across regions. Any variation in retail margins, transport costs, local non-traded input prices and mark-ups across regions will mean that the difference between the border prices of traded goods and retail prices will not be uniform across regions within a country (Parsley & Wei, 2007; Nicita, 2009; Atkin & Donaldson, 2014).

Second, the effects of tariffs on domestic retail prices may vary according to the substitutability and importance of an imported product in the domestic consumption of that product. This, in turn, should affect local mark-ups. For example, a lower tariff on a product for which imports constitute a large share of domestic consumption will raise the level of competition faced by local firms from imported varieties in the domestic market. This may induce local firms to reduce their mark-ups for that product and prompt them to charge lower prices in the domestic market. These effects are, however, contingent on the market structure in individual locations, with tariff reductions not necessarily translating into lower retail prices in imperfectly competitive settings where firms enjoy market power. In imperfectly competitive markets, the tariff pass-through is not perfect. In these situations, the effect of a lower tariff may be absorbed entirely by the mark-ups of local retailers, thus having little impact on final retail prices. Alternatively, the imposition of a tariff on imports may be partially absorbed by foreign exporters who respond by lowering their producer price. In this case, the effect of the tariff is not fully passed through to consumer prices in the importing country (Brander & Spencer, 1984; Feenstra, 1989).

Third, tariffs on imported intermediate inputs may also influence final retail prices in domestic markets. Lower tariffs on imported intermediate inputs may prompt firms to alter their output. Alternatively, they may encourage firms to purchase higher quality inputs with a view to producing better quality products for resale at higher prices (Bas & Strauss-Kahn, 2013). Hence, changes in retail prices resulting indirectly from external tariffs may come through variation in product quality or through changes in mark-ups (Bas & Strauss-Kahn, 2013).

Finally, the pass-through effects of tariff changes on domestic retail prices may also vary across different types of products. For example, the pass-through effects of tariffs may be lower for agricultural products. The higher costs associated with transporting agricultural products, coupled with the reality that there are often greater volumes of locally produced agricultural products available in domestic markets compared to foreign varieties, may reduce the influence of tariffs on domestic agricultural prices relative to their impact on other types of products such as manufactures (Cherkaoui, 2011). Alternatively, since agricultural products tend to be more homogenous, they are likely to face more competition in the domestic market which may, in turn, result in greater pass-through of tariff changes onto domestic prices (Cherkaoui, 2011).

Through both their direct and indirect effects on retail prices in domestic markets, tariffs can have an important influence on product market integration. Retail prices are an increasing function of both tariffs and trade costs, meaning that the presence of high tariffs can drive a wedge between the domestic and foreign prices of specific goods and impose prohibitive costs to arbitrage, resulting in cross-border dispersion in product prices. In such instances, high tariffs can serve as a barrier to product market integration by constraining the integration of prices across markets. Alternatively, where tariff liberalization successfully reduces barriers to trade, the forces of arbitrage should serve to eliminate price differentials for consumer goods across countries, facilitating greater product price integration across markets. This should be reflected in lower border effects on international relative prices.

5.4 Empirical Literature

5.4.1 Gravity-style estimates of tariff effects based on the volume of trade flows between countries

A number of studies have examined the relationship between tariffs, trade flows and product market integration using gravity equations. These studies form part of a larger literature inferring trade costs from the volume of trade flows between countries, and estimate gravity equations with tariffs included in the estimations among other directly observed barriers to trade (Anderson & Van Wincoop, 2004). Most studies have focused on developed countries and generally find that tariffs act as a significant barrier to imports, thereby impeding trade flows between countries (Harrigan, 1993; Hummels, 2001; Chevassus-Lozza et al., 2007). In line with this, Head and Ries (2001) show that tariff barriers tend to distort consumption patterns in favour of domestically produced goods over foreign varieties. On the other hand, reductions in tariffs can play an important role in boosting trade flows. Baier and Bergstrand (2001), for instance, estimate that tariff rate reductions accounted for about 25% of the growth in trade observed among a sample of OECD countries between the late 1950s and the late 1980s.

While not discounting the role of tariffs, some studies in the trade volumes literature have emphasised the primacy of other elements – such as NTBs and beyond the border barriers to trade – in constraining bilateral trade flows between countries (Schiff & Winters, 2003; Olper & Raimondi, 2008; Koczan & Plekhanov, 2013). Their results generally imply that policy measures such as improvements in trade-related infrastructure and regulatory reforms may have a greater impact in boosting trade flows compared to reductions in tariff barriers.

5.4.2 Price-based studies

In comparison to the trade volumes literature, disaggregated price-based studies of the relationship between tariffs and product market integration are less prevalent. This is despite the benefits of using product price data rather than information on trade flows as the basis for measuring product market integration. As explained in Chapters 1 and 3, estimates of product market integration based on the volume of trade between countries may be misleading since variation in the level of trade flows can be consistent with a single trade cost and the same degree of product market integration.

Research on regional trade agreements and product market integration

One arm of the price-based literature examines the relationship between tariffs and market integration *indirectly* by using product price level data to compare border effects and relative prices for countries that share membership in a FTA or customs union with those between countries that do not belong to a particular regional trade grouping. In theory, trade or monetary agreements that reduce barriers to trade and arbitrage (by, for example, reducing tariffs or exchange rate volatility) are expected to allow the forces of arbitrage to eliminate price differentials for consumer goods across countries (Parsley & Wei, 2002).

The evidence on the impact of trade agreements on product market integration drawn from studies using actual price level data generally supports the theory.[70] Using a panel of prices for 95 traded goods across 83 cities in 69 different countries for the period from 1990 to 2000, Parsley and Wei (2002) find that price dispersion is significantly lower between countries that share membership in the EU, European Free Trade Association, NAFTA or MERCOSUR trade blocs. They conclude that trade blocs promote goods market integration more effectively than unilateral trade liberalization. Similarly, using annual price level data for 101 tradable goods across 108 cities in 70 countries from 1990 to 2005, Bergin and Glick (2007) observe that price dispersion declines when countries participate in a regional trade agreement.[71]

As Anderson and van Wincoop (2004: 722) note, however, while it is relatively well documented in the literature that FTAs and customs unions reduce trade barriers and even cross-border product price dispersion, "it is less clear what elements of these trade agreements play a role (tariffs, NTBs, or regulatory issues)." In this respect, the precise role played by tariffs in generating cross-border price dispersion is comparatively understudied.

Research on the pass-through effects of tariffs onto prices

Much of the existing literature that has looked directly at the role of tariffs is focused on the price transmission and welfare aspects of tariff liberalization. For instance, several studies have examined the pass-through effects of changes in tariffs on local, import or export prices using different methodologies (Nicita, 2009; Cherkaoui, 2011; De Loecker et al., 2012; Bas & Strauss-Kahn, 2013).

---

[70] This is not, however, the case for all price-based studies. For instance, Engel and Rogers (1996) examine border effects between the United States and Canada using CPI data. They segment their data into sub-samples in order to separate the periods before and after the introduction of the United States-Canadian FTA, and find that the size of the border effect – reflected in the magnitude of the estimated border coefficient – was actually larger in the period after the FTA came into effect. Confronted with this seemingly counterintuitive finding, the authors raise the possibility that informal trade barriers may have contributed to the observed price dispersion.

[71] Bergin and Glick (2007) also find that that cross-country price dispersion declines when countries participate in a currency union. This is backed by similar findings in Parsley and Wei (2002) and Foad (2004) for the EU, who also show that currency union arrangements facilitate deeper product market integration between participating countries. In contrast, however, a number of studies find little evidence that the introduction of the euro has reduced price dispersion across EU member states (Lutz, 2000; Rogers, 2002; Engel & Rogers, 2004). Furthermore, looking at Africa, Parsley and Wei (2002) find that product markets within the CFA zone are "not very integrated", despite the fact that the CFA countries share a common currency.

A key focus in the tariff pass-through literature is on the extent to which movements in border prices arising from tariff changes between importing and exporting countries are passed-through to domestic prices or absorbed through changes in exporter mark-ups (Nicita, 2004). A number of studies find evidence of the pro-competitive effects of trade liberalization on domestic prices in developing countries (see, for instance, Cherkaoui (2011) for Morocco; De Loecker et al. (2012) for India; and Bas & Strauss-Kahn (2013) for China). De Loecker et al. (2012) show that not only do lower tariffs induce Indian firms to reduce their mark-ups through the pro-competitive effect, but they also reduce the marginal costs faced by these firms. However, the authors observe that firms take advantage of the lower marginal costs to raise their mark-ups, meaning that the cost advantage of the input tariff reductions is not fully passed on to consumers through lower prices.

Nicita (2009) measures the effect of tariff liberalization on domestic prices in Mexico. He shows that the liberalization of tariffs in Mexico following the commencement of NAFTA negotiations in the 1990s reduced domestic consumer prices for both agricultural and manufacturing products, but these effects were not uniform across Mexican states. Using regression analysis, the author examines the determinants of the market price of a particular good in a specific region. His explanatory variables include distance (the shortest driving distance between Mexican states) as a proxy for trade costs, a good-specific tariff variable, and an interaction between distance and the tariff variable for each good. The latter is included in order to isolate the marginal effect of tariff movements on the market price of a particular good in a specific region.

By doing so, Nicita (2009) examines both the change in domestic prices induced by a change in tariffs as well as the effect of trade costs on the pass-through of tariff changes to local prices. He reports a positive and significant coefficient on the tariff variable, suggesting that local prices increase with rising tariffs. He finds that the pass-through rate of tariffs onto local prices is higher for agricultural products (33 percent, on average, for the country as a whole) in comparison to their manufacturing (27 percent) counterparts.[72] In addition, he shows that in the case of manufactured products, the pass-through rate is lower for regions located further from the United States border; with regions located closer to the border found to be more exposed to the effects of changes in manufacturing tariffs.[73] Importantly, Nicita (2009) only focuses on distance-related

[72] Campa and Goldberg (2002) report larger pass-through rates for a range of developed economies.
[73] Nicita (2009: 23) explains that: "Taking into account regional differences the tariff pass-through at the border is about 70% for manufacturing. The tariff pass-through declines to about 40% at 1000km and to about 20% at 2000km from the border."

trade costs *within* Mexico, and does not examine the interaction between tariffs and border-related trade costs or *international* trade costs more generally.

Parsley and Wei (2007) examine the impact of tariffs within the context of the role of traded goods prices in real exchange rate movements. Using price observations that match the prices of Big Macs to the prices of their individual ingredients for 34 countries spanning 13 years from 1990 to 2002, the authors employ a systematic panel regression approach to examine the relative importance of deviations in the international prices of traded goods in explaining real exchange rate movements. Interestingly, the authors find that tariffs – measured as the sum of mean tariff rates between country pairs – have a statistically significant impact in reducing the influence that deviations in traded goods prices have on real exchange rates. They attribute this to the reality that the presence of higher tariffs reduces the scope for arbitrage.

Bas and Strauss-Kahn (2013) focus on export prices and employ a difference-in-difference approach to exploit variation in recent changes in input tariffs between 'ordinary' trade regime firms and 'processing trade regime firms in China – with the latter exempted from paying tariffs for at least 30 years. They show that reductions in Chinese tariffs on imported inputs between 2000 and 2006 – which followed China's accession to the World Trade Organization (WTO) and benefited firms under the ordinary trade regime – actually led to these firms raising their export prices. According to the authors, this is because these firms were able to upgrade their inputs at lower cost following the tariff liberalization; which, in turn, enabled them to upgrade the quality of their exported products. In contrast, the firms that did not benefit from lower imported input costs via the tariff reduction (firms that fell under the processing trade regime) reduced their export prices. The authors attribute the latter to a pro-competitive effect arising from the loss of the cost advantage previously enjoyed by processing firms. Following the tariff reductions, the Chinese processing firms faced more stringent competition in export markets from their ordinary trade regime counterparts and, as a result, reduced their mark-ups.

Research on the relationship between tariffs and border effects

Very few price-based studies have looked directly at the potential role played by tariffs in generating border effects that raise *cross-country* product price dispersion. One notable exception is Bergin and Glick (2007). In their multi-country study, Bergin and Glick (2007) model price dispersion between cities as a function of a number of trade friction determinants, including distance, national borders, language differences, tariff barriers and exchange rate volatility. They find that price dispersion increases with tariffs. Importantly, however, the authors measure tariff barriers as the simple sum of the average tariff rates in both countries for cross-country city pairs.[74] As a result, they are not able to exploit variation in tariffs at the product level. Allowing tariffs to vary by product is likely to be important given that there is generally a high level of variation in tariffs across goods at the individual country level (Anderson & van Wincoop, 2004). Furthermore, the authors do not interrogate the marginal effect of tariffs on the border effect.

5.4.3 Price-based studies of the relationship between tariffs and product market integration in Africa

Focusing on Africa, Versailles (2012) examines the effect of tariffs on product market integration *indirectly* by focusing on the impact of the introduction of the EAC Customs Union in 2005. He finds some evidence of improved market integration following the advent of the Customs Union in the form of a reduced border effect between 2004 and 2007, but only between Kenya and Uganda.

Similarly, using price data for three agricultural commodities (maize, rice and sorghum), Brenton et al. (2014) find lower border effects between Central and East African countries that belong to FTAs. Specifically, they show that average border effects are markedly lower between countries that share membership in the EAC (6.2%) or SADC (11.8%) compared to the average between other countries in their Central and East African sample (33.2%).

Edwards and Rankin (2012) analyse regional product market integration across 13 African cities in 12 countries using product level retail price data. They find that price dispersion between countries declines as each country imposes lower average MFN rates. Their results imply that trade reform that lowers average MFN tariff rates can contribute towards reducing dispersion in relative prices between countries and, thereby, promote product market integration. However, the authors do not take into account the role of preferential tariffs.

Only Mudenda (2013) has looked directly at the relationship between tariffs and border effects in Africa. He examines the effect of tariff reforms on domestic price dispersion *within* Zambia. To do so, he uses data on product-specific import weighted tariff rates. He finds that tariffs exert a

[74] This variable is set to zero when the cities are located in the same country or when the two countries share membership in a FTA.

significant negative effect on price dispersion in Zambia, and shows that the tariff effect is stronger than that of distance. Specifically, he reports that a 10% decrease in tariffs reduces price dispersion between Zambian cities by up to 1.9% (compared to 0.2% in the case of an equivalent reduction in the distance separating cities). Interestingly, he also finds some weak evidence that tariff liberalization is a significant source of cross-border price integration for cities located at the border (the coefficient on the tariff-port of entry interaction term is significant at the 10% level) in Zambia compared to adjacent and non-border cities.

Importantly, while Mudenda (2013) investigates the interaction between tariffs and border effects, his focus is confined to *intranational* price dispersion and city border effects within Zambia.[75] Thus, he does not address the question of how tariffs contribute to observed *international* border effects on cross-country product price dispersion. This chapter extends the literature in this direction by examining the relationship between preferential tariff reform and the South Africa-Zambia border effect.

5.5 Data Description

This chapter examines the effect of tariff liberalization on prices and border effects. This is done by evaluating the effect of the reduction in preferential tariffs on Zambian imports from South Africa on domestic prices in Zambia using district-level retail prices for 14 traded products from 2002 to 2009 (see Table 16). The 14 traded products can be categorised broadly into fruit and vegetables, dairy products, meat, non-perishable food products, durable consumer products, medicine, and household electronics.

The period from 2002 to 2009 covers a stage of significant liberalization of tariffs on intra-SADC trade. The sample includes different categories of products according to the SADC tariff phase down schedule: two Category A products (which were liberalized first), 10 Category B products (for which tariffs were to be eliminated by 2008) and two Category C products (whose tariffs were to be eliminated by 2012). There is thus product-level variation in tariffs in the sample over time.

[75] He does, nevertheless, examine the effect of 'external borders' – measured through a dummy equal to one if both Zambian cities in a bilateral city pair have international ports of entry – on product market integration.

