CSAE Working Paper WPS/2012-01

Market Integration and Border Effects in Eastern Africa

Bruno Versailles*

January 2012

Abstract

This paper studies border effects in Eastern Africa by exploiting a consumer price data set covering 4 out of 5 EAC member states, 39 cities and 24 goods over the period 2004-2008. The Law-of-One-Price (LOP) is tested by running level regressions on relative prices with city-pairs the unit of observation. Unsurprisingly, distance plays an important role in explaining relative price movements, both within and between countries. The border effect as measured by the coefficient on a border dummy is significant but smaller compared to what is usually found in the Engel-Rogers literature. Consequently, distance equivalents, at 300 to 6,000 km, are much lower (and more reasonable) than what most papers working in the Engel-Rogers tradition find. Neither the significance nor the size of the border effect is reduced by allowing for nominal exchange rate variation or non-tariffbarriers. The nominal exchange rate pass through is very high across the region, but lower for country-pairs that include Kenya. The advent of the Customs Union in 2005 improved market integration, as measured by a reduced border effect between 2004 and 2008, but only between Kenya and Uganda. This makes sense as Rwanda and Burundi only joined the EAC in 2007. Further, larger departures from the LOP are found during the Kenyan political crisis as regional markets were disrupted. When splicing the data across goods, staple markets are by far the most integrated with the lowest border effects and the lowest distance coefficients. Non-Trade barriers are most important for fruits and vegetables, which could be linked to these goods being perishable. Non-food items show the largest departures from the LOP, which can be attributed to these goods (i) being less comparable, and (ii) being less present in the consumption baskets of the average Eastern African households (i.e. less deep markets).

*I am currently at the Fiscal Affairs Department at the International Monetary Fund and can be reached at bversailles@imf.org. This paper is a reworked version of one of my DPhil chapters at Oxford University. I gratefully acknowledges financial support from the Webb-Medley Fund (Department of Economics, Oxford University) and Old members Travel Fund for Graduates (University College, Oxford University).

Centre for the Study of African Economies
Department of Economics · University of Oxford · Manor Road Building · Oxford OX1 3UQ
T: +44 (0)1865 271084 · F: +44 (0)1865 281447 · E: csae.enquiries@economics.ox.ac.uk · W: www.csae.ox.ac.uk

1 Introduction

The violation of the Law of One Price (LOP) across borders is one of the most enduring stylized facts in the international pricing literature. Obstfeld and Rogoff (2000) list the border effect among one of the six remaining puzzles of international macroeconomics and suggest trade costs as the main culprit. Trade costs broadly defined *'include all costs incurred in getting a good to a final user other than the marginal cost of producing the good itself'* (Anderson and Van Wincoop, 2004, p.691). Of course, trade costs impair the good functioning of markets also within countries. Non-border related trade costs arise *from distance and geographical irregularities interacting with the most efficient transport and communications technology* (Anderson and Van Wincoop, 2002). An often quoted estimate for average trade costs in industrialized countries is 170% (ad valorem equivalent), which breaks down into 55% local distribution costs and 74% international trade costs (Anderson and Van Wincoop, 2004, p.692). There is ample evidence that trade costs are much higher in sub-Saharan Africa, as African countries face considerably higher transport costs and have in general a weaker institutional set-up for dealing with issues like rules of origin and other non-tariff barriers (Portugal-Perez and Wilson, 2008).

Research on border effects using prices broadly falls into two categories. On the one hand there is the sizeable literature on exploring the time series properties of price series to test whether PPP holds[1]. On the other hand there is the more cross-section oriented research which looks at the cross-section properties of the price data. The classic reference for the latter is Engel and Rogers (1996) (ER) who estimate the economic impact of crossing the US-Canada border to be equivalent to shipping a good 75,000 miles[2]. Most authors working in the ER tradition find similarly implausibly large distance equivalents. However, more recent contributions show that these can be attributed to the methodology ER apply and properties of the used data. Thus, Broda and Weinstein (2008) find distance

[1] For an overview of this literature up to the mid-nineties, see Rogoff (1996). For more recent contributions see Crucini and Shintani (2008) or Broda and Weinstein (2008).c

[2] A different strand of the border effect literature uses quantities rather than prices in a gravity model. The classic reference here is McCallum (1995) who shows that the Canada-U.S. border has a decisive effect on continental trade patterns.

equivalents for the US-Canada border to be only 3 miles when working with good-level prices instead of aggregate price indices, while Gorodnichenko and Tesar (2009) show how the ER result to be entirely driven by cross-country heterogeneity in the distribution of within-country price differentials, rather than the between-country price differentials which the border effect is supposed to capture. Rather than working with the standard deviation of the first difference of real exchange rates as the dependent variable as in Engel and Rogers (1996), I use a level-regression specification which has been less popular in the literature because of higher demands on the data, notably the need for disaggregated good-level price data. The few papers that have used the levels-specification in general find much smaller distance equivalents, even though border dummies are still found to be significant[3].

This paper essentially tests the LOP using highly disaggregated good-level price data in 39 cities in four out of five member countries of the East African Community (EAC) and links the breakdown of the LOP to different types of trade costs. The maps in appendix A show the geographical dispersion of the cities[4]. The data comes from national statistics offices and are collected as part of the standard inflation (CPI) measurement process. 24 goods were selected which are comparable and representative for consumer patterns in the region. Hence, I try and bridge the good-specific price literature emanating from developing countries[5], which mostly focuses on micro/market structural issues, with the developed country literature, which uses highly disaggregated price data to learn about macro-issues such as border effects, trade patterns and exchange rate regimes[6]. The monthly data covers the period 2004-2008, which allows for a potential customs union effect of increased integration, especially between Uganda and Kenya[7]. One other ex-

[3] Examples of papers using this levels-specification are Broda and Weinstein (2008) and Engel et al. (2005) for the US-Canada border, Grafe et al. (2008) for five central Asian countries, and Aker et al. (2010) for the Niger-Nigerian border. Of these, Engel et al. (2005) is the only paper that finds a distance equivalent similar in size to the ER-literature.

[4] It was not possible to obtain similar data for Tanzania, despite a visit to the National Bureau of Statistics and numerous follow-up efforts.

[5] See Fackler and Goodwin (2001) for an overview of that literature or Rashid and Minot (2010) for a focus on Africa.

[6] As in e.g. Broda and Weinstein (2008).

[7] The customs union between Uganda, Kenya and Tanzania went into effect in January 2005, and was extended to Burundi and Rwanda in 2009.

ternal 'event' that might have a bearing on market integration is investigated: the Kenya election of December 2007 and the ensuing violence in the first months of 2008. As Kenya is a transit country for many goods into landlocked Uganda, Rwanda and Burundi[8] markets in these countries were severely disrupted. This is the first paper that tries to measure border effects in an African context, using *more* than a handful of (typically) food-related goods. Another novelty is the use of data on Non-Tariff Barriers using the World Bank's Doing Business database.

I find the border coefficient to be significant, moving prices between 13 and 20 percent away from the LOP benchmark. Further, a distance of 100 km between two cities adds another 13 percent to departures from the LOP. This is much larger than the distance coefficients reported in Broda and Weinstein (2008) or Engel et al. (2005) for the US-Canada border. Hence my distance equivalents are much lower than in the ER literature, but this is mainly because the distance coefficient is large. The message is clear: borders play a role in Eastern Africa, but distance, both between and within countries, is a much more fundamental aspect of trade compared to Northern America. Border coefficients related to borders of neighbouring countries are sensibly lower than those for countries that are further apart. Neither the border coefficient nor the distance coefficient changes substantially when adding variables like a non-tariff barrier index or the nominal exchange rate. Neither of these variables have much explanatory power. The coefficient on the nominal exchange rate is very low implying a high pass through throughout the region. The advent of the Customs Union in 2005 improved market integration, as measured by a reduced border coefficient between 2004 and 2008, but only between Kenya and Uganda. Further, larger departures from the LOP are found during the Kenyan political crisis as regional markets were disrupted. When splicing the data across goods, staple markets are by far the most integrated with the lowest border and distance coefficients. Non-Trade barriers are most important for fruits and vegetables, which could be linked to these goods being perishable. Non-food items show the largest departures from the

[8] The major overland trade arteries in Eastern Africa are the northern and the central corridors (see map A3). Currently, regional traffic is concentrated along the better developed northern corridor (World Bank (2010)).

LOP, which can be attributed to these goods (i) being less comparable, and (ii) being less present in the consumption baskets of the average Eastern African households (i.e. less deep markets).

The paper is structured as follows. After this introduction, section 2 introduces a framework that will help to set up the LOP testing strategy. Section 3 introduces the price data set. It includes a discussion of how the 24 goods were retained for the empirical analysis and an assessment of how comparable these goods are across the four countries. Appendix B gives more detail on the dataset and explains how missing observations were dealt with. The section ends with some summary statistics. Section 4 derives the main empirical specification and discusses the explanatory variables to be used in the regression analysis of section 5. Section 6 concludes.

2 Purchasing Power Parity and the law of one price

Purchasing Power Parity (PPP) states that, once converted to a common currency, price levels should be equal across countries. The underlying mechanism is that if goods market arbitrage manages to enforce parity in prices across a sufficient range of goods (the law-of-one-price or LOP), then one should also see high correlation in aggregate price levels. The concept of *Absolute LOP* thus refers to the price of identical goods being the same across locations when denominated in a common currency:

$$P_{uct} = E_{cc't}P_{uc't} \tag{1}$$

with P_{uct} the price of good u in city c at time t, $P_{uc't}$ the price of the same good in a different city c' and $E_{cc't}$, the nominal exchange rate, which equals one when both cities are in the same country. When testing whether (1) holds, researchers often have to work with data from different country sources. It is likely then that these countries do not define goods in exactly the same way. Hence, a lot of tests are actually based on a slightly different version of (1):

$$P_{uct} = E_{cc't}P_{u'c't} \tag{2}$$

If this equation holds, *Approximate Absolute LOP* holds, with u and $u^{'}$ belonging to a similar, but not identical product category[9]. If departures from the LOP are then found in the statistical analysis, a genuine border effect could be confused with using dissimilar goods. This is especially likely when using data from different countries, which normally come from different sources. It also shows the problem with using price indices instead of good-level prices, as indices like the CPI are constructed with different goods and different weights in different countries. Alternatively, the *relative LOP* hypothesis can be tested, which implies looking at first differences of log price levels, and testing whether prices tend to remain a constant level apart, where lower case letters describe the logs of upper-case variables:

$$\Delta p_{uct} = \Delta e_{cc't} + \Delta p_{uc't} \tag{3}$$

Going back to equation (1), the real exchange rate for good u at time t can be defined as:

$$Q_{ucc't} = \frac{P_{uct}}{E_{cc't} * P_{uc't}} \tag{4}$$

If absolute LOP holds, then equation (1) holds, and expression (4) is equal to one.

From a theoretical perspective, the breakdown of PPP/LOP, both within and across countries, can have many origins. Researchers have found it useful to make a distinction between 'real' and 'nominal' sources of this breakdown[10]. 'Real' sources include:

1. **Transportation Costs**, often proxied by distance between regions or cities, are associated with moving goods from one region to another.

[9] A recent paper mentions that in the US most product groups (e.g. eggs) are comprised of over 200 different Universal Product Categories (UPC's or barcodes). For eggs for example, there are 2275 different barcodes (Broda and Weinstein, 2008, p10).

[10] For a more in-depth discussion of these issues, see Engel and Rogers (2001) or Cecchetti et al. (2002).

2. **Trade Barriers**, including tariff and non-tariff barriers.
3. **Presence of non-traded goods in the general price level**. Intuitively, one would expect arbitrage-type pressures to be smaller for non-tradable goods. Crucini and Shintani (2008), exploiting the time series dimension of micro-price data, find some evidence for this, but the results from papers working in the Engel-Rogers border effect tradition are largely inconclusive. Of course this is not a 'root cause' for departures from the LOP benchmark. It is the characteristics of goods in terms of transport costs, tariffs and NTBs that actually make a good non-tradable, with tradability being a matter of degree rather than purely dichotomous. Still, it is instructive to keep in mind not to expect the LOP to hold for 'goods' such as haircuts.
4. **The existence of non-traded inputs**, like marketing and distribution services that are part of final goods prices. One of the reasons that researchers have not been able to come up with conclusive evidence related to the previous source of PPP/LOP breakdown could be that a substantial amount of non-traded inputs go into certain traded goods. A famous paper by Sanyal and Jones (1982) model all retail (consumer) goods as a composite of traded and non-traded inputs. Final goods are then all essentially non-tradable, due to costs associated with retail/distribution/marketing-type services.
5. **Imperfect Competition**. There is a large literature arguing not to expect equality of prices across different locations because producers set different prices, for example because they have local monopoly power in segmented markets (pricing to market).

The main <u>'nominal'</u> cause of the PPP/LOP breakdown is related to the nominal exchange rate. Floating exchange rates, combined with sticky prices can incite producers selling abroad to set prices in the currency of the (foreign) consumer rather than their own. This is called incomplete exchange rate pass through or Local Currency Pricing (LCP). If LCP applies, changes in the nominal exchange rate do not affect goods prices in the local market. Feenstra and Kendall (1997) find evidence for this in explaining deviations from LOP and most border effect studies find this channel to be an important determinant of observed price dispersion. Intuitively, fixing exchange rates between countries would thus entail welfare gains, as deviations to the LOP would be minimised. However, *overall* welfare does not increase necessarily, as the *inefficiency (and hence the welfare loss) is due to price stickiness, and the choice of exchange rate regime does not necessarily alter that* (Engel and Rogers, 2001, p32)[11]. Other welfare issues to be considered are the standard loss of monetary policy autonomy under fixed exchange rates and the ease of relative price adjustment under floating exchange rates. Devereux and Engel (1998) contrast these welfare trade offs with the 'real' causes of the failure of the LOP, which can be interpreted unequivocally as deadweight welfare losses.

