TI 2017-117/V
Tinbergen Institute Discussion Paper

Bridges

Revision: April 2019

Wouter Zant[1]

[1] VU Amsterdam

Bridges

Wouter Zant*

Abstract

We use spatial maize prices to estimate to what extent bridges in Mozambique lead to transport cost reductions and attribute these reductions to road distance and road quality. The applied methodology allows for potentially oligopolistic traders with spatially varying mark-ups. For identification we exploit the introduction of a road bridge over the Zambezi river, which creates variation in trading itineraries between markets. Estimations are based on monthly maize prices, in 24 markets, for five years before and after the introduction of the bridge. Estimates of the reduction of transport costs, averaged over itineraries, vary from 6% to 10%. Results are robust for non-random bridge placement and various other threats, and supported by observed transport cost data. Reduction in transport costs for particular itineraries varies more (up to 21%) and is roughly for two-third due to road distance and for one-third due to road quality.

JEL code: D23, D61, O13, O18, Q13, R41

Key words: transport costs, infrastructure, bridges, agricultural markets, Mozambique, sub-Sahara Africa

***** Wouter Zant is associate professor at the VU University Amsterdam and research fellow of the Tinbergen Institute, the Netherlands; mailing address: Vrije Universiteit, De Boelelaan 1105, room 10A-79, 1081 HV Amsterdam, The Netherlands; e-mail address: wouter.zant@vu.nl; tel: +31 20 598 9592.

Introduction

Trade costs are important for developing countries, both for rural development and economic growth. Most sub-Saharan African countries are plagued with high to extremely high trade costs, which has major implications for the operation of these economies. Since the seminal theoretical work of Key, de Janvry and Sadoulet (2000) and de Janvry and Sadoulet (2006) there is a consistent basis for explaining the rationality of subsistence farming and the key role of transport costs in this outcome, where transport costs affect both inputs and output. Their framework goes a long way in explaining low input levels, low productivity and low technological progress in agriculture. Various subsequent contributions have supplied empirical evidence supporting the idea that transport costs constitute a major driver in subsistence farming (see for example Omamo, 1998). Simultaneously, low trade costs or major reductions in trade costs lead to improvements in the operation of markets, moving agricultural produce more easily from low-price rural surplus areas to high-price urban deficit areas and increasing welfare of consumers. A wide range of studies have empirically investigated to what extent infrastructure – be it road infrastructure (roads, bridges, etc.), rail infrastructure (railroads) or communication infrastructure (ICT, mobile phones) – have improved the efficiency of markets, raised the welfare of households and alleviated poverty. Results so far are mixed, suggesting in some cases that both producers and consumers realize welfare gains (see for example Jensen, 2007), while in others welfare gains for farmers are not evident and traders are likely to benefit most (see for example Fafchamps and Minten, 2012; Zant, 2017). Improved operation of markets, notably of food markets, will, nevertheless, always be helpful in increasing food security.

The topic of the current study is related to several lines of research in the empirical literature, on trade costs (Atkin and Donaldson, 2015), on the impact of transport infrastructure (Donaldson, 2010; Banerjee, Duflo and Qian, 2012; Casaburi et al. 2013; Brooks and Donovan, 2017; Volpe Martincus, Carballo and Cusolito, 2017; Zant 2018), on the operation of markets (Tostão and Brorsen, 2005; Cirera and Arndt, 2008; Zant, 2013, 2017), on prices and transport cost (Minten and Kyle, 1999), and on transport costs and behaviour of households (Jacoby, 2000; Renkow, Hallstrom and Karanja, 2005; Jacoby and Minten, 2009) and on the impact of rail infrastructure on crop prices (Zant, 2018). We follow recent work on the estimation of trade costs with potentially oligopolistic traders (see Atkin and Donaldson, 2015), and investigate specifically to what extent bridges lead to lower trade costs, using the spatial variation of maize prices between major maize markets in Mozambique. Maize is the most widely traded staple food in Mozambique. Estimations allow for spatially varying trader mark-ups and the empirical strategy explicitly addresses identification of source markets and destination markets, and homogeneity of the traded product. For identification we exploit the introduction of a road bridge over the Zambezi river, in August 2009 between Caia and Chimuara, which removed a major obstruction to north-south trade and created the required variation in shortest trading itineraries between markets. We use monthly maize price data of 24 major maize markets, extending five years before and after the introduction of the new bridge. We find a reduction of transport costs, averaged over itineraries, varying from 6% to 10%, and roughly for two-third due to road distance and for one-third due to road quality. The current study is similar to Zant (2018), which investigates the impact of railways on agricultural crop prices in Malawi, exploiting a natural experiment created by the collapse of a railway bridge. Here we look at how the construction of a road bridge, added to the existing road network in Mozambique affects trade costs. This study connects the literature on estimation of impacts of infrastructure with studies on the use of spatial prices to estimate transport costs.