Table 16: Product sample for tariff analysis

Product category	Product	Unit	Category for SADC tariff liberalization
fruit and vegetables	onions	1 kg	C
	potatoes	1 kg	B
dairy products	cheddar cheese	1 kg	B
	margarine	250 g	B
meat	beef brisket	1 kg	C
	rump steak	1 kg	B
non-perishable food products	baked beans	420 g	B
	biscuits	200 g	B
	rice	1 kg	A
	spaghetti	500 g	B
durable consumer products	shoe polish	50 ml	B
medicine	cough syrup	100 ml	A
household electronics	electric iron	each	B
	electric kettle	each	B

The raw price data for the 14 products is drawn from the database of monthly average retail prices collected by statistical agencies for the computation of each country's CPI (which is also used in Chapters 2, 3 and 4). The monthly prices are recalculated net of VAT and converted to common currency United States dollar prices. Outliers are then systematically removed by deleting monthly product price observations that are either three times greater than the median monthly price for that particular product or less than one third of the median monthly price.

The remaining monthly price observations for each district in the sample are used to compute an annual average price for each product *i* in district *j* in year *t*.[76] This is done in order to smooth out month-to-month price fluctuations and approximate the stable long-run product prices in each district; while dampening the influence of seasonal factors or temporary promotions on product prices in particular districts. The final dataset of annual product prices spans 53 districts spread across the three SADC countries. This total is divided into 10 South African districts, six districts in Botswana and 37 in Zambia.[77]

[76] Monthly price observations for the electric iron and electric kettle in the Botswana districts are only available in the raw data for the years between 2006 and 2009. In order to obtain annual prices for these districts in all years, the growth rate in log prices for the years in which the data is available (2006 to 2009) is used to extrapolate backwards and obtain prices for each year between 2002 and 2005. The estimation results involving the Botswana districts in section 5.7 are robust to the exclusion of these two products from the sample.

[77] The district coverage varies by product across years, with price observations not available for all districts in all years.

In order to analyse the relationship between tariffs and national border effects in the SADC region, product-level MFN and preferential tariff data relevant to trade between the SACU countries (South Africa and Botswana) and Zambia is obtained from two sources. Applied MFN tariffs at the Harmonized System (HS) six-digit code level for SACU are drawn from the WTO's Integrated Database. In turn, data on Zambia's MFN tariffs and preferential tariff phase-down offers to South Africa at the HS eight-digit code level is obtained from the Zambian Revenue Authority.[78] The specific MFN and preferential tariff rates for each product in the sample are extracted from the full tariff databases by manually constructing a concordance to link each product to unique six-digit and eight-digit HS codes based on the HS code description that most closely matches the product in question (see Table D.1 in Appendix D for a mapping of the 14 products to HS codes and accompanying descriptions). The final tariff dataset varies by type of tariff, product and year.

5.6 Methodology and Empirical Specification

A production-consumption pairs approach is employed in this chapter in order to examine the contribution of preferential tariffs to the South Africa-Zambia border effect. The baseline production-consumption pair model is specified as follows:

$$p_{i,c,t} = \beta_0 + \beta_1 p_{i,p,t} + \beta_2 \ln(dist_{pc}) + \beta_3 \ln(dist_{pc})^2 + \beta_4 border_{SA-Zam} + \beta_5 border_{SA-Zam} * ln(1 + ptariffZamSA_{i,t}) + \beta_6 \ln(nontradedprice_{c,t}) + \gamma_t + \gamma_i + \varepsilon_{pc,t} \quad (2)$$

where $p_{i,c,t}$ is the log price of product *i* in consumption district *c* at time *t*; $p_{i,p,t}$ is the log price of product *i* in the production location at time *t*; $\ln(dist_{pc})$ is the log of the shortest road distance between the production location and consumption district *c*, and $\ln(dist_{pc})^2$ is log distance squared;[79] $border_{SA-Zam}$ is a dummy variable equal to one if consumption district *c* is located in Zambia and zero otherwise; $ptariffZamSA_{i,t}$ is Zambia's preferential tariff offer to South Africa for product *i* at time *t;* $\ln(nontradedprice_{c,t})$ is the log dollar price of a men's haircut in consumption district *c* at time *t*; γ_t and γ_i are time and product fixed effects, respectively; and $\varepsilon_{pc,t}$ is the regression error term.

The application of the production-consumption pair model is particularly relevant to this study given the dominance of South African exports in bilateral trade flows within the region. Southern

[78] The Zambian MFN and preferential tariff data series were constructed by Dale Mudenda.
[79] The inclusion of the log distance squared term accounts for potential nonlinearities in the relationship between distance from the production location and the price in the consumption district.

Africa represents one of the few settings in the world where well-grounded assumptions can be made on the direction of trade flows at the individual product-level. The assumption that products are produced in South Africa for consumption in the region allows for a focus on spatially relevant supplier to market pairs, while excluding potentially irrelevant market pairs between which trade in products does not occur. In the case of the latter, bilateral product price gaps are not directly informative of the magnitude of the cost of trade between market pairs (this is discussed in detail in Chapter 3). In contrast, price comparisons between product source and consumption destinations are "informative about international relative to local trade barriers" (Anderson & Van Wincoop, 2004: 740).

Given the dominant status of the Gauteng province as South Africa's economic engine, it is assumed that all 14 products originate from the Gauteng region. In order to ensure that a price observation is included for each product in all years from 2002 to 2009, an average Gauteng price for each product and year is calculated using the prices for all districts in Gauteng for which a price observation is available in that particular year.[80] Bilateral road distances from Pretoria to all other districts included in the sample are used to account for the distance between the production location in Gauteng and the various consumption locations.[81]

In essence, the specification in equation (2) measures the extent to which the price of a particular product in consumption district *c* is related to the price of that product at its source (Gauteng), the transportation cost incurred in moving the product from Gauteng to the consumption district (which is proxied by the bilateral distance between district pairs), and the cost associated with crossing the border into Zambia in the case of the Gauteng-Zambia district pairs. The price of a non-traded service (men's haircut) in the consumption districts is included in the specification to account for the influence of the cost of non-traded inputs on the final retail prices of products in those districts. This is necessary since final retail prices in a particular location are a function of the cost of both traded and non-traded inputs (Crucini et al., 2005a and 2005b).

---

[80] For the period from 2002 to 2008, the Gauteng product price averages for each year are calculated using some combination of the prices in the Pretoria, Vanderbijlpark, Vereeniging and Witwatersrand districts (depending on which districts prices are observed for a particular product and year). In turn, the 2009 averages for each product are calculated using some combination of prices in the City of Johannesburg Metro, Ekurhuleni Metro, Pretoria, Vanderbijlpark, Vereeniging and Witwatersrand districts. The difference is due to changes in the sample of districts covered in the raw data between 2008 and 2009.

[81] As before, distance is measured as the shortest road distance between district pairs. These distances are calculated from Google Maps using the longitude and latitude coordinates of the main city in each district (or the mid-point in districts not defined by a particular city).

The key coefficients of interest in equation (2) are β_4 and β_5. The coefficient on the South Africa-Zambia border dummy measures the additional impact of crossing the border on the price of product *i* in consumption district *c* located in Zambia relative to the price in a consumption district in South Africa, after accounting for the influence of the distance from the production location (Gauteng). Without the inclusion of the interaction between the border dummy and the preferential tariff variable, β_4 would measure the unique effect of crossing the international border on the price of product *i* in the consumption district. Since the preferential tariffs only apply when crossing the Zambian border, the inclusion of the preferential tariff variable allows the effect of the border on the price in the consumption district to vary for different values of the preferential tariff rate. The coefficient on the interaction term, β_5, thus measures the marginal effect of the preferential tariff rate on the effect of crossing the national border on the retail price in the consumption district. It is anticipated that the effect of the border in raising the price of product *i* in the consumption location will be greater for higher values of the preferential tariff rate.

In analysing the effect of crossing the national border on the price of product *i* in consumption district *c* using the specification in equation (2), the bilateral district pairs involving Gauteng and another district *within* South Africa serve as a control. Importantly, however, as South Africa and Zambia are not contiguous countries, products exported from South Africa must first travel through either Botswana or Zimbabwe on their way to Zambia. Since South Africa and Botswana are both members of SACU, South African products are not subject to tariffs when crossing the border into Botswana. Thus, the border effect between South Africa and Botswana should be unrelated to any tariff effects on relative prices.

The addition of Gauteng-Botswana district pairs deals with potential omitted variable bias by accounting for the effect of crossing the Botswana-South Africa border on relative prices, and makes it possible to separate out these effects from the tariff effect on relative prices between South Africa and Zambia. It is anticipated that once accounting for the cost of trade, the price of product *i* in a consumption district in Botswana should be equal to the price of that product in Gauteng, $p_{i,p,t}$; whereas the price of product *i* in a consumption district in Zambia should be equal to: $p_{i,p,t}\left(1 + ptariffZamSA_{i,t}\right)$.

The augmented specification with the inclusion of the Gauteng-Botswana district pairs is thus:

$$p_{i,c,t} = \beta_0 + \beta_1 p_{i,p,t} + \beta_2 \ln\left(dist_{pc}\right) + \beta_3 \ln\left(dist_{pc}\right)^2 + \beta_4 border_{SA-Botswana} + \beta_5 border_{SA-Zam} + \beta_6 border_{SA-Zam} * ln(1 + ptariffZamSA_{i,t}) + \beta_7 \ln(nontradedprice_{c,t}) + \gamma_t + \gamma_i + \varepsilon_{pc,t} \quad (3)$$

where $border_{SA-Botswana}$ is a dummy variable equal to one if consumption district *c* is located in Botswana and zero otherwise; and the remaining variables are defined as for equation (2).

The baseline and augmented production-consumption pair specifications outlined above are used in the empirical analysis presented in the following section. The analysis examines the influence of tariffs on border effects in the SADC region by focusing on trade between South Africa and Zambia. This is achieved by empirical estimating the extent to which the presence of tariffs on South African products imported into Zambia contributes to cross-border dispersion in product prices between the two countries.

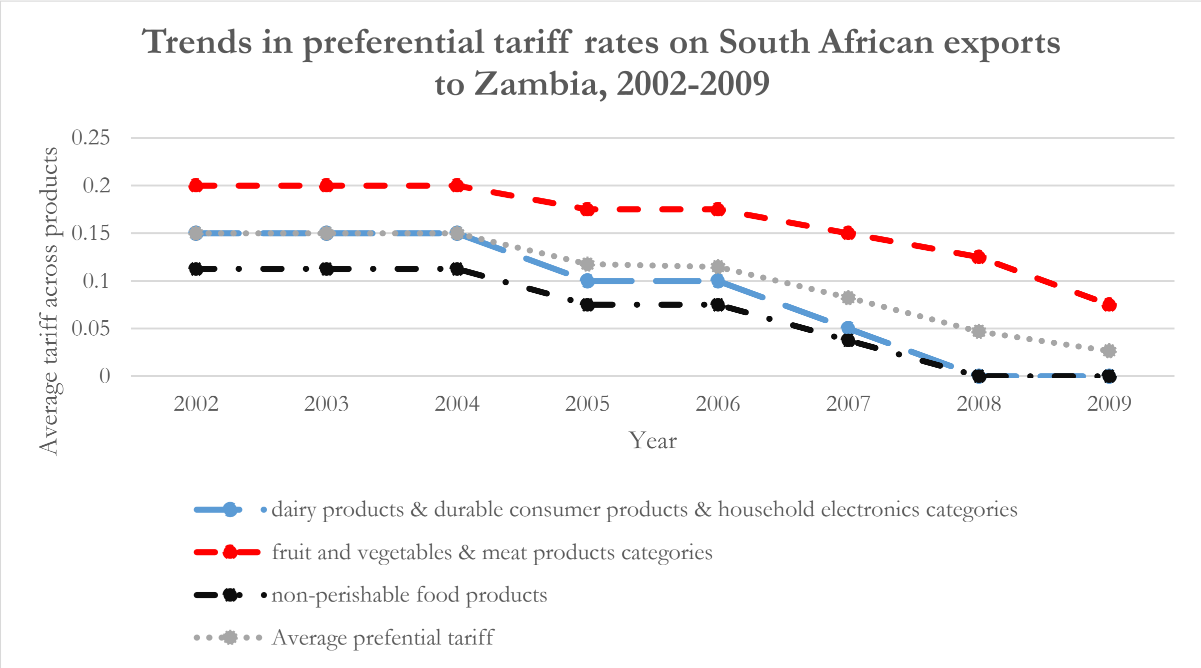
5.7 Empirical Results

Figure 11 plots the average variation over time in preferential tariffs on South African exports to Zambia for specific product categories as well as the average for all 14 products in the sample. Across all product categories, the average preferential tariff rates declined between 2002 and 2009. In all cases, the decline in average preferential tariffs occurred from 2004 onwards, and generally accelerated between 2006 and 2009. This is consistent with the back-loading of Zambia's tariff phase down schedule under the SADC Protocol on Trade.

On average, preferential tariffs on South African exports to Zambia remained highest for fruit and vegetables and meat products across the sample period. These tariffs were notably higher than those applied to dairy products, durable consumer products and household electronics. The average preferential tariffs applied to products in the non-perishable food products category were consistently lower than those in all other categories across the entire sample period.

Zambia's MFN and preferential tariff rates for individual products in each year between 2002 and 2009 are provided in Table D.2 in Appendix D. Zambia's preferential tariffs on South African exports of the two Category A products (rice and cough syrup) were immediately liberalized, meaning that they stood at zero throughout the 2002 to 2009 period. In line with the tariff phase down schedule for Category B products, Zambian preferential tariffs on South African exports of cheddar cheese, margarine, potatoes, rump steak, baked beans, biscuits, spaghetti, shoe polish, electric irons and electric kettles all fell to zero in 2008. Consequently, only the two category C products (onions and beef brisket) were still subject to tariffs when crossing the Zambian border in 2009, but even these declined substantially from 25% in 2008 to 15% in 2009.

Figure 11: Trends in average preferential tariff rates on South African exports to Zambia by product category and year, 2002-2009



The MFN tariff rates reported in Table D.2 in Appendix D show that there was little variation in Zambia's applied MFN tariffs between 2002 and 2009. Indeed, Zambia's MFN tariff rates remained unchanged over this period for 12 of the 14 products in the sample (the only exceptions were rice and cough syrup). The lack of variation in Zambia's MFN tariff rates is particularly important for the empirical analysis in this chapter as it means that any observed changes in price differences between Gauteng and the Zambian districts will not have been driven by Zambia's external tariff. Similarly, over the same period South Africa's MFN tariff rates also remained largely unchanged (see Table D.3 in Appendix D). Indeed, there was no change in South Africa's MFN tariff rates for 12 of the 14 products between 2002 and 2009 (beef brisket and biscuits being the only exceptions where MFN rates fell by 10% and 4%, respectively). This allays concerns that declining South African MFN tariff rates could coincide with changes in the Zambian preferential tariff on South African imports.[82]

[82] Were there to have been changes in the South African MFN tariff rates that coincided with the Zambian preferential tariff phase downs, it would be necessary to account for the effect of changes in the South African MFN tariffs on prices in South Africa and Botswana in the specifications in equations (2) and (3) in order to avoid omitted variable bias. One way in which to do this would be to deflate the South African and Botswana product prices by the South African MFN rate.

Turning to the price data, Table D.4 in Appendix D compares average dollar prices for each of the sampled products by country over the whole period from 2002 to 2009.[83] In the case of South Africa, separate averages are calculated for Gauteng and the remaining districts in the sample, making it possible to isolate the average prices in the designated production location. For some products, the average prices in Gauteng are higher than those in Botswana and/or Zambia. This may be due to the fact that per capita incomes are higher in Gauteng (and South Africa). In general, wages are higher in wealthier countries, which translates into higher non-traded prices (Edwards & Rankin, 2012). Higher non-traded prices will be reflected in higher final retail prices, which are a function of the cost of both traded and non-traded inputs. Table D.5 in Appendix D shows that the average price of a men's haircut (a proxy for non-traded prices) between 2002 and 2009 is substantially higher in Gauteng compared to districts in Botswana and Zambia. The differences in non-traded prices provide one explanation for the higher retail prices observed for some products in Gauteng; and underline the importance of accounting for the price of non-traded inputs in the production-consumption pair regressions.