3 Consumer price data in eastern Africa

The CPI dataset covers 39 cities in four out of five of the EAC member countries—see the maps in appendix A. For three cities in Kenya and one city in Kampala, more than one price series are available, linked to income category and/or geographical location within major cities. Monthly time series are available from October 1997 (Uganda and Kenya), January 2003 (Burundi) and January 2004 (Rwanda). All series run until August 2008, except for Rwanda where the series end in July 2008. This means the four datasets have 55 months (Jan-04 to Jul-08) in common. Appendix B discusses the dataset in more detail with table 7 giving a detailed overview of what the data exactly covers. Table 1 shows the 24 goods from the data set retained for the analysis in the remainder of this paper. These goods were chosen according to the following criteria:

- Data availability across countries (only 'domestique' is not available in all countries).

[11] Devereux and Engel (1998) show how the underlying sticky prices inefficiency makes fixed exchange rates increase the variance of average world consumption.

- Data availability across cities. For Rwanda and Burundi, an imputation method was used to complete data for some good-city-month series (see section B.2 of appendix B).
- Comparability of goods across countries (easiest for food items, harder for most other goods).
- Low persistence of the series. This is particularly a problem for small manufactured goods such as matches, batteries, etc. It is also an issue for services such as doctor fees, education fees, etc. The latter are less surprising as such fees are often set by the public sector. Goods with these characteristics were not retained.

Broda and Weinstein (2008) find an important bias in border effect estimates when non-similar goods are compared. As table 1 shows, this comparability problem will probably be lower for staple foods and fruits & vegetables than for goods in the 'Other' category. Indeed for the first two categories, the original names of the staple foods look very similar. However, for other items the correspondence is less than perfect, e.g. 'beer' is actually Primus beer in Burundi and Rwanda, Tusker in Kenya, whilst there are no details beyond the label 'Brewed bottled Beers' in Uganda.

To start with, I deal with this by omitting high-income price series in Kenya (three areas located in Nairobi) and Uganda (Kampala high-income), as this makes the goods more comparable. Why? At the lower ends of the income distribution, product diversification is normally less pronounced, as good attributes such as packaging, shop location, etc. have less impact on pricing. E.g. shopping for even quite basic goods like staple foods, in a supermarket like Nakumatt, makes quite a big difference price-wise as compared to buying from market stalls. For luxury goods, this effect will be even more pronounced[12].

[12] A nice example from the Kampala high-income series sums up both the high-income claim and the fact that items in the 'other' category will be less comparable. The 'elementary aggregate' shown in the data set for the item 'wine' in July 2008 is 24,800 Ugandan Shillings. Unpacking this number, it is effectively an arithmetic average of the prices of 8 different wines, staggered across four different shops in Kampala, with a standard deviation of just under 8,000 USh. The type of shop where these price quotes were gathered varied substantially: from a top-class hotel to an Uchumi supermarket, whilst the quality of the wine was also substantially different across the eight quoted prices.

Table 1: Matching 24 Goods Across 4 Countries - Data Availability per Country[a,b]

Common name	C[c]	BURUNDI		KENYA		RWANDA		UGANDA	
		Original Name (unit)	C	Original Name (unit)	C	Original Name (unit)	C	Original Name (unit)	C
STAPLES (5)									
RICE	44	Rice - White (1 kg)	7	Rice (1 kg)	22	Rice - local (1 kg)	9	Rice (1 kg)	6
MAIZE FLOUR	43	Maize Flour (1 kg)	7	Maize Flour (2 kg)	22	Maize Flour (1 kg)	8	Maize Flour (2 kg)	6
MAIZE GRAINS	35	Dry Maize Grains (1 kg)	5	Maize Grains (1 kg)	22	Dry Maize Grains (1 kg)	7	Maize Grains (1 kg)	1
POTATOES	45	Potatoes - local (1 kg)	7	Potatoes - local (1 kg)	22	Potatoes - local (1 kg)	10	Irish Potatoes (1 kg)	6
SWEET POTATOES	43	Sweet Potatoes (1 kg)	7	Sweet Potatoes (1 kg)	20	Sweet Potatoes (1 kg)	10	Sweet Potatoes (1 kg)	6
FRUITS and VEGETABLES (7)									
ORANGES	38	Oranges (1 kg)	5	Oranges (1 kg)	22	Oranges - local (1 kg)	5	Oranges (1 kg)	6
ONIONS	45	Onions (1 kg)	7	Onions (1 kg)	22	Onions - Red (1 kg)	10	Onions - Big (1 kg)	6
CABBAGES	43	Cabbage - white (1 kg)	7	Cabbages (1 kg)	22	Cabbage - white (1 kg)	8	Cabbages (1 kg)	6
BANANA RIPE	45	Ripe Bananas - Ibigurube (1 kg)	7	Banana Ripe (1 kg)	22	Ripe Bananas - Kamara (1 kg)	10	Banana Bogoya (1 kg)	6
BANANA COOKING	45	Green Bananas (1 kg)	7	Banana Cooking (1 kg)	22	Bananas Cooking - local (1 kg)	10	Banana Ndizi (1 kg)	6
TOMATOES	45	Tomatoes (1 kg)	7	Tomatoes (1 kg)	22	Tomatoes - Fresh (1 kg)	10	Tomatoes - Big (1 kg)	6
CARROTS	44	Carrots (1 kg)	6	Carrots (1 kg)	22	Carrots (1 kg)	10	Carrots (1 kg)	6
OTHER FOOD ITEMS (6)									
SUGAR	45	Sugar - local (1 kg)	7	Sugar - local (1 kg)	22	Sugar - imported (1 kg)	10	Sugar - brown(1 kg)	6
SALT	45	Kitchen Salt - imported (1 kg)	7	Table Salt (500 gr)	22	Table Salt - imported (1 kg)	10	Salt (500 gr)	6
EGGS	45	Eggs - Fresh (unit)	7	Eggs (unit)	22	Eggs - Fresh (unit)	10	Eggs (2 units)	6
MARGARINE	42	Blue Band from Kenya (250 gr)	7	Blue Band (500 gr)	22	Blue Band (500 gr)	7	Margarine (500 gr)	6
FISH NOT FRESH	35	Dry Ndagara - local (1 kg)	7	Fish Non Fresh (1 kg)	15	Ndagara - imported (1 kg)	7	Fish - Dried Tilapia (1 kg)	6
MILK	41	Milk - non-pasteurized (1 l)	7	Milk - Fresh Packed (0.5 l)	21	Milk - Fresh Packed (1l)	7	Milk - Fresh (1 l)	6
OTHER (6)									
BEER	43	Primus (72 cl)	7	Tusker (500 ml)	20	Primus (72 cl)	10	Brewed bottled beers (0.5 l)	6
SODA	45	FANTA (30 cl)	7	FANTA (30 cl)	22	Fanta citron (30 cl)	10	Bottled Soda (30 cl)	6
CIGARETTES	43	Supermatch - local (20 units)	7	Sportsman (1 packet)	21	SM (1 packet)	9	Cigarettes (1 packet)	6
CHARCOAL	44	Charcoal (1 kg)	7	Charcoal (1 kg)	21	Charcoal (1 kg)	10	Charcoal (sack/kg)	6
DOMESTIQUE	31	Cost of a domestic helper in the popular quarter	7	Maid (per month)	15	House Worker	9	N/A	0
PETROL	34	Petrol (1 l)	7	Petrol - Super (1 l)	15	Petrol - Super (1 l)	6	Petrol (1 l)	6

[a] Names of goods from the Burundi and Rwanda CPI baskets are translated from French.
[b] When countries employ different units, prices are recalculated to the unit that is most common.
[c] C stands for number of cities that report price data for the respective good.

Table 2: Weights in the CPI basket of the 24 retained goods (all in percentages)[a]

Good Groupings	Burundi	Kenya[b]	Rwanda	Uganda	Average	KEN High Income[c]	UGA High Income[d]
Staples	7.5	10.3	8.6	6.0	8.1	4.3	2.8
Fruits & Vegetables	4.9	4.2	4.6	1.9	3.9	2.8	1.2
Other Food Items	4.7	12.7	5.3	6.6	7.3	7.9	3.8
Other	10.9	4.4	6.9	7.7	7.5	2.4	6.2
% of total CPI basket	**28.0**	**31.6**	**25.4**	**22.2**	**26.8**	**17.4**	**14.0**
food items as % of CPI food basket	**42.8**	**35.4**	**45.3**	**47.0**	**41.1**	**48.4**	**42.9**

[a] These are % taken from a total CPI basket summing to 100.
[b] Average of Nairobi Low-Income basket and non-Nairobi basket.
[c] Nairobi High Income basket.
[d] Kampala High Income basket.

Another way of showing that higher income groups have more diversified consumption patterns is by looking at their consumption baskets (table 2). In Kenya the lowest number of items for which price quotes are available is in arguably the poorest area (the Kibera-slum in Nairobi), whilst the highest is for Mombasa city centre. The penultimate row of table 2, which has the weights attached to consumption goods in the respective countries' CPI baskets, shows that the 24 goods account for a lower percentage of the total CPI basket in the high-income series compared to the low income series.

How representative are these goods? Is there a case for saying that conclusions reached from analysing the 24 retained items, can be generalised to how integrated these economies are at a macro-level? The diet within the EAC across countries is quite similar, i.e. tastes are not so different, with more similarity found for poorer households. Table 2 has some evidence on this, with 23 out of 24 goods consumed in all four countries, making up on average 27% of the (low-income) total CPI basket. The 18 food items make up on average 41% of the CPI food basket (for the low-income groups).

How much are these goods traded across the region? A quote from a recent study on Non-Tariff Barriers indicates that the 24 goods are capturing most of the heavily traded goods in the region: *Food and live animals as a group continue to dominate the formal intra-EAC trade of almost all EAC members, except the exports of Kenya (..). The two other important groups - especially for the new EAC members - are beverages and tobacco,* *and inedible crude materials* (World Bank, 2008, p.iv). Anecdotal evidence on informal trade also suggests the majority of the 24 goods to be heavily traded across borders (Bank of Uganda and UBOS, 2007).

Table 3 shows summary statistics of the absolute value of the log of good real exchange rates $Q_{ucc't}$ (see equation 4 above), which is the dependent variable in the subsequent regression analysis, across country-pairs, good groupings and time. A zero entry in the table implies the LOP hypothesis holds perfectly (remember we are taking logs) with higher values showing larger deviations from the LOP. The average mean for within-country country-pairs is 0.243. This implies that the absolute value of the natural log of the relative price (expressed in the same currency) is $e^{0.243}$ or 1.275. I.e. 27.5% higher than the LOP benchmark. For the highest between-country pair (Burundi-Kenya) this rises to 66% ($e^{0.509}$). Grafe et al. (2008) cite very similar numbers for central Asian countries, with the departure from LOP for their set of 31 goods varying between 30-50%. Aker et al. (2010) report slightly smaller numbers for the Niger/Nigeria border, but they only report price differentials for locations within 50 km of the border (see their Figure 2), and still report peaks in the difference in log of prices of more than three.

Both the mean and the standard deviation of city-pairs from the same country are substantially lower compared to city-pairs from different countries. Within the latter category the mean and the standard deviation are higher for countries that do not share a border, with the two countries furthest apart geographically (Burundi and Kenya) having the highest deviations from the LOP. Further, the means for between-country city-pairs are higher for all four good categories. The last two columns show means before and after the Kenyan election crisis, to investigate possible knock-on effects of the crisis for the wider (landlocked) region. For city-pairs involving a Kenyan city, mean real exchange rates are on average 12% higher in 2008 compared to the 2004-2007 period. For most other country-pairings, the 2008 means are also higher.

Table 3: Overview good real exchange rates by country pair and good-category[a]

	Obs	Mean	St.dev.	Max-Min	Mean (pre-2008)[d]	Mean (2008)[d]
WITHIN-COUNTRY	**367,675**	**0.243**	**0.281**	**2.59**	**0.241**	**0.256**
By Country-pair (avg. distance)						
BUR-BUR (138 km)[b]	26,180	0.253	0.257	1.82	0.254	0.250
KEN-KEN (392 km)	276,155	0.246	0.293	2.59	0.244	0.264
RWA-RWA (163 km)	47,190	0.226	0.237	2.08	0.226	0.223
UGA-UGA (317 km)	18,150	0.221	0.244	1.45	0.219	0.235
By good category[c]						
Staple Food	79,365	0.231	0.248	1.87	0.229	0.245
Fruits and Vegetables	116,325	0.396	0.327	2.59	0.395	0.397
Other Food items	90,915	0.142	0.197	2.20	0.140	0.155
Other	81,070	0.149	0.221	1.64	0.145	0.179
BETWEEN-COUNTRY	**779,790**	**0.443**	**0.379**	**3.28**	**0.436**	**0.492**
By Country-pair (avg distance)						
BUR-KEN (1575 km)	186,450	0.509	0.411	3.10	0.499	0.589
BUR-UGA (912 km)	50,105	0.425	0.339	2.28	0.424	0.436
KEN-RWA (1323 km)	244,420	0.467	0.401	3.28	0.455	0.549
BUR-RWA (302 km)	79,750	0.363	0.310	2.37	0.363	0.364
KEN-UGA (792 km)	154,000	0.391	0.341	2.47	0.390	0.392
RWA-UGA (658 km)	65,065	0.405	0.348	2.48	0.400	0.442
By good category[c]						
Staple Food	159,170	0.401	0.412	3.10	0.387	0.498
Fruits and Vegetables	241,835	0.528	0.405	3.28	0.519	0.589
Other Food items	197,615	0.372	0.295	2.20	0.371	0.383
Other	181,170	0.445	0.373	2.23	0.441	0.479

[a] The dependent variable is the absolute value of the log of good real exchange rates. The statistics shown are the mean, standard deviation, and the difference between the maximum and the minimum of the series, averaged over time and across country-pairs or good categories (as indicated).
[b] In brackets is the average distance between cities for the respective country-pair, in km.
[c] For the composition of these groups, i.e. which items belong to which group, see table 1.
[d] When the figure in this column is underlined, it implies the means of the respective real exchange rates before and after 2008 are significantly different at the 1% level (through performing a standard t-test).