In the remainder of this paper we discuss, in Section 1, the maize market in Mozambique, the Mozambique road network and transport services, and the role of the Zambezi river and Zambezi bridges in the Mozambique economy. In Section 2 we explain the theory underlying the empirical estimation and we present details on data and data sources. In Section 2 we set out the empirical strategy. In Section 4 we present and discuss estimation outcomes and verification with trade cost data. In Section 5 we discuss potential threats and robustness checks of estimation outcomes. We conclude with quantifying benefits of the Zambezi bridge for specific itineraries, attributing cost reductions to road distance and road quality, and a presenting a summary of findings, in Section 6.

1. Maize Markets, Transport Services. Prices and the Zambezi

Maize Markets

Maize is the most important staple food of Mozambique: it is widely produced, marketed, exported and consumed. In all provinces two third of rural households produce maize. Despite widespread subsistence farming – only around 30% of production is traded on the market – maize is three times more marketed than cassava. Also, maize has a budget share of similar size to all other staple foods[1] jointly (Tschirley et al., 2006). The calorie share of maize in the average Mozambique diet ranges from 25% to 39%, corresponding with a per capita (annual) consumption of 60 to 85 kg, although, particularly in the south, and in the Maputo region, the maize share is lower due to substitution with rice (Tschirley et al., 2006). Per capita dietary needs also form an indication of the share of production available for trade between regions: the provinces Niassa in the north, and Manica and Tete in the center are therefore potential surplus areas with sufficient production to supply maize to other regions (see Figure 2).

These surplus regions also characterize the geography of maize production: production of maize is concentrated in the central and northern part of Mozambique. The Northern provinces Niassa, Cabo Delgado, and Nampula have better rainfall distribution (see Figure 2)

[1] Staples in Mozambique are maize, rice, cassava, wheat, sorghum, millet, sweet potatoes beans and groundnuts.