5.7.1 Country average product prices and tariffs

As a starting point for the empirical analysis of the product price and tariff data, a simple examination of the relationship between the average product prices in each country is undertaken. It is anticipated that both the Botswana and Zambian prices will be correlated with the price of the product at its source in South Africa. In addition, the price in Zambia should be a function of both the price in South Africa and the preferential tariff on South African imports into the Zambian market. Consequently, simple price regressions are run using the following basic specification:

$$p_{i,C,t} = \beta_0 + \beta_1 p_{i,SA,t} + \beta_2 \, ln(1 + ptariffZamSA_{i,t}) + \gamma_t + \gamma_i + \varepsilon_{pc,t} \quad (4)$$

where $p_{i,C,t}$ is the log of the average US dollar price of product *i* in consumption country C (either Botswana or Zambia) in year *t*; $p_{i,SA,t}$ is the log of the average US dollar price of product *i* in South Africa in year *t*; $ptariffZamSA_{i,t}$ is Zambia's preferential tariff offer to South Africa for product *i* at time *t* (this variable is excluded when considering the relationship between the South African and Botswana prices); γ_t and γ_i are time and product fixed effects, respectively; and $\varepsilon_{pc,t}$ is the regression error term.

---

[83] Trends in the average prices for each product in each country and year over the sample period were also graphed in order to provide a check on the price data. For illustrative purposes, a selection of these graphs for individual products (baked beans, cheddar cheese, beef brisket and shoe polish) are reproduced in Figure D.1 in Appendix D.

The regression results from the estimation of equation (4) are presented in Table 17. As expected, the results confirm that product prices in both Botswana and Zambia are correlated with prices in South Africa.[84] Furthermore, the estimation results in column (2) show that the average product price in Zambia is positively and significantly related to the preferential tariff on South African imports into the country, meaning that a higher preferential tariff on South African imports of product *i* will raise its average retail price in the Zambian market. Specifically, a 1 percentage point increase in the average preferential tariff results in a 2.7% increase in the Zambian price.[85]

Table 17: Estimation of the relationship between average annual product prices and preferential tariffs in South Africa, Botswana and Zambia

	(1)	(2)
ln(South African price)	0.133**	0.336***
	(0.0619)	(0.0871)
ln(1+ptariffZamSA)		2.923***
		(0.777)
constant	-0.656***	-0.438***
	(0.0864)	(0.163)
Observations	112	112
R-squared	0.995	0.991

Notes: The dependent variable is the log of the average United States dollar price of product *i* across districts in Botswana (column (1)) and Zambia (column (2)) in year *t*. All regressions are estimated with year and product fixed effects. Robust standard errors are reported in parentheses. Significance at the 10 percent, 5 percent and 1 percent levels is denoted by *, ** and *** respectively.

5.7.2 Tariffs, Relative Prices and Border Effects in the SADC Region

The empirical analysis now moves on to a more formal econometric analysis of the relationship between tariffs, relative prices and border effects using the specifications presented in equations (2) and (3). Initially, the specification in equation (2) is estimated with the Gauteng-Botswana bilateral district pairs excluded from the analysis. The regression results are reported in Table 18.

[84] In an alternative regression (results not reported here), average product prices in Zambia were also found to be correlated with prices in Botswana.

[85] The average preferential tariff on Zambian imports from South Africa is 9.6% over the whole 2002 to 2009 period.

Table 18: Baseline production-consumption pair estimation excluding Gauteng-Botswana pairs

	(1)	(2)
ln(Gauteng price)	0.288***	0.305***
	(0.0382)	(0.0380)
log distance	0.719*	0.714*
	(0.409)	(0.416)
log distance squared	-0.0580*	-0.0576*
	(0.0316)	(0.0321)
$border_{SA\text{-}Zambia}$	0.258***	0.149**
	(0.0614)	(0.0659)
$border_{SA\text{-}Zambia}$*ln(1+ptariffZamSA)		0.971***
		(0.224)
ln(non-traded price)	0.0676***	0.0553**
	(0.0226)	(0.0244)
constant	-2.567*	-2.545*
	(1.299)	(1.325)
Observations	3,729	3,729
R-squared	0.911	0.912

Notes: The dependent variable is the log price of product *i* in consumption district *c* at time *t*. All regressions are estimated with year and product fixed effects. Robust standard errors (reported in parentheses) are clustered by district pair. Significance at the 10 percent, 5 percent and 1 percent levels is denoted by *, ** and *** respectively.

The preferential tariff variable is initially excluded from the regression in column (1). The coefficients on the Gauteng price variable, the South Africa-Zambia border dummy and the non-traded price variable are all positive and statistically significant at the 1% level, indicating that the retail price of product *i* in a district in Zambia is influenced by the price of that product at its source (Gauteng), the cost associated with moving the product across the border into Zambia, and by non-traded prices in the destination district. The price of product *i* in the consumption district in Zambia is also influenced by the distance between that district and the product's source – which serves as a proxy for the cost of transporting the product from Gauteng to its final retail destination in Zambia.[86] The negative and statistically significant coefficient on the distance squared term, however, indicates that the impact of distance-related transport costs on the price of the product in the destination district begins to decline beyond a certain distance from the product's source location. Both the coefficients on the log distance and log distance squared terms are, however, imprecisely estimated. The regression results for column (1) show that, conditional on the other

[86] The coefficient is, however, only weakly significant (at the 10% level). This is likely due to a high level of correlation between the distance variable and the South Africa-Zambia border dummy, with the cross-country bilateral district pairs all separated by greater distances than the district pairs within South Africa.

explanatory factors, crossing the border into Zambia adds 29.4%,[87] on average, to the final retail price of product i in the Zambian market.

The coefficient on the border-preferential tariff interaction term introduced in column (2) is positive and statistically significant at the 1% level. This indicates that the magnitude of the effect of crossing the national border on the price in Zambian districts varies for different rates of the preferential tariff, with higher (lower) tariffs leading to higher (lower) retail prices in Zambia. The estimation results imply that a 1% increase in $(1 + ptariffZamSA_{i,t})$, which is roughly equivalent to a 10% increase in the mean tariff,[88] results in a 0.97% increase in the Zambian price. Hence, preferential tariffs matter for final retail prices in Zambia. In fact, the estimation results suggest that there is very close to a perfect pass through of preferential tariffs onto Zambian retail prices.

When the preferential tariff variable is added to the regression in column (2), the coefficient on the border dummy falls. This indicates that the South Africa-Zambia border effect on product prices is due, in part, to the presence of tariffs on South African imports into Zambia. In fact, a substantial portion of the border effect is due to tariffs. The inclusion of the preferential tariff variable causes the additional increase in prices in the Zambian market caused by the border effect to fall from 29.4% to 16.1%.

The finding that preferential tariffs are a significant contributor to the South Africa-Zambia border effect has potentially important implications for policies designed to boost product market integration. It implies that the removal of tariffs can be effective in reducing border effects. Given this finding, it is possible that the implementation of tariff phase downs may have coincided with a decline in the border effect over time. This would be consistent with greater product market integration. The results presented in Chapter 4, however, indicate that the magnitude of the aggregate South Africa-Zambia border effect actually increased between 2006 and 2009 (see section 4.3.2). To examine whether this was also the case for the longer time period from 2002 to 2009, the South Africa-Zambia border dummy is interacted with a time trend in the following simple production-consumption pair specification:

$$p_{i,c,t} = \beta_0 + \beta_1 p_{i,p,t} + \beta_2 \ln(dist_{pc}) + \beta_3 \ln(dist_{pc})^2 + \beta_4 trend + \beta_5 border_{SA-Zam} + \beta_6 border_{SA-Zam} * trend + \beta_7 \ln(nontradedprice_{c,t}) + \gamma_i + \varepsilon_{pc,t} \qquad (5)$$

[87] Calculated as: exp(0.258)-1.
[88] As indicated earlier, the mean preferential tariff rate is 9.6% across all products over the 2002 to 2009 period.

where $trend$ is the time trend variable; and the remaining variables are defined as in equation (2). In this specification, the coefficient on the border dummy, β_4, measures the border effect at the start of the sample period in 2002; while the coefficient on the border-trend interaction term, β_5, captures the trend in the border effect over time. The preferential tariff variable is excluded from the specification as the objective here is to estimate the trend in the aggregate border effect.

The estimation results are presented in Table D.6 in Appendix D. The positive and significant coefficient on the border-trend interaction term shows that the aggregate South Africa-Zambia border effect *increased* between 2002 and 2009. This is consistent with the findings presented in Chapter 4 (for the 2006-2009 period); and implies that the tariff phase downs have not, ultimately, lead to greater product market integration between South Africa and Zambia. One explanation for this could be that while tariffs are important, they have become less important in driving integration as other factors such as NTBs have become more apparent. This is consistent with the findings in some other studies in the broader market integration literature, which have shown that tariff preferences have become less important compared to other factors – such as NTBs and beyond the border barriers to trade – as drivers of economic integration (Schiff & Winters, 2003; Olper & Raimondi, 2008; Hartzenberg, 2011; World Bank, 2012; Koczan & Plekhanov, 2013).

Nevertheless, the finding that tariffs account for a significant portion of the South Africa-Zambia border effect suggests that without the phasing down of tariffs, the increase in the border effect over the sample period would have been more substantial. This has important policy implications, suggesting that simply removing tariffs can be effective in reducing border effects. Hence, tariff liberalization remains an important policy tool in promoting product market integration.

Controlling for the effect of crossing the South Africa-Botswana border on product prices

To test the robustness of the core result of a positive tariff contribution to the border effect, the Gauteng-Botswana district pairs are now added to the sample in order to estimate the augmented specification introduced in equation (3). As indicated in section 5.6, the inclusion of these bilateral pairs accounts for potential omitted variable bias by controlling for the impact of crossing the Botswana border on product price differences between South Africa and Zambia. This makes it possible to separate out the specific tariff effect on product prices in the Zambian districts. The regression results with the inclusion of the Gauteng-Botswana district pairs are reported in Table 19.

Comparing the estimation results in Tables 18 and 19, the estimated coefficient on the Gauteng price is marginally smaller once the Gauteng-Botswana district pairs are included in the regression, while the coefficient on the South Africa-Zambia border dummy variables is larger. Hence, the South Africa-Zambia border dummy is not capturing a lower South Africa-Botswana border effect. The sign and significance of the coefficients on the log distance and log distance squared variables remain unchanged. When the preferential tariff variable is included in column (2) of Table 19, the coefficient on the South Africa-Zambia border dummy still declines, but now by approximately half the decline observed when only the cross-country pairs involving South Africa and Zambia are included in the regressions. In this case, the South Africa-Zambia border effect adds 27.6% to the final retail price in Zambia.

Table 19: Production-consumption pair estimation including Gauteng-Botswana pairs

	(1)	(2)
ln(Gauteng price)	0.257***	0.271***
	(0.0346)	(0.0349)
log distance	0.734*	0.728*
	(0.418)	(0.424)
log distance squared	-0.0580*	-0.0577*
	(0.0325)	(0.0330)
$border_{SA\text{-}Botswana}$	0.0811**	0.0831**
	(0.0331)	(0.0353)
$border_{SA\text{-}Zambia}$	0.312***	0.244***
	(0.0601)	(0.0655)
$border_{SA\text{-}Zambia}$*ln(1+ptariffZamSA)		0.774***
		(0.159)
ln(non-traded price)	0.100***	0.0990***
	(0.0235)	(0.0261)
constant	-2.756**	-2.737**
	(1.320)	(1.342)
Observations	4,413	4,413
R-squared	0.911	0.911

Notes: The dependent variable is the log price of product *i* in consumption district *c* at time *t*. All regressions are estimated with year and product fixed effects. Robust standard errors (reported in parentheses) are clustered by district pair. Significance at the 10 percent, 5 percent and 1 percent levels is denoted by *, ** and *** respectively.

Furthermore, the coefficient on the border-preferential tariff interaction term is now smaller. This suggests that the marginal effect of the preferential tariff on the South Africa-Zambia border effect in the baseline specification was overestimated due to omitted variable bias. Now, a 1% increase in $(1 + ptariffZamSA_{i,t})$, results in a 0.77% increase in the Zambian price. Thus, even after the inclusion of the Gauteng-Botswana pairs, the close to perfect pass-through of preferential tariffs onto Zambian domestic prices still exists.

Accounting for product-level variation in import intensity

The analysis presented thus far has assumed homogeneity across products in the effect of tariffs on the border effect. However, the tariff effects may differ depending on the importance of the South African product in Zambia's total imports. To date, the price-based literature on border effects has largely assumed common border effects regardless of the extent to which products are actually traded across borders. This section contributes to the empirical literature on border effects by allowing for variation in import intensity across traded products.

In order to do so, product-level data on the US dollar value of imports into Zambia in 2002 (the base year in the sample) is obtained from the United Nations International Trade Statistics (Comtrade) database. The data is used to calculate the ratio of Zambian imports of product *i* from South Africa to total Zambian imports of that product from the world (hereafter referred to as import intensity). The resulting product-specific import intensity values all fall somewhere within the range between zero and one (see Table D.7 in Appendix D).[89,90] There is a relatively even distribution of products across the range of import intensity shares, with no clear pattern across most product categories. The product with the lowest import intensity from South Africa is potatoes, accounting for just 1% of total Zambian imports of that product. South African exports of shoe polish and baked beans both have shares of less than 20% in Zambia's total imports. For the remaining 11 products, imports from South Africa all account for more than one quarter of Zambia's total imports of those products. The shares of imports of electric kettles, rice, cheddar cheese, spaghetti, rump steak, cough syrup and beef brisket from South Africa all exceed 50%. In the case of rump steak, cough syrup and beef brisket, imports into Zambia are sourced entirely from South Africa.

[89] The median value of the South African share in Zambia's total imports across all 14 products in the sample is 0.37.
[90] It is important to note that this is not a measure of the share of imports from South Africa in Zambian consumption because domestic output is not included.

In order to examine the effect of import intensity, two additional interactions are added to the specification initially introduced in equation (3):

$$p_{i,c,t} = \beta_0 + \beta_1 p_{i,p,t} + \beta_2 \ln(dist_{pc}) + \beta_3 \ln(dist_{pc})^2 + \beta_4 border_{SA-Botswana} + \beta_5 border_{SA-Zam}$$
$$+\beta_6 border_{SA-Zam} * importintensity_i + \beta_7 border_{SA-Zam} * ln(1 + ptariffZamSA_{i,t})$$
$$+ \beta_8 border_{SA-Zam} * ln(1 + ptariffZamSA_{i,t}) * importintensity_i + \beta_9 \ln(nontradedprice_{c,t})$$
$$+ \gamma_t + \gamma_i + \varepsilon_{pc,t} \quad (6)$$

where $importintensity_i$ measures the share of South African imports of product *i* in Zambia's total imports of that product in 2002; and the remaining variables are defined as for equation (3).

The first term is an interaction between the South Africa-Zambia border dummy and the import intensity variable. It is anticipated that the coefficient on this variable will be negative. Products that are traded more heavily across borders are likely to benefit from more established cross-border transit and trading networks and customs clearance procedures.[91] Put simply, they are likely to be more heavily traded precisely because the cost associated with trading across the border are lower for these products. Hence, the effect of the border on relative prices should be smaller for more heavily traded products.

The second term added to the specification is an interaction between the import intensity variable and the border-preferential tariff interaction. For products where only a small proportion of Zambian imports are sourced from South Africa, the influence of the preferential tariff on the prices of those products in the Zambian market may be relatively limited.[92] At the same time, the influence of the preferential tariff may be comparatively greater for products with a higher import intensity from South Africa. Hence, the coefficient on this interaction term is expected to be positive.