4 Estimation Strategy and Dependent Variables

4.1 Deriving reduced form regression specifications

Good-level real exchange rates $Q_{ucc't}$ are now linked to their determinants, to end up with reduced form regressions to be estimated. Allowing for trade costs and differences in currency between locations, define $T_{ucc't}$ as the iceberg cost of trade between locations c and c' per unit of good u. Then the LOP effectively boils down to defining the following 'band of inaction':

$$\frac{1}{T_{ucc't}} \leq Q_{ucc't} \leq T_{ucc't} \tag{5}$$

Equation (5) implies that the price difference between our two locations cannot be higher than the cost of trading. Note that the cost of trade here is defined as everything that would make equation (4) not hold, i.e. all the reasons given in the previous subsection for the LOP breakdown are subsumed into the T-term. A consequence of the assumption of iceberg trade costs[13] is that $q_{ucc't}$ is uniformly distributed across the band of inaction $[-t_{ucc't} \ , \ t_{ucc't}]$[14]. Trade costs at time t could then be specified as:

$t_{ucc't}$= f(distance, border, tariffs, NTBs, E, trade costs specific to c or c')

NTBs are Non-Tariff Barriers and E is (as before) the nominal exchange rate. How can this be related to the real exchange rate? The simplest way, equivalent to testing an absolute version of the LOP, is assuming that $p_{uct} - e_{ucc't} - p_{uc't} = f(t_{ucc't})$, with the deviation of the real exchange rate from unity revealing the size of the trade costs. A log-linear regression specification capturing long-run deviations from the LOP is then:

$$\begin{aligned} \left| ln\, Q_{ucc't} \right| = \beta_0 + \beta_1 \, ln\, d_{ucc'} + \beta_2 \, B_{ucc'} + \beta_3 \, TARIFF_{ucc't} + \\ \beta_4 \, NTB_{ucc't} + \beta_5 \, ln\, E_{cc't} + \sum_{k=1}^{m} \gamma_k D_k + \epsilon_{ucc't} \end{aligned} \tag{6}$$

[13] There are other specifications possible, and the iceberg assumption has come under some criticism recently. For an overview see Anderson and Van Wincoop (2004).
[14] Again, lower-case letters indicate logs.

The unit of observation is city-pairs, with $ln\ d_{ucc'}$ the log of the distance between locations c and $c^{'}$ and $B_{ucc'}$ a dummy variable which takes the value of one if locations c and $c^{'}$ are in different countries and zero otherwise[15]. The D_k's are m city dummies which take on a value of one for D_c and $D_{c'}$ and a value of zero for all other locations. This is to pick up any city fixed effects. $\epsilon_{ucc't}$ is the error term. This is our central estimation equation, to which I will return at several points in the paper.

This brings us to the standard specification used in the Engel-Rogers price-based border effect literature, which uses the standard deviation of the first difference of the LHS of (6) as the dependent variable[16]:

$$st.\, dev.\, (\Delta q_{ucc't}) = \lambda_0 + \lambda_1\, ln\ d_{ucc'} + \lambda_2\, B_{ucc'} + \sum_{k=1}^{m} \gamma_k D_k + \epsilon_{ucc't} \qquad (7)$$

Gorodnichenko and Tesar (2005) give two reasons as to why Engel & Rogers (ER) construct the LHS variable like this. First, taking the standard deviation implies that the dependent variable becomes scale invariant, i.e. it does not matter in which units the components that make up $q_{ucc't}$ are measured, whilst E($q_{ucc't}$) does not have to be zero. This transformation is necessary because Engel & Rogers work with price indices and not with good-level prices[17]. Second, ER use the standard deviation of the *first difference* of real exchange rates. The reason for this is that CPI indices are persistent, which could bias border effect estimates by confounding price shocks with price movements related to prices moving gradually back to their equilibrium. If the real exchange rates follow a random walk, then this confusion would be eliminated by differencing[18]. In general, using indices like the CPI would only permit to investigate short-run deviations to the LOP anyway, as the normalization involved makes a levels regression meaningless.

[15]The distance and border variables are considered time invariant. This might be less than obvious in times or places where borders change (see e.g. the papers by Morshed (2003) or Morshed (2007)) or where road-quality measures are incorporated into the distance measure.

[16]I have left out variables for tariff and non-tariff barriers and nominal exchange rate variation, as most studies in this vein do not include these.

[17]As price indices are *normalised* averages of prices for different goods.

[18]Gorodnichenko and Tesar (2005) call this the persistence effect. See pages 9-12 of their paper for an in-depth treatment.

What are the implications of this for our own work on East Africa? First of all, our data set does have good-level prices, whilst persistence is, at least for some of the goods I will be analysing, quite low. Hence, there is no immediate reason for us to emulate the Engel-Rogers specification. Instead I will use our monthly price data to estimate an equation like (6), which will give us more intuitive results regarding departures from the LOP and its main determinants. Some reflections on the time series properties of the time-varying variables in equation (6) are given in section (4.4) which introduces the data used for the nominal exchange rate.

4.2 Distance

The maps in appendix A give the reader a feel for distances between countries and cities. Distance is measured in kilometer. The first column of table 3 shows the average distance between two cities by country-pair. I experiment with four different distance measures[19]. The first distance measure is the log of the great circle distance between two cities[20]. The second metric is road distance which calculates the road distance between two cities[21]. Road distance follows the Northern Corridor roads (map A3 in appendix A.). For both these measures, distances within Kenyan cities are set equal to zero (remember for three cities in Kenya more than one price series are available—see section B.1 in appendix B. Measures three and four are linked to the previous straight distance and road distance variables respectively, but now with distances within Kenyan cities different from zero. All these within-city distances are calculated as straight lines; with the distances between these cities in the sample the same as for measures one and two. All these distance measures are highly correlated, with correlation coefficients between all possible pairs higher than 0.95 and significant at the 0.1 % level. These various metrics will be used for assessing the robustness of the regression analysis.

[19] In each case, logs are taken to reflect the consensus that the relationship between distance and price dispersion is concave.

[20] In developing this metric, I used the googlemaps distance measurement tool.

[21] From the East Africa Road Map by Freytag and Berndt.

4.3 Non-tariff barriers

Even though the protocol for the establishment of the EAC Customs Union explicitly addresses the importance of removing NTBs to foster regional integration (EAC, 2004, p16), they are still prevalent. For example, a recent World Bank report documents how *the formal grains trade include regulations requiring proof of origin, phyto-sanitary certificates, and invoices from vendors, 'required side-payments' to police at road-blocks and border crossings* (World Bank, 2008, p38). The same paper also reports on *The Dairy Wars*, which refers to cross-border trading problems in the EAC in that sector. Other research mentions the importance of avoiding delays in the issuance of certificates of origin for traders of perishable products, such as dairy products, fruits and vegetables (Mugisa et al., 2008, p56).

One of the innovations in this paper is that the regression analysis includes a variable that captures non-tariff barriers (NTBs). This variable is constructed using the three indicators on *Trading Across Borders* from the World Bank's annual Doing Business report[22]. The drawbacks of these indicators for our purposes are two-fold: (i) data is only available from 2005 onwards[23], and (ii) the data is yearly and not monthly. In the regression analysis I use an index-type variable constructed out of the three variables. In a first step, indices are constructed for each of the three variables individually, with Burundi-2005 the benchmark (i.e. the value of each indicator is set at 100 for Burundi-2005). After this, the indices for two different countries are added to get values per country-pair—or are doubled if cities are part of the same country. Between 2005 and 2009, progress was made throughout the region on indicators related to *number of documents* and *time*, but less so on the container price.

[22] (i) Number of documents entrepreneurs/traders have to fill out for clearance before they can import/export a good, (ii) Number of days it takes to import/export, and (iii) the cost of a container in USD.

[23] Technically, these indicators are only available from 2006. However, the time at which these indicators are collected are often more than half a year before that. Hence, the data points I construct bring the data 1 year forward.

4.4 Nominal exchange rates

All prices are converted into US Dollars, using monthly averages of local currency/USD nominal exchange rates from the IMF's International Finance Statistics (IFS). The log of these averages is used as the nominal exchange rate regressor. When estimating an equation like (6), the time-varying variables on the LHS (good real exchange rates) and RHS (nominal exchange rate) should be of similar orders of integration[24] for the regression results to make sense. Instead of exhaustive unit root testing[25], some arguments are given as to why this should not be a major problem in the proposed estimation strategy.

First, there is only one time-varying variable on the RHS of equation (6), i.e. time series issues have only ramifications for the regression specifications including the nominal exchange rate. Second, as the analysis in this paper essentially tries to explain the cross-section variation, monthly time dummies are included in all regression specifications to try and deal with time-specific heterogeneity. Third, one of the criteria for selecting the 24 retained goods was their persistence characteristics. Time series of food prices for example, are in general less persistent and thus less likely to contain unit roots.

Even if unit root tests were to be performed, these can be anticipated to be inconclusive for two reasons: (i) testing good real exchange rates for a unit root is in itself a test for market integration. Hence, one would expect the LOP to hold for some goods and city-pairs but not for others, depending on good and city characteristics. (ii) Methodologically, *unit root tests are now well known to have very low power, and to be unable to distinguish between random-walk behaviour and very slow mean-reversion in the PPP-consistent level of the real exchange rate* (Caporale and Gil-Alana, 2010, p10). Hence, as our time series is quite short, it is unlikely that meaningful results would emerge from unit root tests. Concluding, some units within the different possible panel dimensions (city-pairs and goods), will probably exhibit unit root behaviour, but this will not consti-

[24] As I only have annual data for our Doing Business variable, it should be interpreted more like a dummy variable.

[25] One could of course subject the LHS of equation (6) to a battery of unit root and co-integration tests, which would be an alternative for testing for market integration, using the time dimension of the data-set.

tute a major problem for drawing conclusions across our data-set.

When testing the nominal exchange rate for unit roots, using the Dickey Fuller Generalised Least Squares (DFGLS) univariate testing procedure[26], four out of six of the bilateral exchange rates were found to contain a unit root, with only the KEN-UGA and RWA-UGA country-pairs rejecting the unit root null. When deploying panel unit root tests, one get conflicting messages, as predicted by Caporale and Gil-Alana (2010) cited above. Using the Pesaran (Pesaran, 2007) or Fischer (Maddala and Wu, 1999) tests, the null of a unit root cannot be rejected, whereas the Hadri test (Hadri, 2000) cannot reject the null of stationarity. Even though nominal exchange rates are mostly thought to be I(1), there are authors who find long memories in nominal exchange rates (see e.g. Gil-Alana (2003) or Caporale and Gil-Alana (2010)). In the African context, Odedokun (2000) finds a non-trivial number of nominal bilateral USD exchange rates to be stationary. Concluding, some bilateral exchange rates in the data will exhibit I(1) behaviour, but this will not play too important a role. By trying to learn LOP and PPP lessons at the aggregate level by exploiting a lot of micro price data, there will always be some difficult methodological choices, as one cannot model all the individual goods and city-pairs—unlike for example the crop-specific work.

5 The border effect - regression analysis

5.1 Dealing with country heterogeneity

Before running the actual regressions, one last criticisms is that most papers using the ER methodology fail to accommodate for the country heterogeneity effect: *If relative price variability across locations within the same country differs systematically country-by-country, the border effect measured by a regression comparing within-country and cross-*

[26] As proposed by Elliott et al. (1996).

country price dispersion will be confounded by the divergence between two countries' internal price distributions (Gorodnichenko and Tesar, 2009, p220).

Appendix C shows how non-parametric kernel distributions fit our data, and whether city-pair relative prices exhibit different distributions conditional on which country-pair they belong to. Figure C.1 shows how within-country price distributions are very similar, with the Rwanda and Kenya distributions almost impossible to distinguish from each other. This eases our worries regarding the country heterogeneity effect. Figures C.2 and C.3 hint at the presence of border effects. Figure C.2 suggests distributions for country-pairs that share one border are slightly different from the Rwanda internal distribution (other within-country densities not shown), whilst Figure C.3 indicates that departures from those internal distributions might even be larger for country-pairs that are more than one border away.

5.2 The border effect: pooled regression

Considering our data availability, our preferred regression specification looks like this:

$$\begin{aligned} \left| ln\ Q_{ucc't} \right| &= \beta_0 + \beta_1\ ln\ d_{ucc'} + \beta_2\ B_{ucc'} + \beta_4\ NTB_{ucc't} \\ &+ \beta_5\ ln\ E_{ucc't} + \sum_{k=2}^{45} \gamma_k D_k + \sum_{m=2}^{55} \zeta_m D_m + \sum_{p=2}^{24} \xi_p D_p + \epsilon_{ucc't} \end{aligned} \tag{8}$$

Compared to equation (6), there is no reliable data on tariffs, and dummy variables were added for the different time periods and for the 24 goods. This level-specification is closest to the theory I want to test, which is the long-run border effect and the related effects of customs union and exchange rate pass-through. Standard errors reported below are clustered by city-pair. The underlying hypothesis for this choice is that there is a high probability that errors are not independent within the city-pairs dimension, mainly because one suspects persistence to be important in the time series dimension, as the time-span covered by the dataset is relatively short.