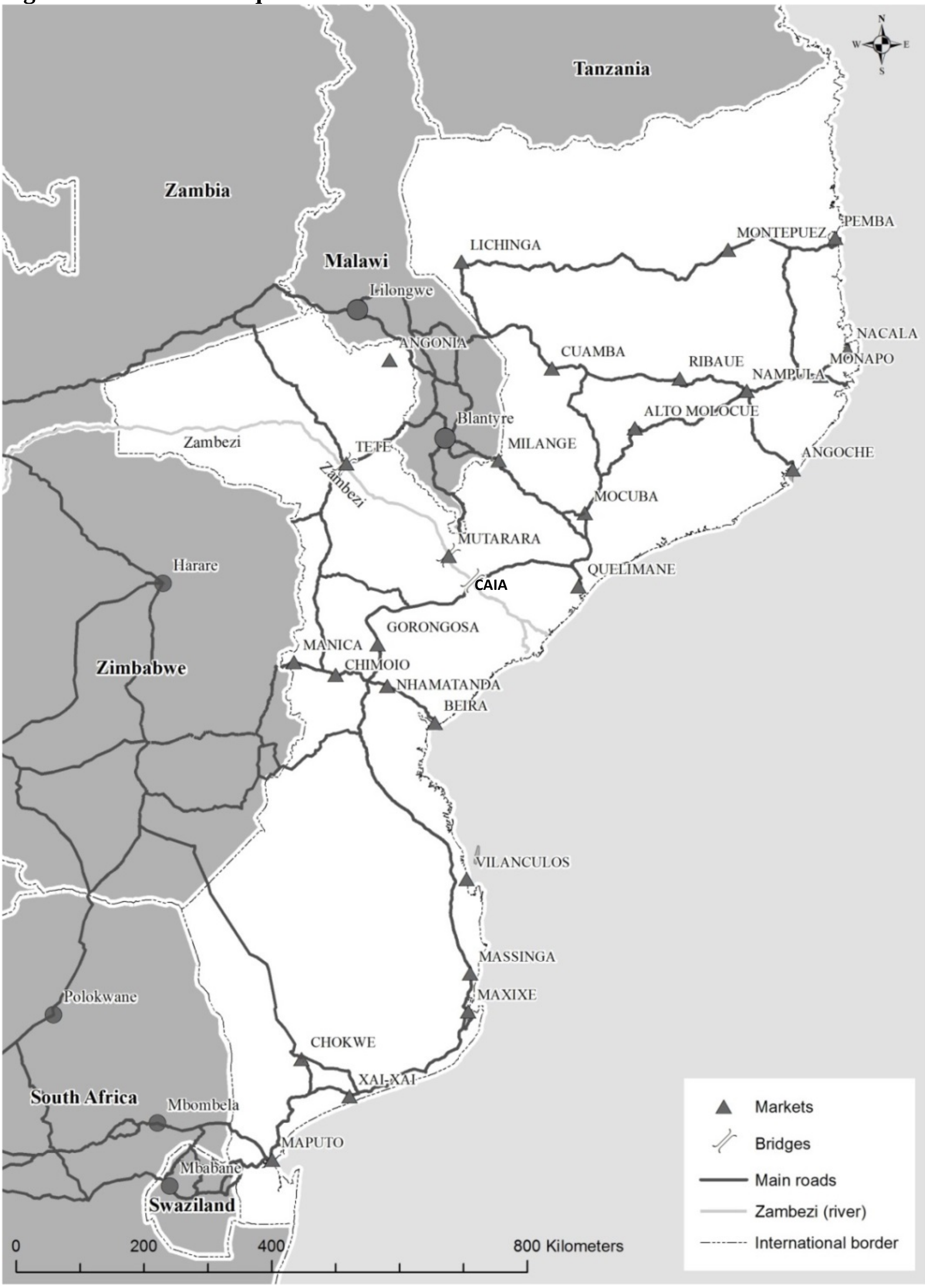
and better soil fertility, while the Southern region has unfavorable weather conditions (see Figure 2) and suffers from occasional pests (Abdula, 2005). Around 60-70% of Mozambique maize production originates from provinces north of the Zambezi river. With a higher total population in the north, per capita production north and south of the Zambezi river is closer together (60kg versus 50kg, average 2005-2014). Most agricultural production in Mozambique is rain-fed. Drought and flooding cause occasional drops in production, and corresponding hikes in prices. In the 1999-2000 crop season, maize production declined 18 percent, primarily due to floods that devastated large areas of the Center and South of the country (Abdula, 2005). Drought and harvest failure in one region may also coincide with good harvests elsewhere, making domestic trade a key factor in food security. Major production, assembly and wholesale markets in the central region are Angonia, Manica and Chimoio, and in the north Alto Molocue, Montepuez, Mocuba and Ribaue (Figure 1). The major retail terminal markets, nearly all located on the coastline, are, from south to north, Maputo (including Matola), Xai-xai, Maxixe, Massinga, Beira, Quelimane, Nacala and Pemba (see Figure 1). The relatively large concentration of population in the Maputo province (see Figure 2) makes this area a major destination of maize trade.

How is the value chain from maize producers to consumers organized? What is size and structure of the various stages in this value chain? And which agents in particular are the key drivers of spatial arbitrage and integration between markets? From producer to consumer, we find retailers, itinerant traders, large scale assemblers, wholesale traders and millers active in the maize value chain. Close to the end of the chain, near consumers, are millers: these are primarily involved in earning returns by value addition rather than earning returns on trade and transport. Wholesale traders take an intermediate position: they buy from assemblers and supply to mills of various sizes. This activity may entail gains from price differences between markets, but is likely to have a large component of value added through collecting, sorting, quality grading and distribution. Retailers specialize in local trade, collecting maize from nearby markets and selling, beyond maize, a wide range of other agricultural crops. The key agents in Mozambique that drive spatial market integration are traders – mostly informal itinerant traders but also large scale assemblers – and transporters (Zovala, 2014; De Vletter and Polana, 2001). Informal itinerant traders are, however, dominant: farmers sell most of their surplus maize to informal small-scale itinerant traders right after harvest, and most of the maize traded in assembly and retail markets in Mozambique, both north, central and south, is supplied by informal traders. Informal traders also carry out most of the marketing functions between the rural producers and the urban consumers: they supply their own working capital, hire storage facilities in source markets and arrange truck transport once a sufficient number of bags with maize is collected. The following quote characterizes the informal itinerant trader:

"Itinerant traders, commonly known as "Mamanas", are normally comprised of women coming from the southern region of the country, mainly from Maputo city, to buy and bulk maize in surplus maize villages of Central and Northern Mozambique with the objective of reselling it in the maize deficit markets of southern Mozambique. Itinerant traders set up buying points in surplus areas. After acquiring large volumes of maize, they rent trucks to transport their maize to southern markets" (Zavale, 2014).