The regression results from the estimation of equation (6) are presented in Table 20. The coefficient on the South Africa-Zambia border dummy increases in magnitude with the inclusion of the border-import intensity interaction. Importantly, however, the coefficient on the South

[91] South African companies such as Shoprite and Woolworths spend significant sums of money on import licensing permits and paperwork for the administration of compliance with rules of origin and forwarding requirements when exporting products to Zambia (Charalambides, 2013). Due to economies of scale, these costs are likely to fall for products that are exported in larger volumes. Shoprite has also invested heavily in the development of local supply chains, information technology and sourcing processes to overcome logistical difficulties in exporting its products to African markets. For example, the company's dedicated fresh fruit and vegetable distributor (Fresh Mark) operates in Zambia, with depots in Lusaka and Kitwe. These investments and initiatives are designed to reduce the cost of trading products across the Zambian border.

[92] In these instances, domestic prices in Zambia may be more affected by Zambia's MFN tariff.

Africa-Zambia border-import intensity interaction is negative and statistically significant. This indicates that, as anticipated, the border effect on Zambian prices is significantly lower for products that are traded more heavily across the border. For example, in the case of a product that is sourced entirely from South Africa (in other words where South Africa's share of total Zambian imports of that product equals 1), the border effect is 0.069.[93] In this case, the South Africa-Zambia border effect adds just 7.1% to the final retail price in Zambia (compared to the 27.6% on average across all products). This result highlights the importance of accounting for heterogeneity in the extent to which products are actually traded across borders when computing border effects. To date, this has not been done in the border effects literature.

Table 20: Production-consumption pair estimation accounting for import intensity from South Africa

ln(Gauteng price)	0.264***
	(0.0356)
log distance	0.845**
	(0.407)
log distance squared	-0.0670**
	(0.0316)
$border_{SA\text{-}Botswana}$	0.0851**
	(0.0347)
$border_{SA\text{-}Zambia}$	0.529***
	(0.0689)
$border_{SA\text{-}Zambia}$* import intensity	-0.460***
	(0.0581)
$border_{SA\text{-}Zambia}$*ln(1+ptariffZamSA)	0.542**
	(0.252)
$border_{SA\text{-}Zambia}$*ln(1+ptariffZamSA)*import intensity	0.211
	(0.354)
ln(non-traded price)	0.103***
	(0.0262)
constant	-3.221**
	(1.287)
Observations	4,413
R-squared	0.916

Notes: The dependent variable is the log price of product *i* in consumption district *c* at time *t*. The regression is estimated with year and product fixed effects. Robust standard errors (reported in parentheses) are clustered by district pair. Significance at the 10 percent, 5 percent and 1 percent levels is denoted by *, ** and *** respectively.

93 Calculated as: 0.529-0.46.

Consistent with the previous specifications, the coefficient on the border-preferential tariff interaction remains positive and statistically significant at the 1% level. Once accounting for variation in import intensity across products, a 1% increase in $(1 + ptariffZamSA_{i,t})$ results in a 0.54% increase in the Zambian price. The tariff pass-through onto domestic prices is thus marginally lower, on average, after taking into account variation in import intensity across products. The estimated coefficient on the additional border-preferential tariff-import intensity interaction is positive but not statistically significant. This suggests that the intensity with which products are imported from South Africa compared to other foreign varieties has no additional influence on the average marginal effect of preferential tariffs on the South Africa-Zambia border effect.

5.8 Conclusion

This chapter has examined the relationship between preferential tariffs and border effects in the SADC region by focusing on trade between South Africa and Zambia. In doing so, the chapter addresses the lack of studies in the international literature on the contribution of tariffs to border effects on international relative prices.

The relationship is analysed empirically using a production-consumption pair specification that assumes a unidirectional flow of trade from South Africa to Zambia. This makes it possible to account for the spatial relationship between production and consumption, thus facilitating more accurate measurement of the border effect and its tariff component, and differentiating the analysis from much of the existing research on tariffs in the product market integration literature. Furthermore, by using disaggregated price level data for narrowly defined products the analysis in this chapter avoids the compositional and aggregation problems that affect studies based on aggregate price indices.

The tariff contribution to the South Africa-Zambia border effect is assessed by exploiting variation in tariffs at the product level, providing an advantage over many other studies in the literature that use average tariff rates across products. By using aggregate tariff data, these studies are likely to face endogeneity problems arising from the difficulty of excluding product-specific effects influencing tariffs. This chapter deals with the causality concerns that affect other studies by using data on prior agreed tariff phase downs.

Similarly, by using product-level data on the share of imports from South Africa in Zambia's total imports, the analysis in this chapter accounts for the possibility that the effect of national borders and tariffs on cross-country relative prices may vary according to the intensity with which the products are traded across the border. The importance of import intensity for border and tariff effects on relative prices has not been examined in the literature. The chapter thus contributes to the international literature by providing new insight into the relationship between import intensity and national border effects on relative prices. In line with theoretical expectations, it is shown that the effect of the South Africa-Zambia border on product prices in the Zambian market is lower for products that are traded more heavily across the border. This highlights the importance of accounting for heterogeneity in the extent to which products are actually traded across borders in future research on border effects.

The core finding in this chapter is that preferential tariffs account for a significant portion of the border effect on product price dispersion between South Africa and Zambia. In the baseline specification, the inclusion of the preferential tariff variable causes the additional increase in prices in the Zambian market generated by the border to fall from 29.4% to 16.1%. This result remains qualitatively the same after controlling for trade between South Africa and Botswana – which is not subject to tariffs and is a likely transit route for South African products on route to the Zambian market.

The finding that preferential tariffs account for a significant portion of the South Africa-Zambia border effect also remains qualitatively the same after accounting for variation in the intensity with which products are traded between South Africa and Zambia. However, variation in import intensity is shown to be important for the magnitude of the border effect. For example, in the case of a product that is sourced entirely from South Africa, the South Africa-Zambia border effect adds just 7.1% to the final retail price in Zambia, compared to 27.6% on average across all products.

The chapter also finds that there is an almost perfect pass-through of preferential tariffs onto domestic prices in Zambia. Depending on the specification, a 1% increase in $(1 + ptariffZamSA_{i,t})$ results in a between 0.54% and 0.97% increase in the Zambian retail price. This suggests that tariff liberalization can have potentially positive welfare implications by reducing Zambian retail prices. Lower retail prices result in higher real wages, meaning that households generally see their incomes rise. The benefit that domestic consumers derive from lower retail prices is likely to translate into increases in domestic demand and consumption. In the case of food products, lower retail prices may boost food consumption, potentially leading to lower levels of malnutrition and poverty.

From a product market integration perspective, the finding that preferential tariffs are a significant contributor to the South Africa-Zambia border effect has potentially important implications for policies designed to boost product market integration. It implies that tariff reforms can be effective in reducing border effects. In the case of trade between South Africa and Zambia, however, the fact that the border effect actually increased between 2002 and 2009 suggests that the phasing down of tariffs on Zambian imports of products from South Africa did not, ultimately, lead to greater integration of the two countries' product markets. One possible explanation for this finding is that while tariffs are important, they have become less important in driving integration as other factors such as NTBs and beyond the border barriers to trade have become more apparent. Future research should examine the role played by NTBs and other barriers to trade in driving a wedge between product prices and constraining progress towards greater integration of product markets in the SADC region.

Nevertheless, the results presented in this chapter indicate that without the phasing down of preferential tariffs, the increase in the border effect over the sample period would have been higher. This suggests that tariff reform should not be discounted as an instrument to drive greater product market integration.

6. General Conclusion and Policy Implications

6.1 Summary of Key Findings

This thesis extends the price-based empirical literature on border effects and product market integration to the SADC region. The analysis draws on a unique dataset of district-level monthly average retail prices spanning a number of districts in Botswana, Malawi, South Africa and Zambia. By using disaggregated product price level data, the empirical analysis presented in this thesis is able to evaluate product market integration from the perspective of absolute deviations from the LOP within and between SADC countries, while also accounting for heterogeneity across products. The analysis thus avoids the compositional and aggregational issues that affect similar studies in the empirical literature that are based on CPI data.

The thesis is comprised of four main chapters. The first main chapter (Chapter 2) provides a descriptive analysis of the extent to which product prices are integrated within and between the four SADC countries. The results provide new insight into the degree to which retail product prices are dispersed across districts in the region. Large and persistent deviations from the LOP are found both within and between the SADC countries over the period from 2006 to 2009. The analysis shows that even within the four countries product prices are highly dispersed and have not, on average, become more integrated.

The analysis reveals that retail product prices are far more dispersed *between* the four SADC countries compared to across districts *within* each country. Within the four countries, product prices differed between districts by 34% on average in 2009; while between-country price deviations averaged 77% in that year. The general result of greater dispersion in prices between countries is consistent with the findings of similar empirical studies for other regions in Africa as well as in other parts of the world. In terms of magnitude, while the price dispersion estimates for the four SADC countries are larger than estimates reported in the literature for North American and European countries, they are quite similar to those computed in comparable work in the empirical literature on developing regions. They are, however, somewhat larger than other estimates for different regions in Africa.

The descriptive analysis in Chapter 2 also reveals evidence of considerable heterogeneity in price dispersion across products. This is the case even within smaller sub-groups of highly homogenous and less homogenous products. This highlights the importance or accounting for product heterogeneity in studies of product market integration.

Looking at trends in price dispersion between countries in the SADC region, when all four countries are considered together, there is no clear evidence to suggest that product prices became more integrated between 2006 and 2009. The only exception was the COMESA countries (Malawi and Zambia), between which there was increased price integration over the same period.

Overall, the descriptive evidence presented in Chapter 2 points to a degree of fragmentation of markets in the region. This is particularly evident between markets located in different SADC countries. Nevertheless, the descriptive analysis does provide interesting insights into factors that reduce price dispersion between countries. Specifically, it is shown that both geographic proximity and membership in additional trade agreements are important drivers of product price integration. This is evident from the fact that absolute price deviations are lower, on average, in contiguous countries and in those that share membership in the SACU.

Chapter 2 then moves on to a more formal empirical analysis of the association between product price integration between countries in the SADC region and distance, borders and membership in regional trade groupings. This is examined using the standard approach in the literature; thus making it possible to test whether key stylized features in the international literature hold in the SADC context. The estimation results indicate that absolute price dispersion between district pairs in the SADC region increases the further apart the districts are from each other, and is higher in the case of districts separated by a national border. The results from the preferred regression specification indicate that 100km in distance between districts causes prices to deviate from the LOP by 17.5%, while crossing a national border in the SADC region adds 11.8%, on average, to relative prices between countries.

The estimation results in Chapter 2 also provide interesting insights into product heterogeneity in border effects. The average border effect in the SADC region is shown to be smaller for highly homogenous products compared to those considered to be less homogenous. Specifically, the average SADC border effect is around 3 percentage points smaller when estimated using only a sub-set of highly homogenous products.

The econometric analysis of distance and border effects on product price dispersion presented in Chapter 2 is based on the standard empirical approach in the price-based literature. This standard approach and certain recent methodological advances to estimating transaction costs and border effects is critically evaluated in the second main chapter (Chapter 3) of the thesis. The evaluation identifies various different sample selection effects that bias estimates of distance and border effects in the existing literature. Building on the quantile regression approach initially introduced by Borraz et al. (2012), the sensitivity of the distance and border coefficients to these sample selection biases is tested using actual product price data for the four SADC countries.

First, the results indicate that the standard pooled OLS estimates reported in much of the existing literature suffer from a sample selection problem which biases the estimated distance and border coefficients *downwards* relative to the true cost of trade. This is because by pooling price gap observations for all possible bilateral pairs, they include observations in their regressions for which the arbitrage constraint is not binding and in which the bilateral price gaps are less than the actual cost of trade. Two methods have been proposed in the literature to deal with this inequality constraint problem: a production-consumption pair method and a quantile regression approach.

The chapter highlights the impact of two additional selection biases that are not dealt with in the Borraz et al. (2012) application of the quantile regression approach. It shows that by pooling absolute price differences for different types of products in the same distance bins and selecting one observation from each bin for inclusion in the regression, the standard application of the quantile regression approach involves selecting products for which distance and borders have the greatest impact on relative prices. This product sample selection bias raises the estimates of the average distance and border effect coefficients.

The chapter also shows that the standard approach of pooling relative prices for all possible bilateral market pair combinations regardless of the distance between them does not account for the possibility that transaction costs are more likely to exceed price gaps at larger distances. This raises the likelihood of including market pairs in the sample for which the equality constraint is not binding and where the observed price gaps are less than the actual transaction costs of trade. By progressively relaxing restrictions on the maximum distance separating district pairs in the sample to test the sensitivity of the modified quantile regressions to different distance cut-offs, the chapter demonstrates empirically that the inclusion of these observations in the quantile regressions results in a sample selection bias which lowers the estimated distance coefficient. In contrast, the border estimates increase in magnitude as the distance between district pairs increases, which may be due to higher levels of unobserved product heterogeneity, greater variation in unobserved district characteristics or the influence of remote (non-contiguous) border effects at larger distances.

It is also shown that by not accounting for variation in within-product quality across districts, the price-based regressions attribute the portion of the price gap that is actually generated by unobserved product heterogeneity to the role of distance and border-related trade costs. This results in omitted variable bias which raises the estimated distance and border coefficients.

By using data on actual product prices for the SADC region to assess the sensitivity of the distance and border coefficients to these different sample selection biases, this thesis makes an important methodological contribution to the price-based literature on border effects and product market integration. The results form the basis for the development of novel extensions to the quantile regression methodology to allow for the analysis of cross-country border effects; and to account for sample selection bias arising due to product and distance sample selection effects.

Chapter 4 applies a version of the extended quantile regression approach, which deals with the general sample selection bias in the standard regressions in the literature as well as the product sample selection bias in the Borraz et al. (2012) application of the quantile regression model, in order to obtain precise estimates of average and individual border effects in the SADC region. This contributes to the empirical literature by extending the application of the quantile regression approach to the study of cross-country border effects. It also improves on the accuracy of other estimates of border effects in Africa – provided in Versailles (2012), Nchake (2013), Aker et al. (2014) and Brenton et al. (2014) – that are derived using variants of the standard approach in the literature and, hence, subject to the sample selection biases identified in this thesis.

The precisely estimated border effect results presented in Chapter 4 reveal significant variation in border effects on price dispersion across the individual country pairs. The border effect is found to be smallest for Botswana and Zambia, where crossing the border increases price dispersion between the two countries by 22.6% relative to the average level of price dispersion within the two countries. The effects on price dispersion of the Botswana-South Africa (23.1%), Malawi-Zambia (26.6%), South Africa-Zambia (32.4%), Botswana-Malawi (35.7%) and Malawi-South Africa (56.4%) borders are consecutively larger.

The results obtained from the application of the extended quantile regression approach in Chapter 4 also provide insight into a number of different dimensions of border effects across different country groupings in the SADC region. Border effects tend to be comparatively lower between SADC countries that share membership in additional regional trade agreements. The borders between the SACU (23.1%) and COMESA (26.6%) countries generate smaller increases in price dispersion compared to the average effect of crossing a border between the rest of the SADC countries (36.2%). Clear differences are also found in the magnitude of border effects for contiguous and non-contiguous SADC countries. The additional price dispersion generated by crossing the border between contiguous countries is 24.1%; compared to 41.1% for countries separated by more than one national border. This points to the presence of incremental border effects on relative prices as products are traded across multiple borders. The ranking of these border effects across different regional trade agreements and for contiguous versus non-contiguous countries is shown to be robust across different specifications.

There is also some evidence to suggest that these incremental border effects in the SADC region are more substantial for perishable products. Again, this highlights the importance of accounting for product heterogeneity in border effects.

The analysis in Chapter 4 also reveals that there was generally little change in the magnitude of border effects in the SADC region between 2006 and 2009. In fact, the average border effect on cross-country price dispersion for the SADC region as well as the individual border effects generally increased by small margins (with the exception of the Malawi-Zambia border). The main insight that can be drawn from this is that border effects in the region generally did not fall, despite the liberalization of tariffs on intra-SADC trade between 2006 and 2009.