Table 4: Border effect pooled regressions

ALL GOODS (24 items)

VARIABLES	Specific. 1	Specific. 2	Specific. 3	Specific. 4	Specific. 5	Specific. 6	Sp.3 w/o 2008
Distance	0.0286***	0.0189***	0.0170***	0.0251***	0.0291***	0.0255***	0.0176***
	(0.00312)	(0.00345)	(0.00351)	(0.00315)	(0.00306)	(0.00309)	(0.00353)
Border dummy	0.136***			0.186***	0.148***	0.199***	
	(0.00953)			(0.00983)	(0.000966)	(0.00996)	
Dummy Incremental		0.0775***					
		(0.00480)					
Burundi-Kenya dummy			0.240***				0.221***
			(0.0211)				(0.0202)
Burundi-Rwanda dummy			0.111***				0.110***
			(0.00968)				(0.00912)
Burundi-Uganda dummy			0.167***				0.165***
			(0.0126)				(0.0122)
Kenya-Rwanda dummy			0.217***				0.198***
			(0.0156)				(0.0151)
Kenya-Uganda dummy			0.158***				0.153***
			(0.0156)				(0.0151)
Rwanda-Uganda dummy			0.168***				0.162***
			(0.00906)				0.00888
Nominal Exchange Rate				0.0202***		0.0206***	
				(0.00214)		(0)	
Doing Business					0.000168***	0.000150***	
					(0)	(0)	
Observations	1,147,465	1,147,465	1,147,465	1,147,465	897,109	897,109	1,001,424
R-squared	0.316	0.317	0.319	0.317	0.318	0.319	0.311
Dist. Eq. 1 (total)	116			1,653	162	2,450	
Dist. Eq. 2 (total)	292 km			4,159 km	407km	6,165 km	
Dist. Eq. 1 (Bur-Ken)			1,352,746				284.025
Dist. Eq. 2 (Bur-Ken)			3,663,685 km				769,235 km
Dist. Eq. 1 (Bur-Rwa)			685				518
Dist. Eq. 2 (Bur-Rwa)			1,439 km				1,088 km
Dist. Eq. 1 (Bur-Uga)			18,463				11,790
Dist. Eq. 2 (Bur-Uga)			46,293 km				29,561 km
Dist. Eq. 1 (Ken-Rwa)			349,656				76,880
Dist. Eq. 2 (Ken-Rwa)			924,558 km				203,285 km
Dist. Eq. 1 (Ken-Uga)			10,874				5,962
Dist. Eq. 2 (Ken-Uga)			26,700 km				14,640 km
Dist. Eq. 1 (Rwa-Uga)			19,582				9,942
Dist. Eq. 2 (Rwa-Uga)			46,746 km				23,735 km

Notes: Dependent variable is the absolute value of the log of good relative prices (=LHS of equation 8). All regressions are run with a constant, 23 good dummies (‘rice’ being dropped), 44 city dummies (‘Bujumbura’ is dropped), 54 month-dummies (‘Jan-04’ dropped), except for the regressions involving Doing Business indicators for which we do not have data for 2004, and 3 country-specific dummies (‘Rwanda-Rwanda’ dummy dropped). p-values in parentheses are clustered by city-pair. Significance levels are: *** p<0.01, ** p<0.05, * p<0.1

Dist. Eq. stands for Distance Equivalent. The correct expressions are [$x = exp(\beta_2/\beta_1)$] for Distance Equivalent 1 and [$x = \overline{distance} * exp(\beta_2/\beta_1 - 1)$] for Distance Equivalent 2.

To account for the 'Country Heterogeneity Effect' discussed in the previous section, I follow Gorodnichenko and Tesar (2005) who suggest using dummies for city-pairs that are both in the same country. One cannot include both a border-dummy and dummies for all four countries, as that would imply perfect collinearity. The same authors make a case for a conservative approach (i.e. which would make the border effect as small as possible) by choosing the country which has the lowest difference between within-country and between-country city-pairs' price variation. In our case this is Rwanda (calculated from table 3), hence regressions are run without a 'Rwanda-Rwanda' dummy variable. Further, we construct the 'Dummy Incremental' variable to capture the hypothesis that crossing more than 1 border should add to the border effect—kind of a double border effect. A value of one for this variable indicates no borders crossed, a two indicates one border crossing (i.e. BUR-RWA, RWA-UGA and UGA-KEN city-pairs), a three two borders (BUR-UGA and RWA-KEN city-pairs) and a four three borders (BUR-KEN city-pairs)[27].

Table 4 reports results from six regression specifications: (i) with distance and the border dummy as regressors, (ii) with the incremental dummy instead of the border dummy, (iii) with six bilateral dummies to capture bilateral border effects[28], (iv) specification one with the nominal exchange rate added, (v) specification one with doing business indicator added, (vi) specification one with both exchange rate and doing business indicators. The number of observations for the first four specifications are 1,147,465 covering 20,863 citypair-good combinations, each for 55 months[29].

The lower part of the table presents distance equivalents as suggested by Engel and Rogers (1996) for distance equivalent 1 and Parsley and Wei (2001) for distance equiv-

[27]This metric is predicated on the fact that imports into Burundi from the rest of the world mainly come through the so-called northern corridor, i.e. via Kenya and not via Tanzania. Of course, for certain staples and fruits and vegetables that are locally grown, one cannot exclude Tanzania-Burundi trade, but the point is that pass-through from Kenya to Burundi will rather happen via Uganda and Rwanda than via Tanzania (Central Corridor)—see also map A3 in annex A. It is also possible that price signals are well transmitted from Kenya to Burundi, without actual trade taking place.

[28]Note that one cannot include the nominal exchange rate in such a regression as this would be perfectly collinear with the bilateral dummies.

[29]Not all cities have data for all goods. Table 1 gives an overview of which goods are missing for which cities. If there were no missing data our dataset would span 24 (goods) * 55 (months) * (45*44/2) (number of city-pairs) = 1,306,800 observations.

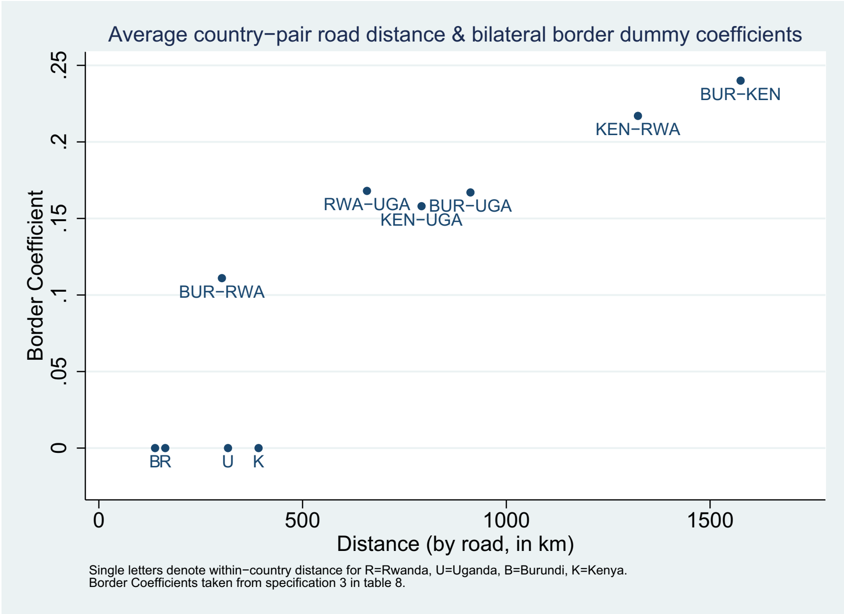
alent 2. The footnote to the table shows how to calculate these indicators. The former calculates the equivalent of the added distance induced by the border, but is unitless (i.e. one reaches the same number whether distance is measured in miles or in km). Hence the second measure, which measures how much extra distance one needs to add to the average distance between two countries to generate as much price dispersion as one actually observes.

Specification one shows both distance and the border dummy to be highly significant. A distance of 100 km between two cities allows for departures of around 13% from the LOP benchmark[30]. This is much higher than the distance coefficients reported by Broda and Weinstein (2008) or Engel et al. (2005) for the US-Canada border, showing, unsurprisingly, that distance plays a much bigger role in Eastern Africa. The border dummy is positive and highly significant, moving prices more than 13% away from the LOP hypothesis. This is higher than the 7% estimated by Engel et al. (2005) for the US-Canada border, and higher than the 2-3% found by Aker et al. (2010) for the Niger-Nigeria border. Alternative measures for the border dummy (specifications two and three) similarly show that countries that are further apart have higher deviations from the LOP. Specification three for example shows prices being moved away from the LOP equilibrium by 24% for Burundi-Kenya city-pairs, but only by 11% for Burundi-Rwanda pairs. Figure 1 shows how these coefficients are related to distance. The further away two countries are, the higher is their border dummy coefficient. The four observations near the origin depict average distances within our four countries, for which the border coefficients are zero. Note that distance is measured in km, whilst the border effect relates to the log of real exchange rates. Hence, the linear relationship depicted understates the extent to which for example Kenya and Burundian city-pairs are disconnected.

Distance equivalents are quite low (with the exception of some country-pairs in specification three) compared to similar estimates in the ER tradition, but are similar to the findings of the previously discussed papers using level-specifications. One interpreta-

[30]Calculated as $0.0286 * ln(100)$.

Figure 1: Distance and bilateral border coefficients



tion of this is that good markets are more integrated in the long-run, but that short-run deviations, which is what the ER-standard deviation approach picks up, are significant.

Can one 'explain' the border dummy and distance equivalents by adding the nominal exchange rate or measures for non-tariff barriers? The nominal exchange rate is highly significant and thus explains departures from the LOP, but it does not reduce the border effect coefficient, nor does it reduce the distance equivalents. In both specifications including the exchange rate (four and six), the coefficient is very close to zero, implying a high pass-through from foreign to domestic prices. In other words, I do not find much evidence of local currency pricing in our Eastern Africa price data. These findings corroborate the conventional wisdom that pass-through is relatively rapid and complete in countries that are small and/or less developed (Frankel et al., 2005, p1) and is counter-evidence for the claim by the same authors that the pass-through recently has decreased in developing countries. The NTB variable is also highly significant, but again reduces neither the size nor the significance of the border dummy. The inclusion of these variables also does not increase the performance of the regression in terms of R^2.

How robust are these results? First, returning to the country heterogeneity effect, none of the conclusions change when using a different combination of benchmark dummies, with the exception of using ‘Burundi-Burundi’ as the benchmark. In that case, the border coefficient actually decreases between specification one and four and between one and six. The size of this change is small, such that this does not warrant a revision of the conclusions reached so far, especially as Burundi is arguably the most ‘peripheral’ of the four countries[31]. Second, results changed neither quantitatively nor qualitatively, when using either of the three other distance measures discussed in section (4.2). Third, running regressions with a time trend instead of the 54 time dummies, does not have noticeable effects. In specification 1, the time trend was found to be positive and significant, but very small (0.00058). Fourth, a robustness check regarding the goods included in the regression was performed, where I ran 24 regressions with one good deleted every time—again, this did not change any of the above conclusions. Finally, as the final two specifications do not have data for 2004, the four other regressions were run without 2004. Results are barely different, with both the distance coefficient and the border dummy slightly larger compared to the baseline case.

5.3 A customs union effect? Border effects & the time dimension

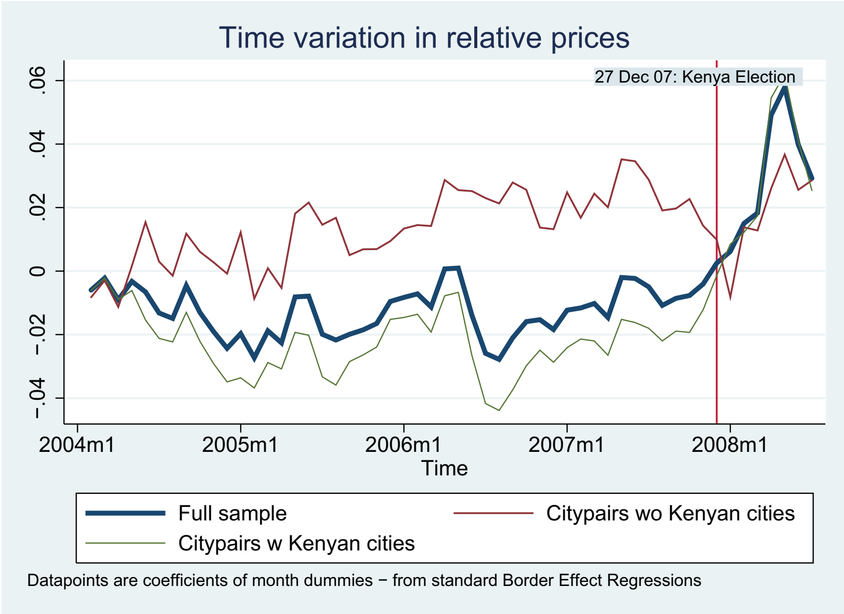
To make a start on gauging the effect of the 2008 Kenya election crisis, the last column of table 4 shows results for regression specification three, leaving out 2008 data. Even though all bilateral border coefficients are lower when leaving out 2008, the size of this decrease does not entice us to change much of our conclusions, with countries that are further apart still reporting the highest border effects. This is as expected, i.e. removing the year of the Kenya crisis (i.e. 2008) reduces the border effect[32].

The time dimension of the regression analysis is depicted in figure 2, which shows the time dummy coefficients from specification six. These coefficients can be interpreted as month-specific shocks that influence departures from the LOP but are not explained by

[31] Burundi only shares a border with Rwanda and was for much of the time of the dataset still embroiled in civil war and its aftermath.

[32] As for the other 5 specifications (not reported), the border effect coefficient again goes down in all specifications compared to the baseline, on average by around 6%.

Figure 2: Relative prices and the time dimension



any of the other explanatory variables. The thick line shows the time dummy coefficients for a regression with all city-pairs in the data set included. It shows a clear increase in the size of LOP-deviations in the last year or so of the data set (roughly August-07 to April-08), with a sharp reduction in the summer of 2008. This shows the impact of the Kenyan elections, with the series only truly moving into 'outlier' territory from January 2008 onwards. As Kenya is a transit country for the transport of many goods into Uganda, Rwanda and Burundi, markets in those 3 countries were severely disrupted. This story holds up well when adding a graph that has time dummy coefficients for a regression which includes city-pairs of those three countries only. This graph only starts going up at the very time of the Kenyan elections. The sharp decrease at the end of our dataset coincides with a return to normalcy in Kenya[33].