The final main chapter of the thesis (Chapter 5) focuses specifically on the relationship between this tariff reform process and border effects in the SADC region. The analysis in Chapter 5 is based on a smaller sample of products to that used in the previous chapters, but spans a longer time period from 2002 to 2009. The core focus is on unpacking the tariff contribution to the South Africa-Zambia border effect over this period. No studies have looked directly at the potential contribution of tariffs in generating border effects on international relative prices and, hence, the analysis presented in Chapter 5 makes an important contribution in extending the literature in this direction.

Moreover, the analysis in Chapter 5 deals with some of the methodological concerns that affect other studies of tariffs and product market integration. It employs a production-consumption pair specification in order to account for the spatial relationship production and consumption in the SADC region. This facilitates more accurate measurement of the border effect. In addition, it draws on product-level tariff data, rather than aggregate data on average tariff rates across products, and is thus able to exploit variation in tariffs at the product level. Similarly, by using data on prior agreed tariff phase downs, the analysis deals with the causality concerns that affect studies using aggregate tariff data which are unable to exclude product-specific effects influencing tariffs.

The estimation results presented in Chapter 5 add several new insights to the border effect literature. The first is to contribute to the understanding of the relationship between import intensity and national border effects. Using product-specific data on the share of imports from South Africa in total Zambian imports, the results in Chapter 5 show that the border effect on Zambian prices is significantly lower for products that are traded more heavily across the border. For example, in the case of a product that is sourced entirely from South Africa, the border adds just 7.1% to the final retail price in Zambia, compared to 27.6% on average across all products. This result highlights the importance of accounting for heterogeneity in the extent to which products are actually traded across borders when computing border effects. To date, this has not been done in the border effects literature.

Turning to the main focus on the relationship between tariffs, prices and border effects. The estimation results reveal an almost perfect pass-through of preferential tariffs onto domestic prices in Zambia. The results presented in Chapter 5 also show that preferential tariffs account for a significant portion of the border effect on product price dispersion between South Africa and Zambia. After accounting for the increase in retail prices in Zambia that is generated by the presence of preferential tariffs, the border effect falls from 29.4% to 16.1%. This result remains qualitatively the same after controlling for trade between South Africa and Botswana. It also remains qualitatively the same after accounting for variation in the intensity with which products are traded between South Africa and Zambia. These results suggest that tariff liberalization can have potentially positive welfare implications by reducing Zambian retail prices

Even so, a simple analysis of the trend in the South Africa-Zambia border effect indicates that the impact of crossing the border in raising Zambian retail prices actually increased between 2002 and 2009. This suggests that the phasing down of preferential tariffs did not, ultimately, lead to greater integration of the two countries' product markets. Nevertheless, given the significant contribution of tariffs to the border effect, the results do suggest that the increase in the border effect over this period would have been more substantial in the absence of any liberalization of tariffs on Zambian imports of South African products.

6.2 Policy Implications of the Findings

The results of this research contribute to the understanding of product market integration in developing countries and the importance of different instruments of regional integration. The period of analysis, which spans eight years from 2002 to 2009, coincides with a period of significant liberalization of tariffs on intra-SADC trade under the SADC Protocol on Trade (which commenced in 2000). While there is some evidence that volumes of intra-SADC trade have grown since 2000, it is not clear whether this growth actually reflects changes in product market integration. The results presented in this thesis, which are based on disaggregated retail product price data rather than trade flows, provide new insight into whether or not the implementation of the SADC Protocol on Trade has coincided with greater integration of product markets in the region. In this respect, the findings in this thesis advance the understanding of the links between trade policy and product market integration in Africa.

In this context, the evidence presented in this thesis of large and persistent absolute price deviations both within and between countries in the SADC region, coupled with the finding that the average SADC border effect and most of the individual border effects in the region have *increased* over time, has important policy implications. On balance, the evidence indicates that markets in the region remain fragmented, with little sign of greater product market integration despite the explicit policy focus on trade reform that has accompanied the introduction of the SADC Protocol on Trade.

What is constraining greater product market integration in the SADC region? The generally large border effects on cross-country dispersion in relative prices reported in this thesis suggest that barriers to trade across national borders in the region are substantial. Hence, border-related barriers to trade are clearly hampering greater regional integration of product markets.

What is less clear is precisely what is causing these border barriers. The border effects reported in this thesis represent an aggregate measure of the impact of all factors involved in crossing national borders on systematic differences in prices between districts in different countries in the SADC region. The final chapter of this thesis unpacks the influence of one possible factor – the tariff contribution to the South Africa-Zambia border effect. The results indicate that the presence of preferential tariffs accounts for a substantial portion of the border effect. From a policy perspective, this finding implies that further tariff reforms could be effective in reducing border effects and thus, in theory, facilitating greater integration of product markets.

Even so, the effect of crossing the border on price dispersion between South Africa and Zambia actually increased between 2002 and 2009, despite the phasing down of preferential tariffs on imports of South African products into the Zambian market. Hence, tariff reforms have not been sufficient on their own to drive product market integration between South Africa and Zambia. This suggests that tariffs only tell part of the story, and that other barriers to trade across the border are present that continue to hinder the integration of product markets in the region.

In this sense, the results presented in this thesis are consistent with recent arguments made elsewhere in the economic integration literature. These arguments have suggested that while tariffs are important, they have become less significant in driving integration as other factors such as NTBs and beyond the border barriers to trade have become more apparent (Schiff & Winters, 2003; Olper & Raimondi, 2008; Hartzenberg, 2011; World Bank, 2012; Koczan & Plekhanov, 2013). In the SADC context, rules of origin are likely to be a key barrier constraining greater product market integration. The complex and restrictive rules of origin and import sourcing requirements contained within the SADC Protocol on Trade hamper trade, raise transaction costs and compromise the ability of producers in the region to benefit from preferential tariff reductions (Erasmus et al., 2004; Brenton et al., 2005; Charalambides, 2013). In this sense, a move towards a simplified and more transparent rules of origin framework for the SADC could play an important role in facilitating greater integration of product markets in the region (Brenton et al., 2005).

6.3 Suggestions for Further Research

Future research could build on the important contribution made in this thesis towards advancing the understanding of the tariff contribution to border effects in the SADC region. This thesis has highlighted the significant contribution of preferential tariffs on Zambian imports of South African products to the South Africa-Zambia border effect. It will be important to establish whether tariffs on intra-SADC trade have had similar effects for other countries in the region. This is particularly relevant given the presence of complex and restrictive rules of origin requirements that may affect the ability of exporters in some SADC countries to benefit from tariff preferences. Owing to its comparatively advanced industrial base, South Africa is able to meet the rules of origin requirements to qualify for preferential tariffs. However, in other SADC countries where producers struggle to meet these relatively onerous requirements, the contribution of tariffs to the border effect may be negligible.

The results presented in this thesis also suggest that further research is required to identify the additional determinants of border effects on cross-country product price dispersion in the SADC. In this regard, it will be interesting to examine the contribution of NTBs to border effects in the region. Existing research suggests that NTBs are pervasive on the African continent, and include onerous procedures regulating cross-border business activity; complex and inefficient customs arrangements; quantitative restrictions; price controls; restrictive rules of origin; limited harmonisation of policies, regulations and procedures across countries; informal roadblocks along key trade corridors; inefficient border management and logistics; strict permit and licensing requirements for cross-border trade; and poorly designed technical regulations and standards (Keane et al., 2010; Brenton et al., 2011; World Bank, 2012; Mbekeani, 2013). Isolating the role of these factors in generating border effects will aid the development of policies designed to reduce barriers to intra-SADC trade and drive product market integration in the region.

References

Aker, J.C., Klein, M.W., O'Connell, S.A. & Yang, M. 2014. Borders, ethnicity and trade. *Journal of Development Economics.* 107(March 2014): 1-16.

Amjadi, A. & Yeats, A. 1995. *Have Transport Costs Contributed to the Relative Decline of Sub-Saharan African Exports? Some Preliminary Empirical Evidence.* Policy Research Working Paper 1559. World Bank.

Anderson, M.A., Davies, M. & Smith, S.L.S. 2014. Ethnic Networks and Price Dispersion. (Unpublished).

Anderson, M.A., Schaefer, K.C. & Smith, S.L.S. 2010. Price Dispersion in Spatial Perspective: Theory and Evidence. (Unpublished).

Anderson, M.A., Schaefer, K.C. & Smith, S.L.S. 2013. Can Price Dispersion Reveal Distance-Related Trade Costs? Evidence from the United States. *Global Economy Journal.* 13(2): 151-173.

Anderson, J.E. & van Wincoop, E. 2003. Gravity with Gravitas: A Solution to the Border Puzzle. *The American Economic Review.* 93(1): 170-192.

Anderson, J.E. & van Wincoop, E. 2004. Trade Costs. *Journal of Economic Literature.* 42(3): 691-751.

Anderson, M.A., Schaefer, K.C. & Smith, S.L.S. 2010. *A Spatial Model of Price Dispersion: Theoretical and Empirical Findings.* Unpublished manuscript.

Anderson, M.A. & Smith, S.L.S. 2004. Borders and Price Dispersion: New Evidence on Persistent Arbitrage Failures. *Mid-West International Economics Meetings, April 30—May2, 2004.* April 30 — May2.

Asplund, M. & Friberg, R. 2001. Deviations from the Law of One Price in Scandinavian Duty-Free Stores. *American Economic Review.* 91(4): 1072-1083.

Atkin, D. & Donaldson, D. 2014. *Who's Getting Globalized? The Size and Implications of Intranational Trade Costs.* Unpublished manuscript.

Bahmani-Oskooee, M. & Gelan, A. 2006. Testing the PPP in the Non-linear STAR Framework: Evidence from Africa. *Economics Bulletin.* 6(17): 1-15.

Baier, S.L. & Bergstrand, J.H. 2001. The growth of world trade: tariffs, transport costs, and income similarity. *Journal of International Economics.* 53(1): 1-27.

Bas, M. & Strauss-Kahn, V. 2013. *Input-Trade Liberalization, Export Prices and Quality Upgrading.* Discussion Paper 2013-13. Sciences Po Department of Economics.

Baulch, B. 1997. Transfer Costs, Spatial Arbitrage, and Testing for Food Market Integration. *American Journal of Agricultural Economics.* 79(2): 477-487.

Behar, A. & Edwards, L. 2011. *How Integrated is SADC? Trends in Intra-Regional and Extra-Regional Trade Flows and Policy.* Development Research Group Policy Research Working Paper 5625. World Bank.

Bergin, P.R. & Glick, R. 2007. Tradability, productivity, and international economic integration. *Journal of International Economics.* 73: 128-151.

Besedes, T. & Prusa, T.J. 2006. Product differentiation and duration of US import trade. *Journal of International Economics.* 70(2): 339-358.

Borraz, F. 2006. Border Effects between U.S. and Mexico. *Journal of Economic Development.* 31(1): 53-62.

Borraz, F., Cavallo, A., Rigobon, R. & Zipitría, L. 2012. *Distance and Political Boundaries: Estimating Border Effects under Inequality Constraints.* Working Paper 18122. National Bureau of Economic Research.

Bradford, S.C. & Lawrence, R.Z. 2004. *Has Globalization Gone Far Enough? The Costs of Fragmented Markets.* Washington, DC: Institute for International Economics.

Brander, J.A. & Spencer, B.J. 1984. Trade Warfare: Tariffs and Cartels. *Journal of International Economics.* 16(3): 227-242.

Brenton, P., Dihel, N., Gillson, I. & Hoppe, M. 2011. *Regional trade agreements in sub-Saharan Africa: supporting export diversification.* Africa Trade Policy Note #15. World Bank.

Brenton, P., Flatters, F. & Kalenga, P. 2005. *Rules of Origin and SADC: The Case for Change in the Mid Term Review of the Trade Protocol.* World Bank Africa Region Working Paper Series No. 83. Washington, D.C.: World Bank.

Brenton, P., Portugal-Perez, A. & Régolo, J. 2014. *Food Prices, Road Infrastructure, and Market Integration in Central and Eastern Africa.* Policy Research Working Paper 7003. World Bank.

Broda, C. & Weinstein, D.E. 2008. *Understanding International Price Differences using Barcode Data.* (Working Paper 14017). National Bureau of Economic Research.

Burstein, A. & Jaimovich, N. 2012. Understanding Movements in Aggregate Product-Level Real Exchange Rates. (Unpublished).

Campa, J.M. & Goldberg, L.S. 2002. *Exchange Rate Pass-through into Import Prices: A Macro or Micro Phenomenon.* Working Paper 8934. National Bureau of Economic Research.

Cavallo, A., Neiman, B. & Rigobon, R. 2013. Currency Unions, Product Introductions, and the Real Exchange Rate. (Unpublished).

Ceglowski, J. 2003. The law of one price: intranational evidence for Canada. *Canadian Journal of Economics.* 36(2): 373-400.

Chahrour, R. & Stevens, L. 2012. Equilibrium Price Dispersion and the Border Effect. (Unpublished).

Chang, T., Chang, H., Chu, H. & Su, C. 2006. Does PPP Hold in African Countries? Further Evidence Based on a Highly Dynamic Non-linear (Logistic) Unit Root Test. *Applied Economics.* 38(20): 2453-2459.

Charalambides, N. 2013. *What Shoprite and Woolworths can tell us about Non-tariff Barriers.* Occasional Paper No. 148. South African Institute of International Affairs.

Cherkaoui, M., Khellaf, A. & Nihou, A. 2011. *The Price Effect of Tariff Liberalization in Morocco: Measuring the Impact on Household Welfare.* Working Paper No. 637. Economic Research Forum.

Chevassus-Lozza, E., Latouche, K. & Majkovic, D. 2007. How Much do Non-Tariff Measures Explain the Border Effect at Entry to the EU Market? The CEECs Agri-Food Exports to EU in the Pre-Accession Period. *American Agricultural Economics Association Annual meeting July 29-August 1, 2007.*

Crucini, M.J. & Shintani, M. 2008. Persistence in law of one price deviations: Evidence from micro-data. *Journal of Monetary Economics.* 55(3): 629-644.

Crucini, M.J., Shintani, M. & Tsuruga, T. 2008. *Accounting for Persistence and Volatility of Good-Level Real Exchange Rates: The Role of Sticky Information.* Working Paper No. 14381. National Bureau of Economic Research.

Crucini, M.J. & Telmer, C.I. 2012. *Microeconomic Sources of Real Exchange Rate Variation.* Working Paper No. 17978. National Bureau of Economic Research.

Crucini, M., Telmer, C. & Zachariadis, M. 2005a. *Price Dispersion: The Role of Distance, Borders and Location.* 2005 Meeting Papers, Society for Economic Dynamics 767. Society for Economic Dynamics.

Crucini, M.J., Telmer, C.I. & Zachariadis, M. 2005b. Understanding European Real Exchange Rates. *The American Economic Review.* 95(3): 724-738.

Crucini, M.J. & Yilmazkuday, H. 2014. Understanding long-run price dispersion. *Journal of Monetary Economics.* 66226-240.

Crucini, M., Telmer, C. & Zachariadis, M. *Price Dispersion: The Role of Distance, Borders and Location.* 2005 Meeting Papers, Society for Economic Dynamics 767. Society for Economic Dynamics.

De Loecker, J., Goldberg, P.K., Khandelwal, A.K. & Pavcnik, N. 2012. *Prices, Markups and Trade Reform.* Working Paper No. 17925. National Bureau of Economic Research.

Dumas, B. 1992. Dynamic Equilibrium and the Real Exchange Rate in a Spatially Separated World. *The Review of Financial Studies.* 5(2): 153-180.

Edwards, L. & Rankin, N. 2012. *Is Africa Integrating? Evidence from Product Markets.* ERSA Working Paper 292. Economic Research Southern Africa.

Engel, C. 1993. Real exchange rates and relative prices: An empirical investigation. *Journal of Monetary Economics.* 32(1): 35-50.

Engel, C. 1999. Accounting for U.S. Real Exchange Rate Changes. *Journal of Political Economy.* 107(3): 507-538.

Engel, C. & Rogers, J.H. 1996. How Wide Is the Border? *The American Economic Review.* 86(5): 1112-1125.

Engel, C. & Rogers, J.H. 2001. Deviations from purchasing power parity: causes and welfare costs. *Journal of International Economics.* 5529-57.