With the help of table 5, the possible effect of the EAC Customs Union is analysed in two ways. First, regressions in the spirit of the previous six specifications are run for 2004 and then for 2007. Looking at effects by year removes possible seasonal effects.

[33] An alternative explanation is the world food crisis. For this to be an explanation of the unfolding pattern, this phenomenon would need to have a differential impact on markets in different cities, which seems plausible. According to the FAO's Cereals Price Index (see www.fao.org/worldfoodsituation/FoodPricesIndex/en/), food prices starting increasing in the last quarter of 2006, accelerated in the second half of 2007, and reached a peak around May 2008, which sits well with our graphs.

Table 5: Looking for customs union impact

VARIABLES	TOTAL SAMPLE 2004	2007	2004-2007	KENYA-UGANDA citypairs only 2004	2007	2004-2007
Distance	0.0269***	0.0256***	0.0327***	0.0220***	0.0195***	0.0219***
	(0.00348)	(0.00328)	(0.00306)	(0.00329)	(0.00312)	(0.00290)
Border Dummy	0.0937***	0.136***		0.147***	0.120***	
	(0.0100)	(0.0103)		(0.0173)	(0.0176)	
Border-trend			0.00157***			-0.000242**
			(0.000122)			(0.000119)
Observations	250,356	250,356	1,001,424	97,812	97,812	391,248
R-squared	0.323	0.320	0.306	0.351	0.355	0.341
Dist. Eq. 1	33	203		798	709	
Dist. Eq. 2	82 km	510 km		1,959 km	4,733 km	

Notes: The notes to table 4 are still valid, with the exception that good and city-pair dummies are only used when the relevant cities/goods actually enter the regressions.

Next, I construct an interaction term between the border dummy and a time trend (called 'border-trend') and check whether this has a positive sign over the period 2004-2007. Because of the '2008' Kenya crisis effect, the seven months for which there is data in 2008 are left out.

Results seem to contradict what one would expect from the EAC Customs Union. Comparing coefficients for 2004 with 2007, the border effect coefficient goes up, with distance equivalents increasing. Similarly, the border-trend variable, estimated over the sample period 2004-2007 is positive and significant, suggesting the border effect has gotten stronger. Of course, the Customs Union agreement of 2005 was only between Uganda, Kenya and Tanzania. Hence, the final four columns of table 5 show estimates for the Kenya-Uganda country-pair only. Here, there is a CU effect, with both the distance coefficient and the border coefficient diminishing between 2004 and 2007. One can expect the former also to go down with the Customs Union, as such agreements often give new impetus to cross-country infrastructure works for example. As both distance and border dummy coefficients are going down, the CU-effect does not show up in the distance equivalents. Furthermore, the border-trend variable is significant and negative, implying a reduced effect of the border over the 2004-2007 period, which is what one would expect from deeper integration.

5.4 Border effect regressions across the four good categories

Table 6 shows results from running regression specifications one, four and six from table 4 for the four good categories. Strikingly, the border dummy is significant in all 12 regressions reported. In terms of model performance, the 'other goods' category regressions have the highest R^2, followed by the 'staples' category, which still has more than double the R^2 of the 'fruits & vegetables' category.

Looking at specification one, for food items (first three categories), the border dummy increases departures from the LOP benchmark by some 10%. This goes up to 25% for the 'other good' category. Distance drives less of a wedge between prices in different locations for 'staples' compared to the other categories, which is an indication that staple goods are more easily traded across the region. One reason is that staples are a very important part of people's diets across the region, as was shown in table 2, and that markets for these goods are 'deep'[34]. Because of their prominence in household food consumption, some of these staples have been subjected to market interventions, like food aid and/or government subsidies[35]. Distance could then play less of a role in far-off areas, if these areas get relatively higher subsidies or more food aid. Such interventions are much less common for fruits or vegetables. Another reason for the relatively higher distance coefficient for the latter compared to staples is that they are perishable. So, even though regional trade in 'fruits & vegetables' is well-documented[36], according to this estimation, these goods have less integrated markets compared to staple foods both across (border dummy coefficient) and within (distance coefficient) countries.

[34] von Cramon-Taubadel et al. (2009), exploiting more the time series dimension of the data, find maize markets to be well integrated between Ugandan and Kenyan cities, but less so with Tanzania. Distance has a significant effect on the strength of the price transmission within the EAC region. Another study by the University of Michigan similarly find the maize markets in Eastern and Southern African well integrated, even though Govereh et al. (2008) and Haggblade et al. (2008) claim that further regional integration and cooperation would help improve market functioning.

[35] Even though the countries studied here proclaim a policy of non-intervention and market-driven price determination, this is not always implemented uniformly. See for example the Kenya and Uganda country chapters in Anderson and Masters (2009). For the maize market in Kenya, Govereh et al. (2008) find that *Kenyan maize policy in recent decades has involved supporting high prices for farmers through high government procurement prices.*

[36] See for example Bank of Uganda and UBOS (2007) for informal trade related to Uganda's borders.

Table 6: Border effect regressions for good categories

	STAPLE FOOD (5 items)			**FRUITS & VEGETABLES (7 items)**		
VARIABLES	Specific. 1	Specific. 4	Specific. 6	Specific. 1	Specific. 4	Specific. 6
Distance	0.0148***	0.0154***	0.0163***	0.0360***	0.0309***	0.0306***
	(0.00456)	(0.00465)	(0.00457)	(0.00601)	(0.00617)	(0.00608)
Border Dummy	0.101***	0.0923***	0.0879***	0.0804***	0.153***	0.208***
	(0.0138)	(0.0151)	(0.0211)	(0.0182)	(0.0204)	(0.0262)
Nominal Exchange Rate		-0.00355	-0.00443		0.0307***	0.0316***
		(0.00276)	(0.00279)		(0.00451)	(0.00452)
Doing Business			0.000122			-0.00103***
			(0.000272)			(0.000242)
Observations	238,535	238,535	186,491	358,160	358,160	280,016
R-squared	0.442	0.442	0.445	0.194	0.197	0.199
Dist. Eq. 1	920	401	220	9	141	896
Dist. Eq. 2	2,316 km	1,009 km	553 km	23 km	356 km	2,254 km

	OTHER FOOD (6 items)			**OTHER GOODS (6 items)**		
VARIABLES	Specific. 1	Specific. 4	Specific. 6	Specific. 1	Specific. 4	Specific. 6
Distance	0.0242***	0.0197***	0.0195***	0.0376***	0.0327***	0.0343***
	(0.00268)	(0.00259)	(0.00261)	(0.00360)	(0.00344)	(0.00357)
Border Dummy	0.109***	0.171***	0.0655***	0.252***	0.317***	0.310***
	(0.0110)	(0.0132)	(0.0186)	(0.0159)	(0.0154)	(0.0192)
Nominal Exchange Rate		0.0261***	0.0262***		0.0254***	0.0275***
		(0.00296)	(0.00303)		(0.00271)	(0.00287)
Doing Business			0.00203***			0.00112***
			(0.000179)			(0.000210)
Observations	288,530	288,530	225,578	262,240	262,240	205,024
R-squared	0.342	0.346	0.340	0.448	0.451	0.460
Dist. Eq. 1	90	5,885	29	814	16,223	8,416
Dist. Eq. 2	228 km	14,813 km	72 km	2,049 km	40,833 km	21,183 km

Notes: The notes to table 4 are still valid, with the exception that good dummies are only used when the relevant goods actually enter the above regressions. The dependent variable is still the absolute value of the log of good relative prices (=LHS of equation 8).

Neither the nominal exchange rate, nor the doing business indicator seems to have much of an impact on the border effect—similar to the pooled regression. Interestingly, neither of these two indicators is even significant in the staples regression. This points to very high pass-through of the nominal exchange rate, and NTBs not playing a large role when it comes to trading these goods across borders. This is in line with our discussion of NTBs in section 4.3, where the NTBs seemed to be concentrated in non-staple food sectors. The exchange rate result sits rather well with the Engel (1993) finding for industrialised countries that for most goods there are significant deviations from the LOP at the consumer level, with the exception of 'pure' commodities (as opposed to more complex consumer products)[37].

All this points to well-integrated and well-functioning regional markets for staples. I corroborate this by running regression specifications one and four for the five individual goods making up the 'staples' category (results not reported here). I find that the border dummy is not significantly different from zero for sweet potatoes and maize flour and negatively significant for maize grains. These results confirm the evidence gathered in the recent crop-based work on the regional maize market in Eastern and Southern Africa[38].

The category 'other goods' has the highest border effects, which fits well with our prior that these goods are less traded across borders. However, running regressions for the goods in this category separately, one sees that this effect is mainly due to 'domestique' (household services), and 'charcoal', which has the highest border coefficient of all goods. For the latter, this is probably related to repeated production and export bans and other policy interventions in EAC countries to combat deforestation and other charcoal-related environmental damage[39]. The results for the other goods in this category, like cigarettes, petrol, beer and soda are similar to those for the food items. This is not un-

[37] It reasonable to assume that staples are more like a 'pure' commodity than the goods in the other categories. Engel and Rogers (2001) find evidence for this hypothesis in a dataset spanning 11 European countries.

[38] See the references mentioned in footnote 43 in this chapter.

[39] For Kenya, see e.g. ban on moving charcoal from Turkana (http://allafrica.com/stories/200605110704.html), whilst in Rwanda charcoal trade is linked to rebel activity in the DRC which the Government is cracking down on (http://allafrica.com/stories/200708201143.html).

surprising in the light of more anecdotal evidence on trade flows for these goods in the region (see section (3) for a reference). The high border dummy coefficient for the ‘other goods’ category translates into high distance equivalents. For the other three categories, distance equivalents are much lower however and in line with the pooled regression results. Note that the very low results for this indicator for ‘fruits & vegetables’ is more due to the high distance coefficient, rather than a low border coefficient.

6 Conclusion

This paper investigated the behaviour of consumer prices across Eastern Africa between 2003 and 2008, a period in which, through the start of the Customs Union, market integration was supposed to be accelerating. Using level regressions to test the law-of-one price hypothesis, I find that the border effect as measured by the border dummy coefficient, is significant, with distance equivalents ranging from 300 to 6,000 km. This is much lower compared to papers using the Engel-Rogers regression specifications, but in line with the (few) papers using real exchange rate level regressions. I also find that cities that are in countries that do not share a border show higher departures from LOP compared to cities in adjacent countries. Distance plays a big role, both between and within countries, with distance coefficients much higher compared to evidence from the Canada-US border. Neither the nominal exchange rate, nor Non-Tariff Barriers reduce the border effect very much, even though they are both significant variables in the regression analysis. Looking at specific goods, I corroborate earlier findings that markets for staple foods are well integrated. Non-Trade barriers are most important for fruits & vegetables, which can be linked to these goods being perishable. I finally research whether the start of the customs union between Tanzania, Uganda and Kenya has shown a reduction in the deviations to the LOP. I find that there is a positive effect for the Kenya-Uganda border, but that the benefits (in terms of lower departures from the LOP) do not extend to Rwanda or Burundi. Further, the Kenya crisis of early 2008 can clearly be linked to higher departures from the LOP throughout the region.

This chapter is part of a larger research agenda regarding the behaviour of consumer prices in Eastern Africa. Little is known about the real and nominal aspects of good-city-time series like the ones researched in this paper for developing countries. Our finding that a border dummy plays a significant role in explaining real exchange rates, shows more work needs to be done in this field. Particularly interesting for future work seems to be to understand what drives the time variation in the price data (see figure 2 and the discussion in section 5.3), and especially the 2008 developments, involving the international food price shock and the Kenyan crisis. It would also be interesting to take a closer look at what factors drive the divergence and convergence between relative price series and nominal exchange rates. I have hinted that the nominal exchange rate pass-through is high in the region, which lends credibility to the characterisation of these developing countries as small-open economies. The fact that the pass-through was found to be lower for city-pairs that include a Kenyan city strengthens this hypothesis, as Kenya is the largest economy in the region. These issues will come increasingly to the fore with the prospect of Monetary Union. In that respect it will be crucial for EAC member states to improve on the comparability of inflation statistics, which will allow more systematic research on pricing in the region.

References

Jenny Aker, Michael Klein, Stephen O'Connell, and Muzhe Yang. Borders, Ethnicity and Trade. *NBER Working Paper 15960*, 2010.

James Anderson and Eric Van Wincoop. Borders, Trade and Welfare. in Susan Collins and Dani Rodrik (eds.), Brookings Trade Forum 2001, p.207-44, 2002.

James Anderson and Eric Van Wincoop. Trade Costs. *Journal of Economic Literature*, XLII: 691–751, September 2004.

Kym Anderson and William Masters. *Distortions to Agricultural Incentives in Africa*. The World Bank, 2009.

Bank of Uganda and UBOS. The Informal Cross Border Trade Survey Report 2006, May 2007.

Christian Broda and David E. Weinstein. Understanding International Price Differences Using Barcode Data. *NBER Working Paper 14017*, 2008.

Guglielmo Caporale and Luis Gil-Alana. Long Memory and Volatility Dynamics in the US Dollar Exchange Rate. *DIW Berlin German Institute for Economic Research Discussion Paper 975*, February 2010.

Stephen Cecchetti, Nelson Mark, and Robert Sonora. Price Index Convergence Among United States Cities. *International Economic Review*, 43(4):1081–1099, November 2002.

Mario Crucini and Mototsugu Shintani. Persistence in Law of One Price Deviations: Evidence from Micro-data. *Journal of Monetary Economics*, 55:629–644, 2008.

Michael Devereux and Charles Engel. Fixed vs. Floating Exchange Rates: How Price Setting Affects the Optimal Choice of Exchange Rate Regime. *NBER Working Paper 6867*, 1998.

EAC. Protocol on the Establishment of East African Community Customs Union. East African Community Secretariat, Arusha, Tanzania, March 2, 2004.

Graham Elliott, Thomas Rothenberg, and James Stock. Efficient tests for an autoregressive unit root. *Econometrica*, 64(4):813–836, 1996.

Charles Engel. Real Exchange Rates and Relative Prices: An Empirical Investigation. *Journal of Monetary Economics*, 32:35–50, 1993.

Charles Engel and John Rogers. How Wide is the Border? *American Economic Review*, 86(5): 1112–1125, 1996.

Charles Engel and John Rogers. Deviations from Purchasing Power Parity: Causes and Welfare Costs. *Journal of International Economics*, 55:29–57, 2001.

Charles Engel, John Rogers, and Shing-Yi Wang. Revisiting the Border: An Assessment of the Law of One Price Using Very Disaggregated Consumer Price Data. Chapter 8 in Exchange Rates, Capital Flows and Policy, Edited by Rebecca Driver, Peter Sinclair and Christoph Thoenissen, 2005.

Paul L. Fackler and Barry K. Goodwin. Spatial Price Analysis. Chapter 17 in Handbook of Agricultural Economics, Volume 1, Edited by B.Gardner and G.Rausser, 2001.

Robert Feenstra and Jon Kendall. Pass-through of Exchange Rates and Purchasing Power Parity. *Journal of International Economics*, 43:237–261, 1997.

Jeffrey Frankel, David Parsley, and Shang-Jin Wei. Slow Passthrough around the World: A new Import for Developing Countries? *NBER Working Paper 11199*, 2005.

Freytag and Berndt. East Africa Road Map (Kenya, Tanzania, Uganda). German Edition.

Luis Gil-Alana. Stochastic Behaviour of Nominal Exchange Rates. *Atlantic Economic Journal*, 31(2):159–173, June 2003.

Yuriy Gorodnichenko and Linda Tesar. A Re-Examination of the Border Effect. *NBER Working Paper 11706*, 2005.

Yuriy Gorodnichenko and Linda Tesar. Border Effect or Country Effect? Seattle May Not Be so Far from Vancouver After All. *American Economic Journal: Macroeconomics*, 1(1):219–241, 2009.

Jones Govereh, Steven Haggblade, Hunter Nielson, and David Tschirley. Report 1: Maize Market Sheds in Eastern and Southern Africa. Report prepared by Michigan State University for the World Bank under contract No. 7144132, Strengthening Food Security in sub-Saharan Africa through Trade Liberalization and Regional Integration, June 2008.

Clemens Grafe, Martin Raiser, and Toshiaki Sakatsume. Beyond Borders - Reconsidering Regional Trade in Central Asia. *Journal of Comparative Economics*, 36:453–466, 2008.

Kaddour Hadri. Testing for stationarity in heterogeneous panel data. *The Econometrics Journal*, 3:148–161, 2000.

Steven Haggblade, Hunter Nielson, Jones Govereh, and Paul Dorosh. Report 2: Potential Consequences of Intra-Regional Trade in Short-Term Food Security Crises in Southeastern Africa. Report prepared by Michigan State University for the World Bank under contract No. 7144132, Strengthening Food Security in sub-Saharan Africa through Trade Liberalization and Regional Integration, June 2008.

James Honaker, Gary King, and Matthew Blackwell. Amelia II: A Program for Missing Data. Accessed via http://gking.harvard.edu/amelia/docs/amelia.pdf.

ILO. *Consumer Price Index Manual - Theory and Practice*. ILO Publications, August 2004.

ISTEEBU. Bulletin Mensuel des Prix — Fevrier 2007. Technical report, Institut de Statistiques et d'Etudes Economiques du Burundi, March 2007. accessed via www.burundistats.org/downloads/BulletinPrix2007.pdf on 22 April 2009.

Kenya National Bureau of Statistics. The New Kenya Consumer Price Index - User's Guide. Technical report, Ministry of Finance and Planning, January 2002.

Gary King, James Honaker, Anne Joseph, and Kenneth Scheve. Analyzing Incomplete Political Science Data: An Alternative Algorithm for Multiple Imputation. *American Political Science Review*, 95(1):49–69, March 2001.

GS Maddala and Shaowen Wu. A Comparative Study of Unit Root Tests With Panel Data and A New Simple Test. *Oxford Bulletin of Economics and Statistics*, 61:631–652, 1999.

John McCallum. National Borders Matter. *American Economic Review*, 85:615–23, 1995.

Mahbub Morshed. What Can We Learn From a Large Border Effect in Developing Countries? *Journal of Development Economics*, 72:353–369, 2003.

Mahbub Morshed. Is There Really A 'Border' Effect? *Journal of International Money and Finance*, 26(7):1229–38, 2007.

Evarist Mugisa, Chris Onyango, and Patrick Mugoya. An Evaluation of the Implementation and Impact of the EAC Customs Union. Report commissioned by the Directorates of Customs and Trade of the EAC Secretariat, March 2008.

National Institute of Statistics Rwanda. Consumer Price Index - Methodology, 2007.

Maurice Obstfeld and Kenneth Rogoff. The Six Major Puzzles in International Macroeconomics: Is there a Common Cause? in NBER Macroeconomics Annual 2000, by B.S. Bernanke and K.Rogoff (eds.), pp 339-90, 2000.

M.O. Odedokun. Fulfilment of Purchasing Power Parities in Africa: The Differential Role of CFA and non-CFA membership. *Journal of African Economies*, 9(2):213–234, 2000.

David Parsley and Shang-Jin Wei. Explaining the Border Effect: The Role of Exchange Rate Variability, Shipping Costs and Geography. *Journal of International Economics*, 2001.

Hashem Pesaran. A Simple Panel Unit Root Test in the Presence of Cross Section Dependence. *Journal of Applied Econometrics*, 22(2):265–312, 2007.

Alberto Portugal-Perez and John Wilson. Trade Costs in Africa: Barriers and Opportunities for Reform. *World Bank Policy Research Working Paper No.4719*, September 2008.

Shahidur Rashid and Nicholas Minot. Are Staple Food Markets in Africa Efficient? Spatial Price Analyses and Beyond. Paper presented at the COMESA policy seminar "Food price variability: Causes, consequences, and Policy Options", 25-26 January 2010, Maputo, Mozambique., January 2010.

Kenneth Rogoff. The Purchasing Power Parity Puzzle. *Journal of Economic Literature*, XXXIV (II):647–68, 1996.

Kalyan Sanyal and Ronald Jones. The Theory of Trade in Middle Products. *American Economic Review*, 72(1):16–31, 1982.

Ugandan Bureau of Statistics. Uganda CPI - Major topics, 2006.

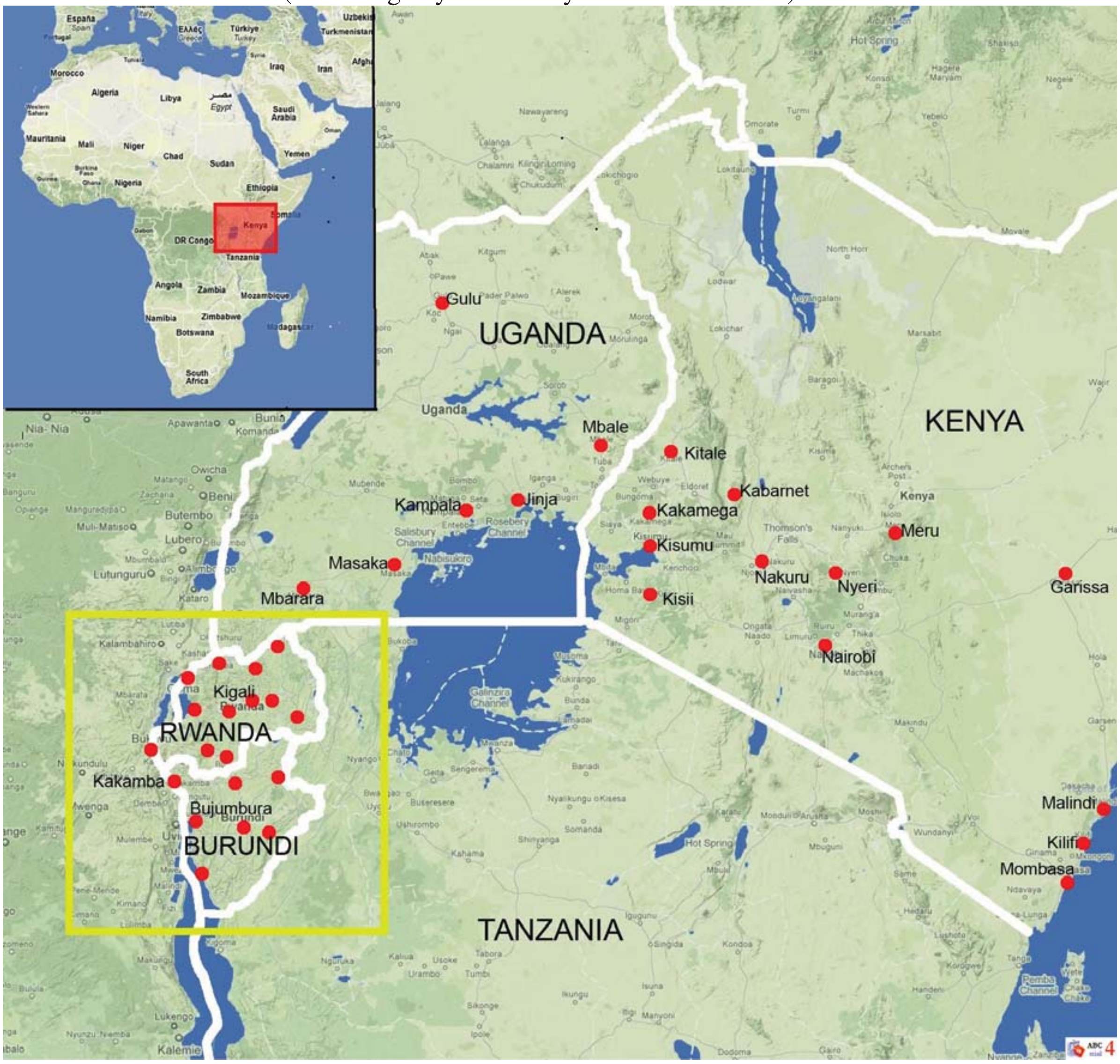
Stephan von Cramon-Taubadel, Rico Ihle, and Joseph Amikuzuno. Price Transmission in the Eastern Africa Maize Market. Report prepared for the World Bank, January 2009.

World Bank. Non-Tariff Measures on Goods Trade in the East African Community - Synthesis Report. Report No. 45708-AFR, World Bank, Washington, DC, 2008.

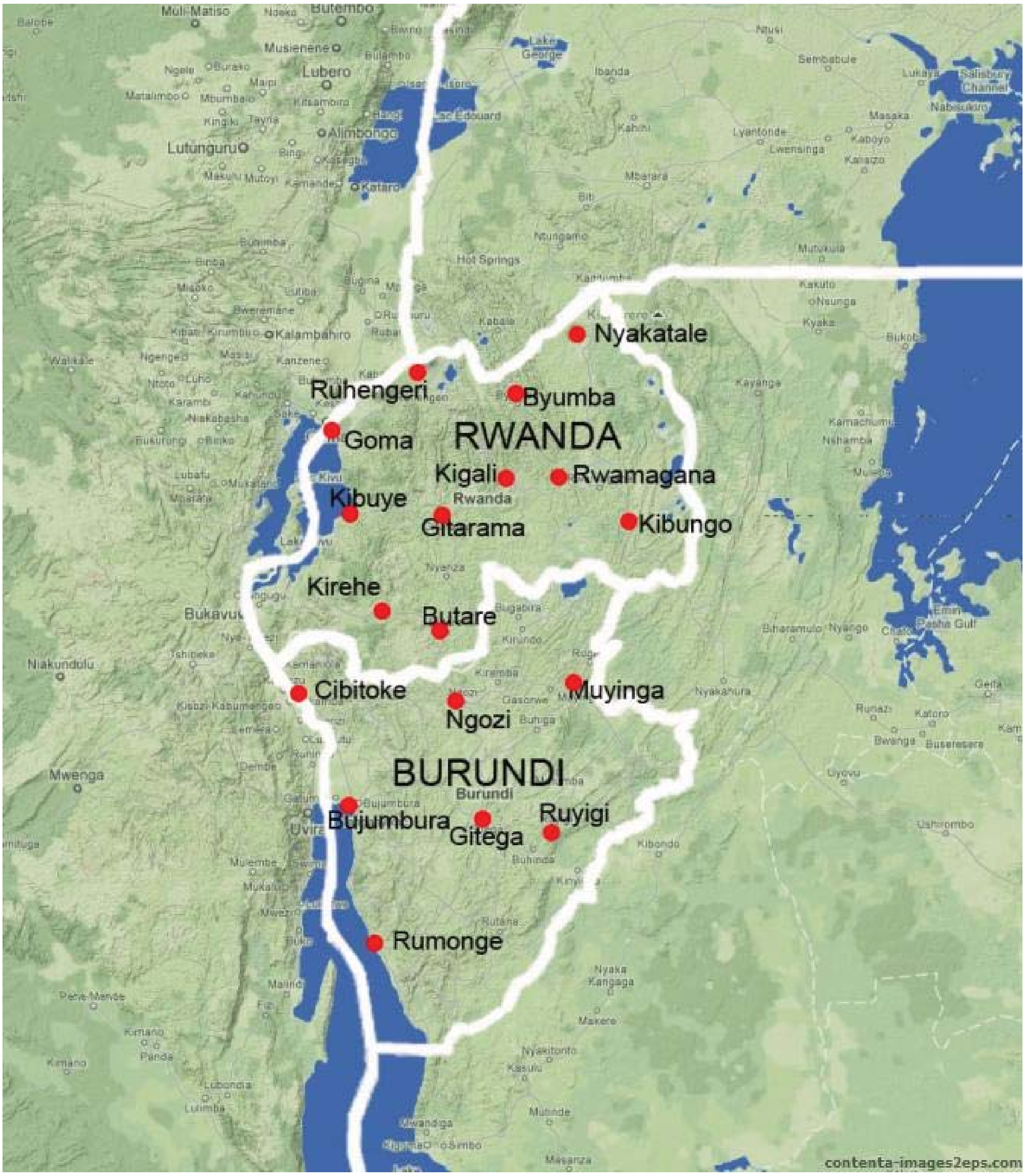
World Bank. East Africa's Infrastructure: A Regional Perspective. powerpoint presentation for the Africa Infrastructure Country Diagnostic accessed at www.infrastructureafrica.org/aicd/library/doc/813/eac-rec-ppt, 2010.