Engel, C. & Rogers, J.H. 2004. European Product Market Integration After the Euro. *Economic Policy.* 19(39): 347-384.

Engel, C., Rogers, J.H. & Wang, S. 2005. Revisiting the Border: An Assessment of the Law of One Price Using Very Disaggregated Consumer Price Data. In *Exchange Rates, Capital Flows and Policy.* R. Driver, P. Sinclair & C. Thoenissen, Eds. London: Routledge. 187-203.

Erasmus, H., Flatters, F. & Kirk, R. 2004. Rules of Origin as Tools of Development? Some Lessons from SADC. (Unpublished).

Evans, C.L. 2001. The Economic Significance of National Border Effects. *American Economic Review.* 93(4): 1291-1312.

Feenstra, R.C. 1989. Symmetric Pass-Through of Tariffs and Exchange Rates Under Imperfect Competition: An Empirical Test. *Journal of International Economics.* 27(1/2): 25-45.

Flatters, F. 2010. *Implementing the SADC FTA: Where Are We? What Next?* Gaborone, Botswana: USAID.

Foad, H. 2004. *Europe Without Borders? The Effect of the EMU on Relative Prices.* San Diego State University Department of Economics Working Papers No. 0017.

Foroutan, F. & Pritchett, L. 1993. Intra-Sub-Saharan African Trade: Is It Too Little. *Journal of African Economies.* 2(1): 74-105.

Frenkel, J.A. 1981. Flexible Exchange Rates, Prices, and the Role of "News": Lessons from the 1970s. *Journal of Political Economy.* 89(4): 665-705.

Froot, K., Kim, M. & Rogoff, K. 1995. *The Law of One Price over 700 Years.* Working Paper 5132. National Bureau of Economic Research.

Gillson, I. 2012. Deepening regional integration to eliminate the fragmented goods market in Southern Africa. In *De-Fragmenting Africa: Deepening Regional Trade Integration in Goods and Services.* P. Brenton & G. Isik, Eds. Washington, D.C.: The World Bank. 89-97.

Giovanni, A. 1988. Exchange rates and traded goods prices. *Journal of International Economics.* 24(1-2): 45-68.

Goldberg, P.K. & Knetter, M.M. 1997. Goods Prices and Exchange Rates: What Have We Learned? *Journal of Economic Literature.* 35(3): 1243-1272.

Gopinath, G., Gourinchas, P. & Hsieh, C. 2010. Estimating the Border Effect: Some New Evidence. (Unpublished).

Gopinath, G., Gourinchas, P., Hsieh, C. & Li, N. 2011. International Prices, Costs, and Markup Differences. *The American Economic Review.* 101(6): 2450-2486.

Gorodnichenko, Y. & Tesar, L.L. 2009. Border Effect or Country Effect? Seattle May Not Be so Far from Vancouver After All. *American Economic Journal: Macroeconomics.* 1(1): 219-241.

Grafe, C., Raiser, M. & Sakatsume, T. 2008. Beyond borders—Reconsidering regional trade in Central Asia. *Journal of Comparative Economics.* 36453-466.

Harrigan, J. 1993. OECD imports and trade barriers in 1983. *Journal of International Economics.* 35(1-2): 91-111.

Hartzenberg, T. 2011. *Regional Integration in Africa.* Staff Working Paper No. 2011-14. World Trade Organization Economic Research and Statistics Division.

Head, K. & Ries, J. 2001. Increasing Returns Versus National Product Differentiation as an Explanation for the Pattern of U.S.–Canada Trade. *The American Economic Review.* 91(4): 858-883.

Hillberry, R.H. 2002. Aggregation bias, compositional change, and the border effect. *Canadian Journal of Economics.* 35(3): 517-530.

Hillberry, R.H. & Hummels, D. 2003. Intranational Home Bias: Some Explanations. *The Review of Economics and Statistics.* 85(4): 1089-1092.

Hillberry, R.H. & Hummels, D. 2008. Trade responses to geographic frictions: A decomposition using micro-data. *European Economic Review.* 52527-550.

Holmes, M.J. 2000. Does Purchasing Power Parity Hold in African Less Developed Countries? Evidence from a Panel Data Unit Root Test. *Journal of African Economies.* 9(1): 63-78.

Horváth, J., Ratfai, A. & Dome, B. 2008. The border effect in small open economies. *Economic Systems.* 32(1): 3233-45.

Hummels, D. 2001. Towards a Geography of Trade Costs. (Unpublished).

Imbs, J., Mumtaz, H., Ravn, M.O. & Rey, H. 2005. PPP Strikes Back: Aggregation and the Real Exchange Rate. *The Quarterly Journal of Economics.* 120(1): 1-43.

Inanc, O. & Zachariadis, M. 2012. The Importance of Trade Costs in Deviations from the Law-of-One-Price: Estimates based on the Direction of Trade. *Economic Inquiry.* 50(3): 667-689.

Isard, P. 1977. How Far Can We Push the "Law of One Price"? *The American Economic Review.* 67(5): 942-948.

Kalenga, P. 2009. Implementation of the SADC Trade Protocol: Some Reflections. (Unpublished).

Kano, K., Kano, T. & Takechi, K. 2013. Exaggerated death of distance: Revisiting distance effects on regional price dispersions. *Journal of International Economics.* 90: 403-413.

Kawai, M. & Wignaraja, G. 2010. *Free Trade Agreements in East Asia: A Way toward Trade Liberalization?* Asian Development Bank Brief No. 1.

Keane, J., Calì, M. & Kennan, J. 2010. *Impediments to Intra-Regional Trade in Sub-Saharan Africa.* London, UK: Overseas Development Institute.

Knetter, M.M. & Slaughter, M.J. 2001. Measuring Product-Market Integration. In *Topics in Empirical International Economics: A Festschrift in Honour of Robert E. Lipsey.* M. Blomstrom & L.S. Goldberg, Eds. Chicago: University of Chicago Press. 15-46.

Koczan, Z. & Plekhanov, A. 2013. *How important are non-tariff barriers? Complementarity of infrastructure and institutions of trading partners.* Working Paper No. 159. European Bank for Reconstruction and Development.

Krugman, P.R. 1987. Pricing to market when the exchange rate changes. In *Real-financial linkages among open economies.* S.W. Arndt & D.J. Richardson, Eds. Cambridge, MA: MIT press. 49-70.

Landry, A. 2013. Borders and Big Macs. *Economics Letters.* 120(2): 318-322.

Lutz, M. 2003. *Price Convergence under EMU? First Estimates.* Economics Discussion Paper No. 2003-08. University of St. Gallen.

Maier, P. 2010. An Analysis of International Price Differentials on eBay. *Contemporary Economic Policy.* 28(3): 307-321.

Mathä, T.Y. 2003. *What to expect of the euro? Analysing price differences of individual products in Luxembourg and its surrounding regions.* Central Bank of Luxembourg Working Paper No. 8. Luxembourg: Central Bank of Luxembourg.

Mbekeani, K.K. 2013. *Understanding the Barriers to Regional Trade Integration in Africa.* Ghana: African Development Bank Group.

McCallum, J. 1995. National Borders Matter: Canada-U.S. Regional Trade Patterns. *The American Economic Review.* 85(3): 615-623.

Mengistae, T. 2012. The Business Environment in Southern Africa: Issues in Trade and Market Integration. In *De-Fragmenting Africa: Deepening Regional Trade Integration in Goods and Services.* P. Brenton & G. Isik, Eds. Washington, DC: World Bank. 113-120.

Moon, W. 2013. *Cross-Border Price Differentials and Goods Market Integration in East Asia.* ADBI Working Paper No. 426. Tokyo: Asian Development Bank Institute.

Morshed, M. 2007. Is there really a "border effect"? *Journal of International Money and Finance.* 26(7): 1229-1238.

Morshed, M. 2011. Border effects in the variability of rice price in the Indian subcontinent: Results from a natural experiment. *Journal of Asian Economics.* 22(4): 295-301.

Mudenda, D. 2013. Tariff reform and Product Market Integration in Developing Countries: Evidence from Zambia 1993-1999. *Biennial Conference of the Economic Society of South Africa, Bloemfontein, University of the Free State.* 25-27 September 2013.

Mudzonga, E. 2008. Implementation challenges for the SADC FTA: tariff and non-tariff barriers. In *Liberalising Trade in Southern Africa: Implementation challenges for the 2008 SADC FTA and beyond.* Midrand South Africa: Institute for Global Dialogue. 15-23.

Nagayasu, J. 2002. Does the Long-Run PPP Hypothesis Hold for Africa? Evidence from a Panel Cointegration Study. *International Monetary Fund Bulletin of Economic Research.* 54181-187.

Nchake, M.A. 2013. Product market price integration in developing countries. Doctor of Philosophy. University of Cape Town.

Nchake, M.A., Edwards, L. & Rankin, N. 2014. Price-Setting Behaviour in Lesotho: Stylised Facts from Consumer Retail Prices. *South African Journal of Economics.* forthcoming.

Nicita, A. 2004. *Who Benefited from Trade Liberalization in Mexico? Measuring the Effects on Household Welfare.* Policy Research Working Paper 3265. World Bank.

Nicita, A. 2009. The price effect of tariff liberalization: Measuring the impact on household welfare. *Journal of Development Economics.* 89(1): 19-27.

Nitsch, V. 2000. National Borders and International Trade: Evidence from the European Union. *The Canadian Journal of Economics.* 33(4): 1091-1105.

Olper, A. & Raimondi, V. 2008. Explaining National Border Effects in the QUAD Food Trade. *Journal of Agricultural Economics.* 59(3): 436-462.

Parsley, D.C. & Wei, S. 2001. Explaining the border effect: the role of exchange rate variability, shipping costs, and geography. *Journal of International Economics.* 5587-105.

Parsley, D.C. & Wei, S. 2002. *Currency Arrangements and Goods Market Integration: A Price Based Approach.* International Finance 0211004. EconWPA.

Parsley, D.C. & Wei, S. 2007. A Prism into PPP Puzzles: The Micro-Foundations of Big Mac Real Exchange Rates. *The Economic Journal.* 117(523): 1336-1356.

Paternoster, R., Brame, R., Mazerolle, P. & Piquero, A. 1998. Using the Correct Statistical Test for the Equality of Regression Coefficients. *Criminology.* 36(4): 859-866.

Portugal-Perez, A. & Wilson, J.S. 2008. *Trade Costs in Africa: Barriers and Opportunities for Reform.* Policy Research Working Paper No. 4719. World Bank.

Rauch, J.E. 1999. Networks versus markets in international trade. *Journal of International Economics.* 487-35.

Ravallion, M. 1986. Testing Market Integration. *American Journal of Agricultural Economics.* 68(1): 102-109.

Redding, S. & Venables, A.J. 2004. Economic geography and international inequality. *Journal of International Economics.* 62(1): 53-82.

Richardson, D.J. 1978. Some Empirical Evidence on Commodity Arbitrage and the Law of One Price. *Journal of International Economics.* 8(2): 341-351.

Robinson, A.L. 2013. Internal Borders: Ethnic Diversity and Market Segmentation in Malawi. *Working Group in African Political Economy National Meeting, May 3-4, 2013.*

Rogers, J.H. 2002. *Monetary union, price level convergence, and inflation: How close is Europe to the United States?* International Finance Discussion Paper No. 2002-74. Board of Governors of the Federal Reserve System.

Rogers, J.H. & Jenkins, M. 1995. Haircuts or hysteresis? Sources of movements in real exchange rates. *Journal of International Economics.* 38(3-4): 339-360.

Rogers, J.H. & Smith, H.P. 2001. *Border Effects within the NAFTA Countries.* International Finance Discussion Papers Number 698. Board of Governors of the Federal Reserve System.

Rogoff, K. 1996. The Purchasing Power Parity Puzzle. *Journal of Economic Literature.* 3(2): 647-668.

Rossi-Hansberg, E. 2005. A Spatial Theory of Trade. *The American Economic Review.* 95(5): 1464-1491.

Schiff, M. & Winters, A. 2003. *Regional Integration and Development.* Washington, DC: World Bank and Oxford University Press.

Schulze, M. & Wolf, N. 2009. On the origins of border effects: insights from the Habsburg Empire. *Journal of Economic Geography.* 9117-136.

Schwartz, J. 2012. Impact of institutions on cross-border price dispersion. *Review of World Economics.* 148(4): 617-645.

Teravaninthorn, S. & Raballand, G. 2009. *Transport Prices and Costs in Africa: A Review of the Main International Corridors.* Africa Infrastructure Country Diagnostic Working Paper 14. Washington, DC: The International Bank for Reconstruction and Development / The World Bank.

Versailles, B. 2012. *Market Integration and Border Effects in Eastern Africa.* CSAE Working Paper WPS/2012-01. Oxford, United Kingdom: Centre for the Study of African Economies.

Wei, S. 1996. *Intra-National versus International Trade: How Stubborn are Nations in Global Integration?* NBER Working Paper 5531. Cambridge, MA: National Bureau of Economic Research.

Wei, S. & Parsley, D.C. 1995. *Purchasing Power Disparity During the Floating Rate Period: Exchange Rate Volatility, Trade Barriers and Other Culprits.* Working Paper No. 5032. National Bureau of Economic Research.

Wolszczak-Derlacz, J. 2008. Price convergence in the EU - an aggregate and disaggregate approach. *International Economics and Economic Policy.* 525-47.

World Bank 2011. *Harnessing Regional Integration for Trade and Growth in Southern Africa.* Washington, DC: World Bank.

World Bank. 2012. *Defragmenting Africa: Deepening Regional Trade Integration in goods and services*. Washington, DC: World Bank.

Appendix A

Table A.1: Matching and classification of sampled products

Classification according to Rauch (1999)	Common product name (unit)	Original product name				Final classification	Product sub-group
		Botswana	Malawi	South Africa	Zambia		
Goods traded on an organized exchange (homogenous goods)	bananas (1kg)	bananas	bananas	bananas	banana	highly homogenous	food
	potatoes (1kg)	potatoes	Irish potato	potatoes	Irish potatoes	highly homogenous	food
	rice (1kg)	rice (Tastic)	rice-other (long grain)	rice	rice local	highly homogenous	food
Reference priced	cabbage (1kg)	cabbage	cabbage	cabbage	cabbage	highly homogenous	food
	onions (1kg)	onions	onions	onions	onion	highly homogenous	food
	oranges (1kg)	oranges	oranges	oranges	oranges	highly homogenous	food
	paraffin (1 litre)	paraffin	paraffin	paraffin	paraffin	highly homogenous	other products
	pineapples (1kg)	pineapples	pineapple	pineapples	pineapples	highly homogenous	food
	tomatoes (1kg)	tomatoes	tomatoes	tomatoes	tomatoes	highly homogenous	food
Differentiated products	bath towel (each)	bath towel	towel-bath (medium)	bath towel	bath towel, cotton, (medium)	less homogenous	clothing and textiles
	blanket (each)	blanket, woolen, 150cms x 200cms	blanket-special, 150cm x 200cm	blankets - made from cotton	single blanket, (Vanguard), 150cm * 200cm	less homogenous	clothing and textiles
	boy's shirt (each)	boys khaki shirt, size 6	shirt local poly/cotton	shirt – short sleeve	boys shirt, 6 to 8 years	less homogenous	clothing and textiles
	brassiere (each)	ladies brassiere, size 36	brassiere-padded, size 38	bra	bra	less homogenous	clothing and textiles