A Maps

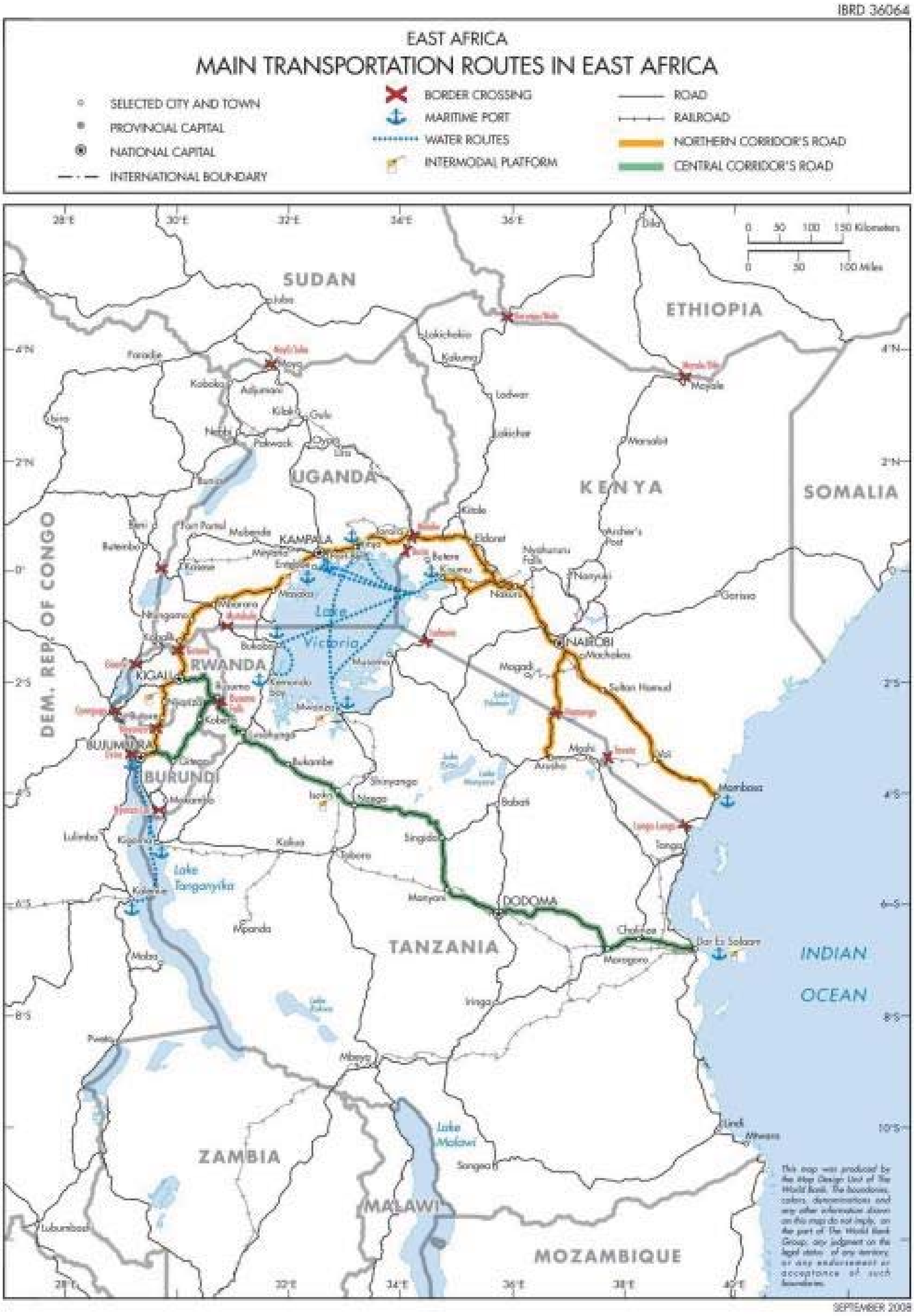
MAP A1: Eastern Africa (indicating city availability in the CPI datasets)



MAP A2: Rwanda and Burundi (indicating city availability in the CPI datasets)



MAP A3: Northern and Central Corridors



B The price data set

B.1 A detailed description of the CPI dataset

Table 7 gives an overview of the coverage of the CPI data. The number of goods for which data is available ranges from 216 in Kenya to 557 in Burundi. These numbers are slightly deceptive however, because (i) not all items are available in all cities, with more remote cities reporting sometimes less than half of the goods for which price data is available in the respective capital city, (ii) some datasets (notably the Burundi one) contain missing values[40], and (iii) for Kenya, there are actual price quotes available, below the so-called 'elementary aggregates' level (see box 1 for an explanation of elementary aggregates). Hence, Kenyan elementary aggregates are disaggregated into more finely defined items, which means that at least for Kenya, I have slightly more goods to work with than the 216 quoted. The following sections present the country-specific datasets in more detail.

BURUNDI. The *Institut de Statistiques et d'Etudes Economiques du Burundi* (ISTEEBU) is responsible for the collection of Consumer Price Data in Burundi. Inflation is reported on a monthly basis, but takes only into account consumer prices gathered in the capital city of Bujumbura. Next to this Bujumbura-index, ISTEEBU also collects prices in six other cities (ISTEEBU, 2007). This is done using a slightly different methodology and for comparison purposes, prices for Bujumbura are also collected using this (latter) methodology. The Burundi data suffers from quite a lot of missing observations. The next section in this appendix outlines a strategy for dealing with this. Because the price data collected for the Bujumbura index has better coverage, those prices are used for Bujumbura. Even though the collection method is slightly different, correlations between the two Bujumbura series is high. Amongst the 24 common goods retained, there are two goods, 'milk' and 'maize grains', that do not appear in the Bujumbura Index. For these items, prices from the other (non-CPI) Bujumbura series were used.

[40] Note that in the CPI overview table (table 7) 'Missing Observations' refers to incomplete series, i.e. series per good-city that have less than 55 observations, but more than zero observations.

Table 7: CPI Country Overview†

	Burundi	**Kenya**	**Rwanda**	**Uganda**
Cities	7 (Bujumbura, Cibitoke, Gitega, Muyinga, Ngozi, Rumonge, Ruyigi)	13 (Nairobi, Mombasa, Kilifi, Malindi, Meru, Nyeri, Nakuru, Garissa, Kisii, Kisumu, Kakamega, Kitale, Kabarnet)	12 (Kigali, Butare, Gisenyi, Ruhengeri, Kibungo, Gikongoro, Cyangugu, Gitarama, Kibuye, Kabuga, Byumba, Umutara)	7 (Kampala, Masaka, Mbarara, Jinja, Mbale, Gulu, Arua)
Areas[a]	N/A	15 (10 in Nairobi, 3 in Mombasa, 2 in Nakuru)	N/A	N/A
Income Groups	1	3 (Nairobi has 3 types of income group)	1	2 (Kampala has a low-income and a high-income index)
Items in Consumer Basket (min-max)[b]	557	216 (98 - 200)	413 (211 - 386)	243 (208 - 223)
Items for which price quotes not collected monthly	N/A	N/A	N/A	Rent, Health & Medical goods, Education, Centrally Administered Prices
Data available 'below' elementary aggregates?	Only for Bujumbura-index (weekly prices available which are then aggregated into monthly averages)	Yes, on average 3 individual price quotes available per item in every of the 25 areas	No	No
Source of Weights	unknown	1993-1994 Urban Household Budget Survey	2001 Household Living Conditions Survey (French acronym EICV)	Uganda National Household Survey 2005/2006
Weights available per city?	No (only weighting scheme available is for Bujumbura-index)	Only for Nairobi Low Income, Nairobi Middle/Upper Income and 'Rest-of-Kenya'	yes (with different weights attached depending on whether goods are locally sourced or imported)	Yes
Observations missing?[c]	Yes	No	Yes	No
Period Available	Jan 2003 - Aug 2008	Jan 1997 - Aug 2008	Jan 2004 - Jul 2008	Jan 1997 - Aug 2008
Frequency	Monthly	Monthly	Monthly	Monthly

† For more on CPI data collection and calculations, see B. Also, for Burundi, see ISTEEBU (2007); for Kenya, see Kenya National Bureau of Statistics (2002); for Rwanda, see National Institute of Statistics Rwanda (2007) and for Uganda, see Ugandan Bureau of Statistics (2006).
[a] In some Kenyan cities, different price indices are constructed for different city areas.
[b] This is the total number of items. It might be that not all cities have all items in their respective baskets, with the most 'developed' cities (typically capital cities), having the most items — i.e. being the most diversified. In brackets are listed those cities with the minimum and maximum number of items for which price series are available in that country. These statistics are not reported for Burundi as it constructs only a CPI index for Bujumbura and data coverage is very patchy.
[c] Missing observations $\approx$ Good-city time series that have less than 55 observations, but more than zero observations.

Box 1: CPI calculations

It is instructive to reflect shortly on the way statistical offices calculate Consumer Price indices (CPIs). A good reference is the ILO CPI manual (ILO, 2004). A country's overall Consumer Price Index is the result of aggregating prices of different goods into one index. Sometimes good prices are also aggregated over different cities. The evolution over time of the country-wide index is then used as a summary statistic of the country's inflation path. In the aggregation process, weights are used to make sure goods that are consumed more often are counted more heavily in the index. These weights are typically expenditure weights calculated from national household surveys. In such surveys, however, expenditures/consumption are only recorded for categories of goods. For example, a household would report how much beer it consumes, but not the exact brand of beer. The lowest level in the aggregation chain for which expenditure weights from household surveys are available, are called 'elementary aggregates'. As the name indicates, these are already averages (aggregates) of real price quotes. Thus, CPI statisticians average up prices to the level of these 'elementary aggregates', using geometric or arithmetic means, and then use weights to construct the higher levels of the aggregation chain, and ultimately the country-wide CPI.

KENYA. The Kenyan National Bureau for Statistics (KNBS) is the institution responsible for Kenyan inflation statistics. Every month, KNBS enumerators collect prices for a basket of consumer goods in 13 cities across Kenya in order to calculate the Consumer Price Index. In the larger cities of Nairobi, Mombasa and Nakuru, there are several areas in which these prices are collected. In Nairobi for example there are 10 separate areas, covering different income groups and geographical locations. For every item in the consumer price basket, on average three price quotes are gathered every month, and aggregated up into elementary aggregates (see box 1). A total of 216 elementary aggregates are collected every month, with 67 items common to the 25 areas in the data set. Kenya is the only country in our sample for which I have actual price quotes, i.e. I can go a level below the 'elementary aggregates'. This means that for some of the 24 goods retained for the analysis, one can drill down one level to make certain goods more comparable with the goods definitions in the other three countries and within Kenya itself. This is done for the following four 'elementary aggregates':

Elementary Aggregate		Replaced by
1. Soda	⇒	Fanta
2. Beer	⇒	Tusker Beer
3. Cigarettes	⇒	Sportsman Cigarettes
4. Petrol	⇒	Petrol - Super

As is explained in more detail in the main text, the Nairobi high-income groups (which cover three areas in Nairobi) are not retained for the final analysis, leaving us with 22 areas (in 13 cities) to work with.

RWANDA. The National Institute of Statistics of Rwanda (NISR), which is responsible for the CPI in Rwanda, collects monthly price data in 12 different cities. The data for Ruhengeri and Byumba were found to be very incomplete for the goods of interest, and were dropped, i.e. no effort was made to recover some data through imputing, as is done for some of the other series. Expenditure weights are city-specific, but are collected one level above the aggregation level that is used in the other countries of our sample. This means weights are given at the level of milk instead of one level below (i.e. fresh milk, or powder milk, etc). To get to the 'elementary aggregate', a geometric mean is taken from the prices of the latter products. Implicitly, this means that these items get the same weighting within their elementary aggregate. Price quotes for a total of 413 items are collected every month, with 105 items common to the 10 retained cities. The city with the most items covered is Kigali City (386) and the lowest number is for the city of Umutara (211).

UGANDA. Similar to Rwanda, Uganda calculates its CPI indices through city-specific expenditure weights, taken from the Uganda National Household Survey. In total, there are seven cities for which data is available. However, because price quotes have only started to be collected recently for Arua, this city is not included in the analysis. There are two series for Kampala, with the Kampala High-Income series only begun a couple of years ago. Hence, the series included in the analysis is the low income one. Price quotes for a total of 243 items are collected every month, with 187 items common to our six cities. Kampala Low/Middle Income series has the most goods (223) and Gulu the least (208). Price data is only available at the 'elementary aggregates' level. For averaging up to the elementary levels, arithmetic averaging is used.

B.2 Dealing with missing observations in Rwanda and Burundi

For the 24 goods retained, the raw price data from Rwanda and Burundi are not balanced panels. The two figures below give an overview of the missing data by good and by city. For each city, reading the graphs from top to bottom gives you the months in which price data is available (white) or missing (grey). For 'milk' in 'Gitarama' (see the Rwanda graph) for example, there are some missing observations in the beginning and at the end of the 55 months of data. For Burundi, most good-city pairs have some missing data, which is mainly due to data missing for entire month-city pairs (e.g. the city of Ngozi for March 2006). In Rwanda, missingness is more related to certain goods, such as oranges, a good which has no observations in five cities.