Classification according to Rauch (1999)	Common product name (unit)	Original product name				Final classification	Product sub-group
		Botswana	Malawi	South Africa	Zambia		
Differentiated products	colour television (each)	television, colour 51 cms screen, remote control/manual	TV recorder (Panasonic 21 ins)	colour TV-screen 50–70cm	television, colour, Philips, 21 inch	less homogenous	machinery, equipment and electronics
	electric iron (each)	electric iron, steam	iron electric-automatic	iron	electric iron, dry, Phillips	less homogenous	machinery, equipment and electronics
	electric kettle (each)	electric kettle, 1.5-2 litre, plastic one	electric kettle, 2.2lts	kettle	electric kettle, 2 litres	less homogenous	machinery, equipment and electronics
	girl's dress (each)	girls' dress, cotton or poly-cotton, size 28	dress-press, size 30	dress - summer - made from natural fabric	girls dress, 6 to 8 years	less homogenous	clothing and textiles
	ladies dress (each)	ladies dress, polyester and cotton	dress-imported, size 38	dress - summer - made from natural fabric	ladies dress, imported, lowest price	less homogenous	clothing and textiles
	margarine (250g)	margarine - not `soft', Butter Cup	margarine-Blue Band	block type margarine	margarine, Butter Cup	less homogenous	food
	men's shirt (each)	men's shirt, polyester/cotton, size 16	shirt/shirt (white)/shirt (Falcon)/shirt-Imported	business shirt – long sleeve	long sleeved shirt imported, lowest price	less homogenous	clothing and textiles
	men's suit (each)	men's suit, 65 polyester 35 cotton (two buttons), size 3	suit	suit - summer - jacket and trousers	gents' two piece suit	less homogenous	clothing and textiles
	men's trousers (pair)	men's trousers, polyester/cotton, size 4	trousers-Winger	trousers - summer - made from natural fabric	pair of trousers, local	less homogenous	clothing and textiles
	refrigerator (each)	electric refrigerator, 250-275 litres	refrigerator	refrigerator - 250 to 299 litre	refrigerator, any, 210 litres	less homogenous	machinery, equipment and electronics
	shoe polish (50ml)	shoe polish, 50ml, Kiwi	shoe polish, Kiwi (black), 50 ml	wax shoe polish	shoe polish, Kiwi, 50mls	less homogenous	other products

Table A.2: Average price in US$ by product and country (across all districts), 2008

	Average product price (in US$) in 2008			
Product	**Botswana**	**Malawi**	**South Africa**	**Zambia**
Highly homogenous products				
Food				
bananas	0.74	0.29	1.00	0.74
cabbage	0.68	0.22	0.79	0.38
onions	1.00	0.45	0.86	1.27
oranges	0.65	0.19	0.83	0.93
pineapples	0.77	0.37	0.75	0.77
potatoes	0.82	0.29	0.96	0.86
rice	1.58	0.61	1.38	1.34
tomatoes	1.18	0.39	1.28	0.83
Other products				
paraffin	1.01	0.87	1.00	1.17
Less homogenous products				
Food				
margarine	0.99	1.08	0.98	1.39
Clothing and textiles				
bath towel	4.46	4.71	5.06	5.75
blanket	17.38	11.68	30.36	7.83
boy's shirt	2.70	2.06	5.15	4.76
brassiere	3.83	1.73	5.28	2.16
girl's dress	5.30	1.63	8.37	7.37
ladies dress	13.03	5.60	15.32	14.96
men's shirt	5.82	3.38	7.75	10.17
men's suit	88.74	37.76	73.61	104.83
men's trousers	9.55	4.41	12.81	7.67
Machinery, equipment and electronics				
colour television	184.62	221.97	151.01	246.52
electric iron	11.84	21.54	16.55	21.53
electric kettle	13.01	22.02	15.86	15.04
refrigerator	403.33	471.59	387.73	290.60
Other products				
shoe polish	0.79	0.81	0.67	0.78

Note: The product price for each country is calculated as the mean price across all districts and all months.

Table A.3: Pooled regressions (all products) excluding South African districts, 2006-2009

	Basic regressions		Accounting for country-heterogeneity effect		Accounting for country-heterogeneity effect (highly homogenous products only)	
	(1)	(2)	(3)	(4)	(5)	(6)
log distance	0.0157***	0.0193***	0.0117***	0.0154***	0.0209***	0.0272***
	(0.00204)	(0.00205)	(0.00203)	(0.00205)	(0.00255)	(0.00255)
border	0.181***	0.174***	0.138***	0.126***	0.121***	0.101***
	(0.00357)	(0.00345)	(0.00463)	(0.00400)	(0.00609)	(0.00493)
COMESA		0.0967***		0.107***		0.166***
		(0.0125)		(0.0132)		(0.0190)
Observations	274,776	274,776	274,776	274,776	121,077	121,077
R-squared	0.270	0.271	0.272	0.274	0.274	0.279
Distance-equivalent of the border effect	101,586.1	8,230.0	132,570.8	3,575.4	326.8	41.0

Notes: All regressions include a constant and are estimated with product and year fixed effects. Standard errors are clustered by district pair and reported in parentheses. The rest of SADC dummy is the omitted category (covering Botswana-Malawi and Botswana-Zambia country pairs). It is not possible to include the SACU dummy in these regressions because Botswana-South Africa is the only SACU country pair in the sample, and is eliminated with the exclusion of South Africa. In the regressions in columns (2) and (3), country-specific dummies are included for 'Botswana-Botswana' and 'Malawi-Malawi' district pairs (with the 'Zambia-Zambia' dummy omitted). Significance at the 10 percent, 5 percent and 1 percent levels is denoted by *, ** and *** respectively.

Table A.4: Trend in mean absolute price dispersion for between-country district pairs excluding all South African districts, 2006-2009

	(1)	(2)
trend	0.00686***	0.00883***
	(0.000759)	(0.000747)
trend*COMESA		-0.0334***
		(0.00409)
Observations	139,677	139,677
R-squared	0.286	0.287

Notes: All regressions include a constant and are estimated with product and country pair fixed effects. Robust standard errors are clustered by district pair and reported in parentheses. Significance at the 10 percent, 5 percent and 1 percent levels is denoted by *, ** and *** respectively. It is not possible to include a year*SACU interaction term in these regressions because Botswana-South Africa is the only SACU country pair in the sample, and is eliminated with the exclusion of South Africa.

Appendix B

Table B.1: Quantile regression estimations with and without accounting for country selection biases

	Standard Borraz approach			**Separate within-country distance bins**			**Separate within- and between-country distance bins**		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	mean	p95	max	mean	p95	max	mean	p95	max
log distance	0.0188***	0.0582***	0.197***	0.0209***	0.0509***	0.145***	0.0371***	0.0727***	0.141***
	(0.00379)	(0.00711)	(0.0131)	(0.00297)	(0.00634)	(0.0109)	(0.00355)	(0.00656)	(0.00963)
border	0.184***	0.293***	0.217***	0.171***	0.307***	0.476***	0.211***	0.338***	0.312***
	(0.00922)	(0.0153)	(0.0372)	(0.00732)	(0.0138)	(0.0289)	(0.00712)	(0.0144)	(0.0230)
Observations	2,044	2,044	2,044	3,279	3,279	3,279	4,754	4,754	4,754
R-squared	0.495	0.560	0.444	0.412	0.379	0.417	0.464	0.451	0.381
Distance-equivalent of the border effect	17805.0	153.6	3.0	3575.3	416.3	26.6	295.1	104.5	9.1

Notes: The dependent variable is the mean, 95th percentile or the maximum of the absolute value of the log price difference between districts *j* and *k* in year *t* for bin *n* (using 500 distance bins). All regressions are estimated with a constant and year and product fixed effects. Robust standard errors (reported in parentheses) are clustered by distance bin. Significance at the 10 percent, 5 percent and 1 percent levels is denoted by *, ** and *** respectively.

Table B.2: Quantile regression estimation using 1,000 distance-border-bins

	(1)	(2)	(3)
	mean	p95	max
log distance	0.0248***	0.0668***	0.181***
	(0.00311)	(0.00639)	(0.0106)
border	0.173***	0.286***	0.254***
	(0.00646)	(0.0128)	(0.0261)
Observations	3,587	3,587	3,587
R-squared	0.519	0.507	0.423
Distance-equivalent of the border effect	1,070.4	72.3	4.1

Notes: The dependent variable is the mean, maximum or 95th percentile of the absolute value of the log price difference between districts *j* and *k* in year *t* for bin *n* (using 1,000 distance-border-bins). All regressions are estimated with a constant and year and product fixed effects. Robust standard errors (reported in parentheses) are clustered by distance bin. Significance at the 10 percent, 5 percent and 1 percent levels is denoted by *, ** and *** respectively.

Table B.3: Quantile regressions using sub-samples of district pairs with distance restrictions (stable sub-set of country borders)

	Panel A				
	(1) 800km	(2) 900km	(3) 1000km	(4) 1100km	(5) 1200km
log distance	0.0982***	0.1000***	0.100***	0.0957***	0.0904***
	(0.00507)	(0.00472)	(0.00448)	(0.00461)	(0.00462)
border	0.105***	0.115***	0.131***	0.149***	0.165***
	(0.0100)	(0.0101)	(0.0107)	(0.0119)	(0.0127)
Observations	48,996	53,034	56,633	59,937	62,951
R-squared	0.277	0.284	0.294	0.298	0.300
Distance-equivalent of the border effect	2.9	3.2	3.7	4.7	6.2
	Panel B				
	(6) 1300km	(7) 1400km	(8) 1500km	(9) 1600km	(10) 1798km
log distance	0.0881***	0.0867***	0.0853***	0.0842***	0.0831***
	(0.00451)	(0.00454)	(0.00447)	(0.00443)	(0.00449)
border	0.176***	0.190***	0.199***	0.204***	0.211***
	(0.0127)	(0.0136)	(0.0138)	(0.0137)	(0.0135)
Observations	65,345	67,578	69,410	70,713	73,029
R-squared	0.305	0.312	0.317	0.320	0.327
Distance-equivalent of the border effect	7.4	8.9	10.3	11.3	12.7

Notes: The dependent variable is the maximum of the absolute value of the log price difference between districts *j* and *k* in year *t* for bin *n* (using 500 distance bins). All regressions are estimated with a constant and year and product fixed effects. Robust standard errors (reported in parentheses) are clustered by distance bin. Significance at the 10 percent, 5 percent and 1 percent levels is denoted by *, ** and *** respectively.

Table B.4: Quantile regression estimations using sub-set of homogenous products (excluding fruit and vegetables)

	(1) mean	(2) p95	(3) max
log distance	0.0234***	0.0590***	0.0604***
	(0.00311)	(0.00447)	(0.00456)
border	0.147***	0.144***	0.148***
	(0.00586)	(0.0106)	(0.0108)
Observations	7,120	7,120	7,120
R-squared	0.249	0.296	0.301
Distance-equivalent of the border effect	534.9	11.5	11.6

Notes: The dependent variable is the mean, 95th percentile or the maximum of the absolute value of the log price difference between districts *j* and *k* in year *t* for bin *n* (using 500 distance bins). All regressions are estimated with a constant and year and product fixed effects. Robust standard errors (reported in parentheses) are clustered by distance bin. Significance at the 10 percent, 5 percent and 1 percent levels is denoted by *, ** and *** respectively.

Table B.5: Comparison of standard pooled OLS regression and quantile regression estimates (accounting for country-heterogeneity effect)

	Pooled OLS regression	Quantile regressions		
		mean	p95	max
	(1)	(2)	(3)	(4)
log distance	0.0420***	0.0185***	0.0583***	0.215***
	(0.00203)	(0.00408)	(0.00700)	(0.0119)
border	0.106***	0.166***	0.264***	0.223***
	(0.00409)	(0.00805)	(0.0208)	(0.0464)
Observations	412,972	2,044	2,044	2,044
R-squared	0.232	0.494	0.562	0.472
Distance-equivalent of the border effect	12.5	7,887.0	92.6	2.8

Notes: For the quantile regressions, the dependent variable is the mean, 95th percentile or the maximum of the absolute value of the log price difference between districts *j* and *k* in year *t* for bin *n* (using 500 distance-border-bins). All regressions are estimated with a constant and year and product fixed effects as well as country-specific dummies for 'Botswana-Botswana', 'Malawi-Malawi' and 'South Africa-South Africa' district pairs (with the 'Zambia-Zambia' dummy omitted). Robust standard errors (reported in parentheses) are clustered by district pair for the pooled OLS regression and by distance bin for the quantile regressions. Significance at the 10 percent, 5 percent and 1 percent levels is denoted by *, ** and *** respectively.

Table B.6: Quantile regression estimations with and without accounting for product selection bias (accounting for country-heterogeneity effect)

	Without accounting for product selection bias			Accounting for product selection bias		
	(1)	(2)	(3)	(4)	(5)	(6)
	mean	p95	max	mean	p95	max
log distance	0.0352***	0.0707***	0.160***	0.0399***	0.105***	0.107***
	(0.00369)	(0.00665)	(0.00960)	(0.00289)	(0.00504)	(0.00513)
border	0.178***	0.253***	0.178***	0.147***	0.177***	0.181***
	(0.00773)	(0.0193)	(0.0292)	(0.00651)	(0.0127)	(0.0129)
Observations	4,754	4,754	4,754	92,631	92,631	92,631
R-squared	0.477	0.476	0.433	0.305	0.386	0.388
Distance-equivalent of the border effect	157.1	35.8	3.0	39.8	5.4	5.4

Notes: The dependent variable is the mean, 95th percentile or the maximum of the absolute value of the log price difference between districts *j* and *k* in year *t* for bin *n* (using 500 distance bins). All regressions are estimated with a constant and year and product fixed effects as well as country-specific dummies for 'Botswana-Botswana', 'Malawi-Malawi' and 'South Africa-South Africa' district pairs (with the 'Zambia-Zambia' dummy omitted). Robust standard errors (reported in parentheses) are clustered by distance bin. Significance at the 10 percent, 5 percent and 1 percent levels is denoted by *, ** and *** respectively.

Table B.7: Quantile regressions using sub-samples of district pairs with distance restrictions (accounting for country-heterogeneity effect)

			Panel A		
	(1)	(2)	(3)	(4)	(5)
	800km	900km	1000km	1100km	1200km
log distance	0.113***	0.114***	0.114***	0.109***	0.102***
	(0.00516)	(0.00478)	(0.00445)	(0.00483)	(0.00501)
border	0.0871***	0.0913***	0.104***	0.119***	0.132***
	(0.0123)	(0.0119)	(0.0119)	(0.0125)	(0.0130)
Observations	48,996	53,034	56,633	59,937	62,951
R-squared	0.303	0.311	0.321	0.324	0.324
Distance-equivalent of the border effect	2.2	2.2	2.5	3.0	3.6

			Panel B		
	(6)	(7)	(8)	(9)	(10)
	1300km	1400km	1500km	1600km	1798km
log distance	0.100***	0.0998***	0.0988***	0.0986***	0.0986***
	(0.00494)	(0.00498)	(0.00492)	(0.00486)	(0.00486)
border	0.141***	0.153***	0.161***	0.164***	0.167***
	(0.0129)	(0.0136)	(0.0136)	(0.0135)	(0.0133)
Observations	65,345	67,578	69,410	70,713	73,029
R-squared	0.328	0.335	0.340	0.344	0.352
Distance-equivalent of the border effect	4.1	4.6	5.1	5.3	5.4

Notes: The dependent variable is the maximum of the absolute value of the log price difference between districts *j* and *k* in year *t* for bin *n* (using 500 distance bins). All regressions are estimated with a constant and year and product fixed effects as well as country-specific dummies for 'Botswana-Botswana', 'Malawi-Malawi' and 'South Africa-South Africa' district pairs (with the 'Zambia-Zambia' dummy omitted). Robust standard errors (reported in parentheses) are clustered by distance bin. Significance at the 10 percent, 5 percent and 1 percent levels is denoted by *, ** and *** respectively.