Most social scientists, when confronted with missing data, use list-wise deletion, which effectively means throwing away potentially valuable observations. Researchers have tried to mend this problem by writing algorithms that calculate 'best guesses' for the missing data-points. I use Amelia II, a sophisticated imputation algorithm, based on the EMB algorithm, developed by King et al. (2001). *'Amelia II performs multiple imputation, a general-purpose approach to data with missing values. Multiple imputation has been shown to reduce bias and increase efficiency compared to listwise deletion'* Honaker et al.. This algorithm is for example also used by von Cramon-Taubadel et al. (2009) for imputing missing values in similar price series.

Even with the AMELIA algorithm at our disposal, a city-time-good series (i.e. 55 months for one good for one city) is only included if it has a minimum number of observations. The cut-off point for inclusion used is 15 observations (or just over 25 percent of observations). AMELIA diagnostics show that this is a reasonable approach, as confidence intervals of point estimates increase quite substantially when the number of actual observations goes to zero.

In total, Burundi has 529 missing observations (out of a total of 55 months * 24 goods * 7 cities=9240). 5 series had less than 15 observations, such that the total number of imputed values is equal to 389. For Rwanda, the equivalent figures are: 465 missing observations (out of a total of 55*24*10 = 13200). 10 series had less than 15 observations, such that the total number of imputed values is equal to 31.

FIGURE B.1: Missing observations for Burundi

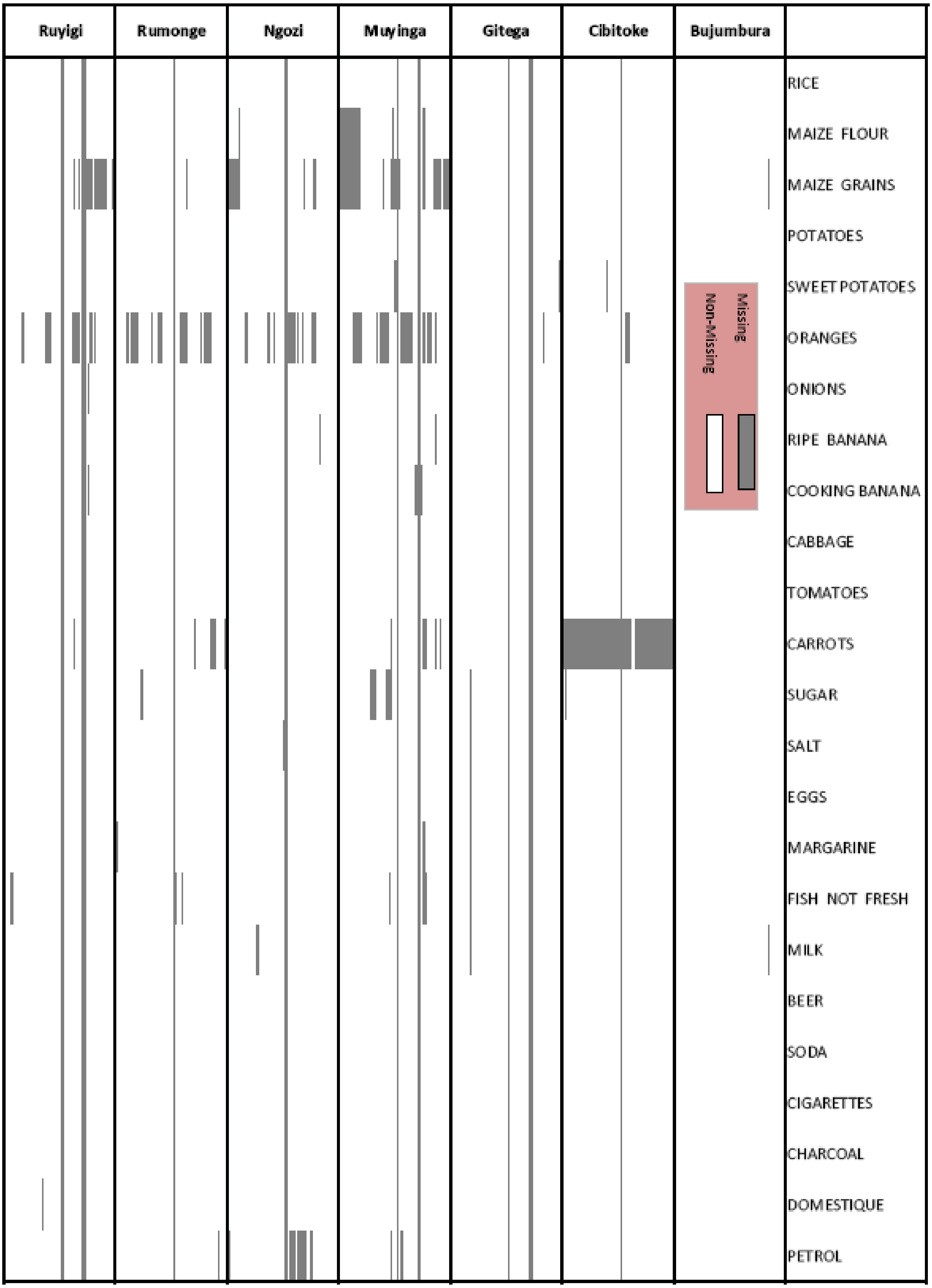
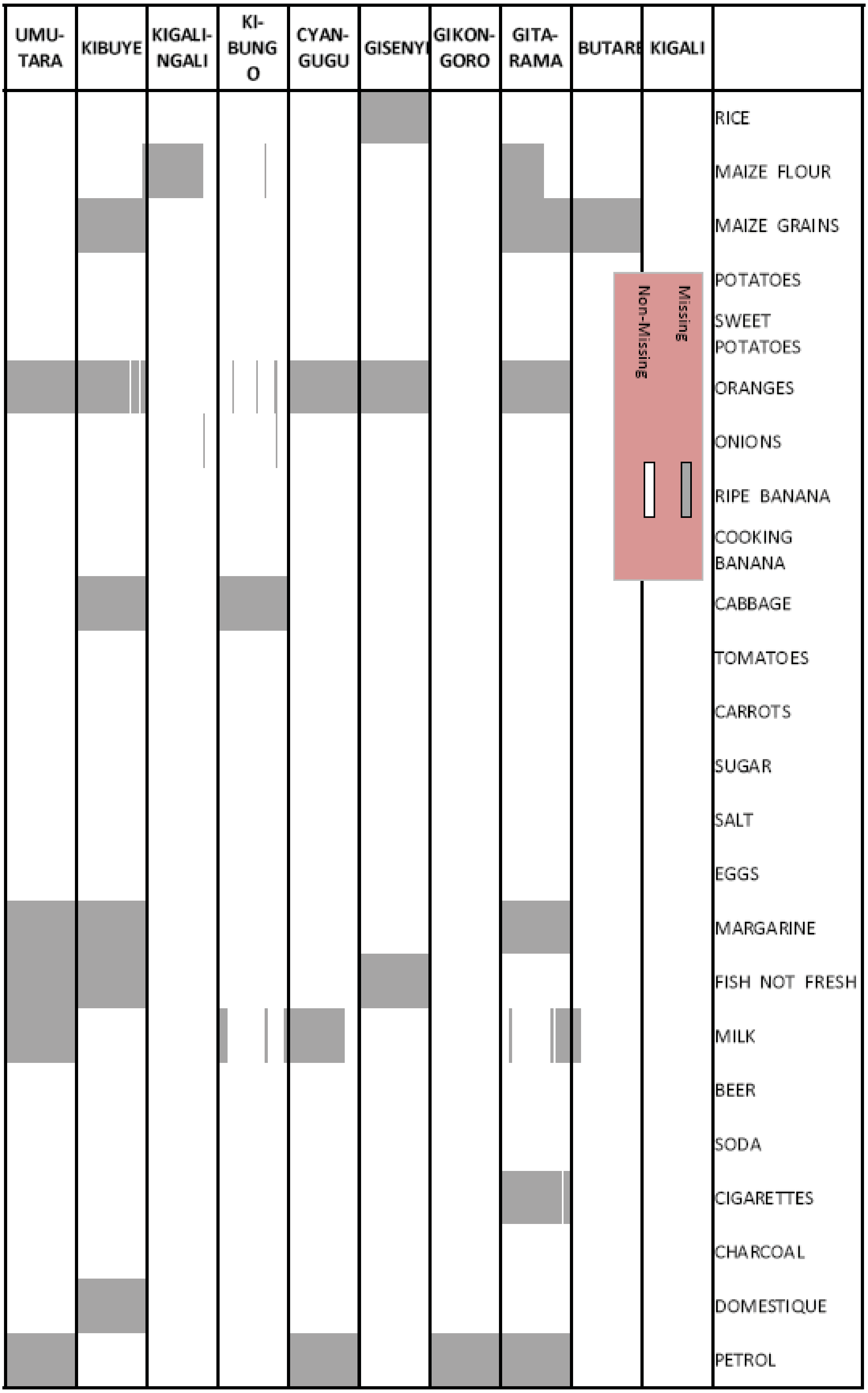


FIGURE B.2: Missing observations for Rwanda



C Kernel densities of relative prices between and within countries

FIGURE C.1

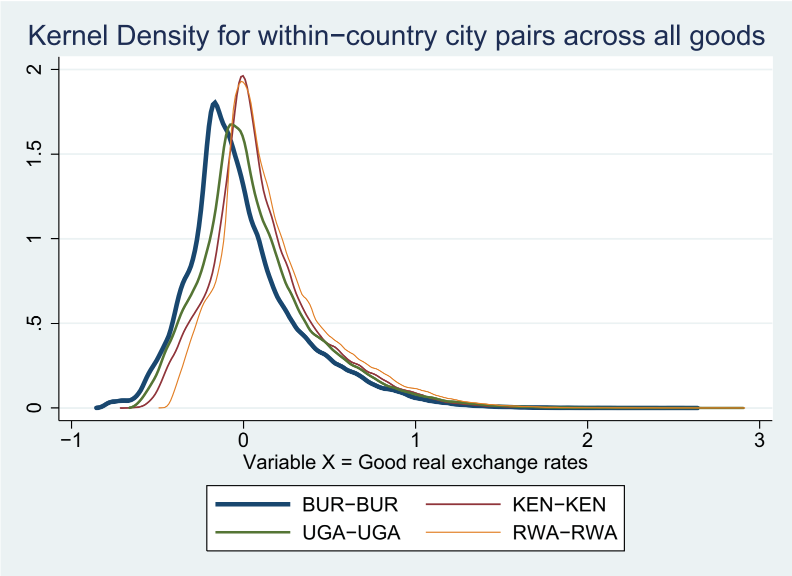


FIGURE C.2

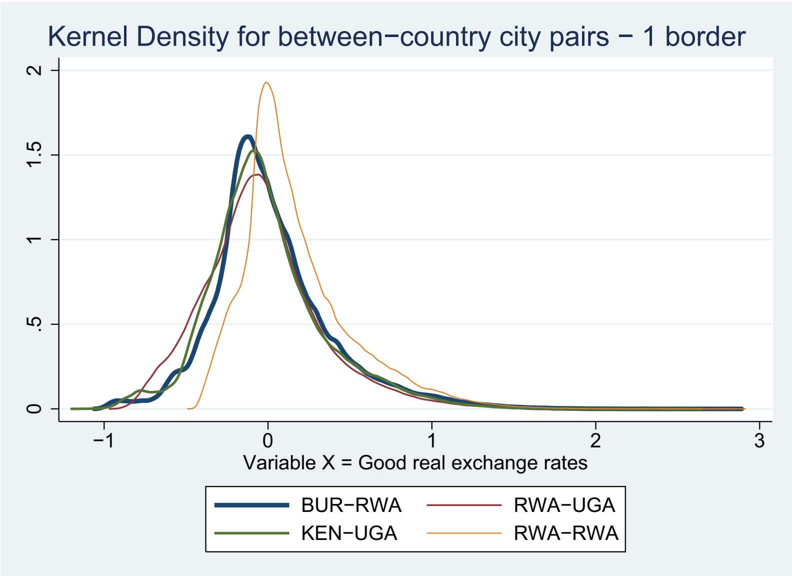
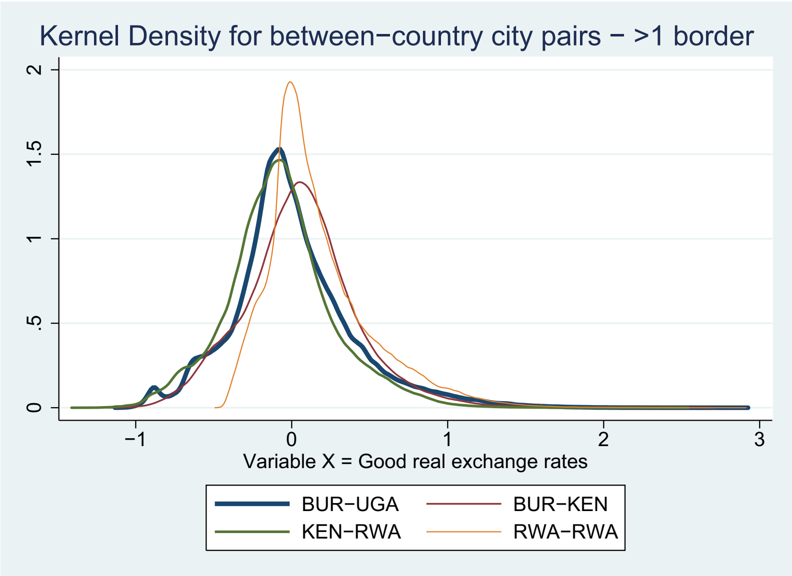


FIGURE C.3



Note: All kernel densities shown here are conditional on distance, plus dummies for cities, goods and months.