Table B.8: Comparison of quantile regression estimates – full sample versus highly homogenous products (accounting for country-heterogeneity effect)

	Full sample			Highly homogenous products			Excluding highly homogenous products		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	mean	p95	max	mean	p95	max	mean	p95	max
log distance	0.0399***	0.105***	0.107***	0.0620***	0.138***	0.141***	0.0257***	0.0852***	0.0859***
	(0.00289)	(0.00504)	(0.00513)	(0.00408)	(0.00628)	(0.00645)	(0.00302)	(0.00504)	(0.00508)
border	0.147***	0.177***	0.181***	0.142***	0.112***	0.119***	0.150***	0.224***	0.226***
	(0.00651)	(0.0127)	(0.0129)	(0.00826)	(0.0151)	(0.0155)	(0.00793)	(0.0140)	(0.0140)
Observations	92,631	92,631	92,631	35,808	35,808	35,808	56,823	56,823	56,823
R-squared	0.305	0.386	0.388	0.309	0.415	0.419	0.316	0.380	0.380
Distance-equivalent of the border effect	39.8	5.4	5.4	9.9	2.3	2.3	342.6	13.9	13.9

Notes: The dependent variable is the 95th percentile or the maximum of the absolute value of the log price difference between districts *j* and *k* in year *t* for bin *n* (using 500 distance bins). All regressions are estimated with a constant and year and product fixed effects as well as country-specific dummies for 'Botswana-Botswana', 'Malawi-Malawi' and 'South Africa-South Africa' district pairs (with the 'Zambia-Zambia' dummy omitted). Robust standard errors (reported in parentheses) are clustered by distance bin. Significance at the 10 percent, 5 percent and 1 percent levels is denoted by *, ** and *** respectively.

Table B.9: Comparison of quantile regression estimates – nearly homogenous versus perfectly homogenous products for Botswana-Zambia district pairs only (accounting for country-heterogeneity effect)

	Full sample of 24 nearly homogenous products			Perfectly homogenous products only		
	(1)	(2)	(3)	(4)	(5)	(6)
	mean	p95	max	mean	p95	max
log distance	0.0122***	0.0990***	0.101***	0.0121***	0.0683***	0.0725***
	(0.00260)	(0.00681)	(0.00701)	(0.00204)	(0.00564)	(0.00597)
border	0.100***	0.183***	0.201***	0.0626***	0.0760***	0.0957***
	(0.00580)	(0.0208)	(0.0221)	(0.00423)	(0.0136)	(0.0149)
Observations	48,703	48,703	48,703	14,978	14,978	14,978
R-squared	0.314	0.390	0.397	0.310	0.365	0.386
Distance-equivalent of the border effect	3,629.0	6.4	7.3	176.5	3.0	3.7

Notes: The dependent variable is the mean, 95th percentile or maximum of the absolute value of the log price difference between districts *j* and *k* in year *t* for bin *n* (using 500 distance bins). All regressions are estimated with a constant and year and product fixed effects as well as country-specific dummies for 'Botswana-Botswana', 'Malawi-Malawi' and 'South Africa-South Africa' district pairs (with the 'Zambia-Zambia' dummy omitted). Robust standard errors (reported in parentheses) are clustered by distance bin. Significance at the 10 percent, 5 percent and 1 percent levels is denoted by *, ** and *** respectively.

Table B.10: Comparison of estimates from pooled OLS, quantile and production-consumption pair regressions (for products where 80% or more of total imports in Botswana/Malawi/Zambia are sourced from South Africa)

	Pooled OLS	Quantile regression	Production-consumption pairs
	(1)	(2)	(3)
log distance	0.0670***	0.121***	0.144***
	(0.00339)	(0.00610)	(0.0195)
border	0.156***	0.252***	0.206***
	(0.00492)	(0.0156)	(0.0318)
Observations	96,151	19,829	1,626
R-squared	0.169	0.350	0.288
Distance-equivalent of the border effect	10.3	8.0	4.2

Notes: For the quantile regressions, the dependent variable is the maximum of the absolute value of the log price difference between districts *j* and *k* in year *t* for bin *n* (using 500 distance bins). All regressions are estimated with a constant and year and product fixed effects. Robust standard errors (reported in parentheses) are clustered by district pair for the pooled OLS regression and production-consumption pair regressions and by distance bin for the quantile regressions. Significance at the 10 percent, 5 percent and 1 percent levels is denoted by *, ** and *** respectively.

Appendix C

Table C.1: Estimation of average SADC border effect by year, 2006-2009

	(1) 2006	(2) 2007	(3) 2008	(4) 2009
log distance	0.0913***	0.112***	0.108***	0.112***
	(0.00485)	(0.00514)	(0.00469)	(0.00503)
border	0.344***	0.391***	0.325***	0.426***
	(0.0129)	(0.0121)	(0.0113)	(0.0112)
Botswana	0.152***	0.229***	0.199***	0.264***
	(0.0140)	(0.0135)	(0.0125)	(0.0132)
Malawi	0.183***	0.193***	0.176***	0.133***
	(0.0227)	(0.0190)	(0.0268)	(0.0364)
Zambia	0.155***	0.209***	0.215***	0.177***
	(0.0130)	(0.0123)	(0.0118)	(0.0114)
constant	-0.143***	-0.331***	-0.259***	-0.219***
	(0.0334)	(0.0341)	(0.0314)	(0.0344)
Observations	22,239	23,901	23,716	22,775
R-squared	0.389	0.402	0.393	0.412

Notes: The dependent variable is the maximum of the absolute value of the log price difference between districts *j* and *k* in year *t* for bin *n* (using 500 distance bins). All regressions are estimated with product fixed effects as well as country-specific dummies for 'Botswana-Botswana', 'Malawi-Malawi' and 'Zambia-Zambia' district pairs (with the 'South Africa-South Africa' dummy omitted). Robust standard errors (reported in parentheses) are clustered by district pair. Significance at the 10 percent, 5 percent and 1 percent levels is denoted by *, ** and *** respectively.

Table C.2: Estimation of individual border effects in the SADC region by year, 2006-2009

	(1) 2006	(2) 2007	(3) 2008	(4) 2009
log distance	0.0827***	0.0869***	0.0910***	0.104***
	(0.00496)	(0.00503)	(0.00478)	(0.00540)
Botswana-Malawi border	0.528***	0.539***	0.458***	0.431***
	(0.0193)	(0.0182)	(0.0178)	(0.0198)
Botswana-SA border	0.269***	0.299***	0.234***	0.428***
	(0.0136)	(0.0122)	(0.0115)	(0.0129)
Botswana-Zambia border	0.357***	0.430***	0.387***	0.415***
	(0.0163)	(0.0162)	(0.0155)	(0.0157)
Malawi-SA border	0.451***	0.643***	0.456***	0.554***
	(0.0239)	(0.0192)	(0.0201)	(0.0212)
Malawi-Zambia border	0.477***	0.412***	0.417***	0.339***
	(0.0176)	(0.0165)	(0.0172)	(0.0180)
SA-Zambia border	0.270***	0.431***	0.316***	0.473***
	(0.0181)	(0.0191)	(0.0175)	(0.0176)
Botswana	0.148***	0.208***	0.188***	0.255***
	(0.0137)	(0.0124)	(0.0119)	(0.0130)
Malawi	0.187***	0.185***	0.174***	0.128***
	(0.0209)	(0.0161)	(0.0253)	(0.0350)
Zambia	0.157***	0.204***	0.214***	0.174***
	(0.0128)	(0.0115)	(0.0114)	(0.0111)
constant	-0.0900***	-0.175***	-0.149***	-0.165***
	(0.0335)	(0.0333)	(0.0317)	(0.0362)
Observations	22,239	23,901	23,716	22,775
R-squared	0.412	0.424	0.413	0.417

Notes: The dependent variable is the maximum of the absolute value of the log price difference between districts *j* and *k* in year *t* for bin *n* (using 500 distance bins). All regressions are estimated with product fixed effects as well as country-specific dummies for 'Botswana-Botswana', 'Malawi-Malawi' and 'Zambia-Zambia' district pairs (with the 'South Africa-South Africa' dummy omitted). Robust standard errors (reported in parentheses) are clustered by district pair. Significance at the 10 percent, 5 percent and 1 percent levels is denoted by *, ** and *** respectively.

Table C.3: Production-consumption pair estimates, 2006-2009

	(1)	(2)	(3)	(4)
log distance	0.0849***	-0.917***	0.0133	-0.684***
	(0.0142)	(0.162)	(0.0119)	(0.197)
log distance squared		0.0758***		0.0560***
		(0.0126)		(0.0160)
border	0.155***	0.187***		
	(0.0197)	(0.0149)		
Botswana-SA border			0.148***	0.176***
			(0.0128)	(0.0149)
Malawi-SA border			0.620***	0.563***
			(0.0284)	(0.0343)
SA-Zambia border			0.257***	0.198***
			(0.0237)	(0.0316)
Botswana				
Malawi				
Zambia				
constant	-0.267***	2.973***	0.166**	2.305***
	(0.0887)	(0.520)	(0.0775)	(0.603)
Observations	7,514	7,514	7,514	7,514
R-squared	0.320	0.335	0.361	0.365

Notes: the dependent variable is the absolute value of the log price difference between district *j* and Pretoria (the production location) in year *t*. All regressions are estimated with year and product fixed effects. Robust standard errors (reported in parentheses) are clustered by district pair. Significance at the 10 percent, 5 percent and 1 percent levels is denoted by *, ** and *** respectively.

Appendix D

Table D.1: Mapping of products to HS codes and descriptions

Product	HS 6-digit code	HS 8-digit code	HS 8-digit description
baked beans	200559	20055900	Other Beans, Prepared or Preserved (Excluding By Vinegar; Not Frozen)
beef brisket	020230	02023000	Frozen, boneless meat of bovine animals
biscuits	190531	19053100	Sweet biscuits
cheddar cheese	040630	04063000	Processed cheese, not grated or powdered
cough syrup	300490	30049000	Medicaments consisting of mixed or unmixed products for therapeutic or prophylactic purposes
electric iron	851640	85164000	Electrical smoothing irons
electric kettle	851610	85161000	Electric instantaneous or storage water heaters and immersion heater
margarine	151710	15171000	Margarine (excl. liquid)
onions	070310	07031000	Onions and shallots, fresh or chilled
potatoes	070190	07019000	Fresh or chilled potatoes (excl. seed)
rice	100630	10063000	Semi-milled or wholly milled rice
rump steak	160250	16025000	Preparations of meat of bovine animals
shoe polish	340510	34051000	Polishes, creams and similar preparations, for footwear or leather, whether or not in the form of paper, wadding, felt, nonwovens, cellular plastics or cellular rubber, impregnated, coated or covered with such preparations
spaghetti	190219	19021900	Uncooked pasta, not stuffed or otherwise prepared, not containing eggs

Table D.2: Zambian preferential and MFN tariff rates for individual products, 2002-2009

			Tariff rate							
Product category	**Product**	**Tariff type**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
dairy products	cheddar cheese	MFN Zambia	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
		Pref. tariff Zam-SA	0.15	0.15	0.15	0.1	0.1	0.05	0	0
	margarine	MFN Zambia	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
		Pref. tariff Zam-SA	0.15	0.15	0.15	0.1	0.1	0.05	0	0
fruit and vegetables	onions	MFN Zambia	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
		Pref. tariff Zam-SA	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.15
	potatoes	MFN Zambia	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
		Pref. tariff Zam-SA	0.15	0.15	0.15	0.1	0.1	0.05	0	0
meat	beef brisket	MFN Zambia	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
		Pref. tariff Zam-SA	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.15
	rump steak	MFN Zambia	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
		Pref. tariff Zam-SA	0.15	0.15	0.15	0.1	0.1	0.05	0	0
non-perishable food products	baked beans	MFN Zambia	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
		Pref. tariff Zam-SA	0.15	0.15	0.15	0.1	0.1	0.05	0	0
	biscuits	MFN Zambia	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
		Pref. tariff Zam-SA	0.15	0.15	0.15	0.1	0.1	0.05	0	0
	rice	MFN Zambia	0.05	0.05	0.05	0.15	0.15	0.15	0.15	0.15
		Pref. tariff Zam-SA	0	0	0	0	0	0	0	0
	spaghetti	MFN Zambia	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
		Pref. tariff Zam-SA	0.15	0.15	0.15	0.1	0.1	0.05	0	0
durable consumer products	shoe polish	MFN Zambia	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
		Pref. tariff Zam-SA	0.15	0.15	0.15	0.1	0.1	0.05	0	0
medicine	cough syrup	MFN Zambia	0.05	0	0	0	0	0	0	0
		Pref. tariff Zam-SA	0	0	0	0	0	0	0	0

			Tariff rate							
Product category	**Product**	**Tariff type**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
household electronics	electric iron	MFN Zambia	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
		Pref. tariff Zam-SA	0.15	0.15	0.15	0.1	0.1	0.05	0	0
	electric kettle	MFN Zambia	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
		Pref. tariff Zam-SA	0.15	0.15	0.15	0.1	0.1	0.05	0	0

Note: *MFN Zambia* is Zambia's MFN tariff on imports from other members of the World Trade Organization, who are not party to a preferential trade agreement with Zambia. *Pref. tariff Zam-SA* is Zambia's preferential tariff offer for imports from South Africa.

Table D.3: South African MFN tariff rates for individual products, 2002-2009

		South African MFN tariff rate							
Product category	**Product**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
dairy products	cheddar cheese	0	0	0	0	0	0	0	0
	margarine	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
fruit and vegetables	onions	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15
	potatoes	0	0	0	0	0	0	0	0
meat	beef brisket	0.4	0	0.4	0	0	0	0	0
	rump steak	0	0	0	0	0	0	0	0
non-perishable food products	baked beans	0	0	0	0	0	0	0	0
	biscuits	0.25	0.25	0.25	0.25	0.25	0.25	0.21	0.21
	rice	0	0	0	0	0	0	0	0
	spaghetti	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3
durable consumer products	shoe polish	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15
medicine	cough syrup	0	0	0	0	0	0	0	0
household electronics	electric iron	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
	electric kettle	0	0	0	0	0	0	0	0

Table D.4: Average US dollar prices by product and country/region over the whole 2002-2009 period

Product	Gauteng	Rest of South Africa	Botswana	Zambia
baked beans	0.52	0.53	0.64	0.95
beef brisket	3.53	3.62	2.93	2.57
biscuits	0.54	0.59	0.95	1.03
cheddar cheese	4.92	5.29	6.88	10.17
cough syrup	2.51	2.46	2.60	1.01
electric iron	17.02	17.85	6.15	19.17
electric kettle	20.66	16.97	6.93	14.64
margarine	1.00	0.65	0.93	0.96
onions	0.89	0.73	0.79	0.98
potatoes	0.91	0.80	0.74	0.62
rice	1.19	1.27	1.27	1.55
rump steak	6.02	6.49	3.44	3.19
shoe polish	0.64	0.65	0.67	0.60
spaghetti	0.74	0.74	0.77	1.07

Table D.5: Average US dollar prices for a men's haircut by country/region, 2002-2009

	Gauteng	Rest of South Africa	Botswana	Zambia
Average US$ price across districts	6.03	5.11	1.92	0.45

Table D.6: Production-consumption pair estimate of the trend in the aggregate South Africa-Zambia border effect

ln(Gauteng price)	0.321***
	(0.0359)
log distance	0.638
	(0.423)
log distance squared	-0.0519
	(0.0326)
trend	0.0267***
	(0.00797)
$border_{SA\text{-}Zambia}$	0.311***
	(0.0683)
$border_{SA\text{-}Zambia}$*trend	0.0213***
	(0.00588)
ln(non-traded price)	0.136***
	(0.0256)
constant t	-2.246*
	(1.337)
Observations	3,729
R-squared	0.906

Notes: The dependent variable is the log price of product *i* in consumption district *c* at time *t*. All regressions are estimated with year and product fixed effects. Robust standard errors (reported in parentheses) are clustered by district pair. Significance at the 10 percent, 5 percent and 1 percent levels is denoted by *, ** and *** respectively.

Table D.7: South Africa's share in Zambia's total imports by product, 2002

Product	Import intensity from South Africa
potatoes	0.01
shoe polish	0.14
baked beans	0.18
onions	0.29
margarine	0.34
biscuits	0.36
electric iron	0.38
electric kettle	0.63
rice	0.77
cheddar cheese	0.85
spaghetti	0.88
rump steak	1.00
cough syrup	1.00
beef brisket	1.00

Note: These shares are calculated as the ratio of the US dollar value of Zambia's imports of product *i* from South Africa to the total value of Zambia's imports of that product from the world in 2002.

Figure D.1: Plots of simple average prices by country and year for selected products, 2002-2009