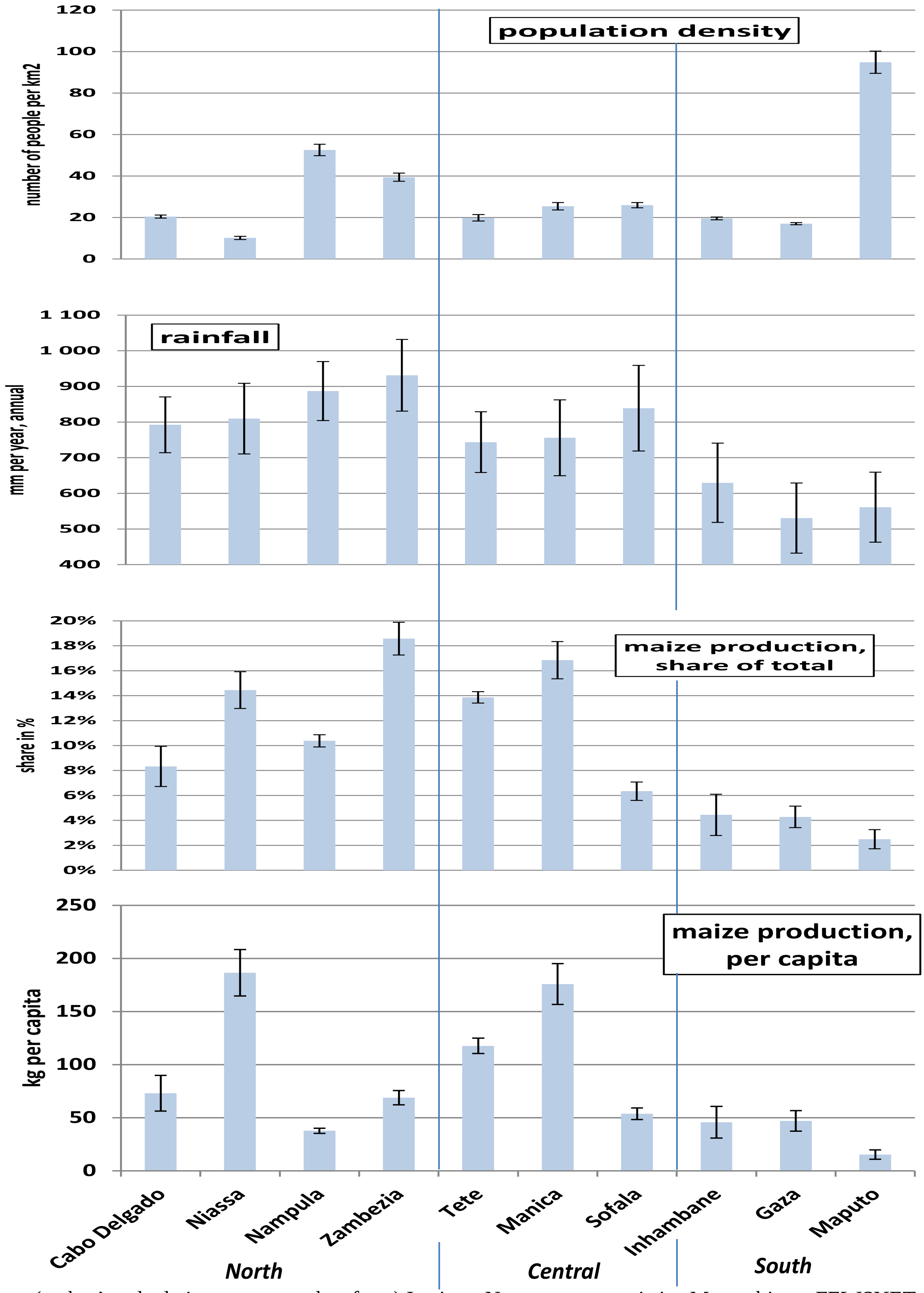
Competition for informal traders comes from formal large warehouse trader companies, who source maize from the same locations and often operate their own (fleet of) trucks (de Vletter and Polana, 2001).

Figure 1 Mozambique: markets and roads



Source: VU-SPINlab

Figure 2 Population density, rainfall and maize production by province



Source: (author's calculations based on data from) Instituto Nacional de Estatistica Moçambique, FEWSNET and Ministry of Agriculture, Early Warning Unit (Aviso Previo); The figure is based on aggregate (average) annual province data. Error bars show 95% confidence intervals. See the maps in the Appendix for the location of provinces.

Note: population data are from 2005 to 2014, rainfall from 1996 to 2012 and production from 2005 to 2012.

Figure 3 Mozambique maize prices, by region, Jan 2005-Dec 2014 (meticais per kg)

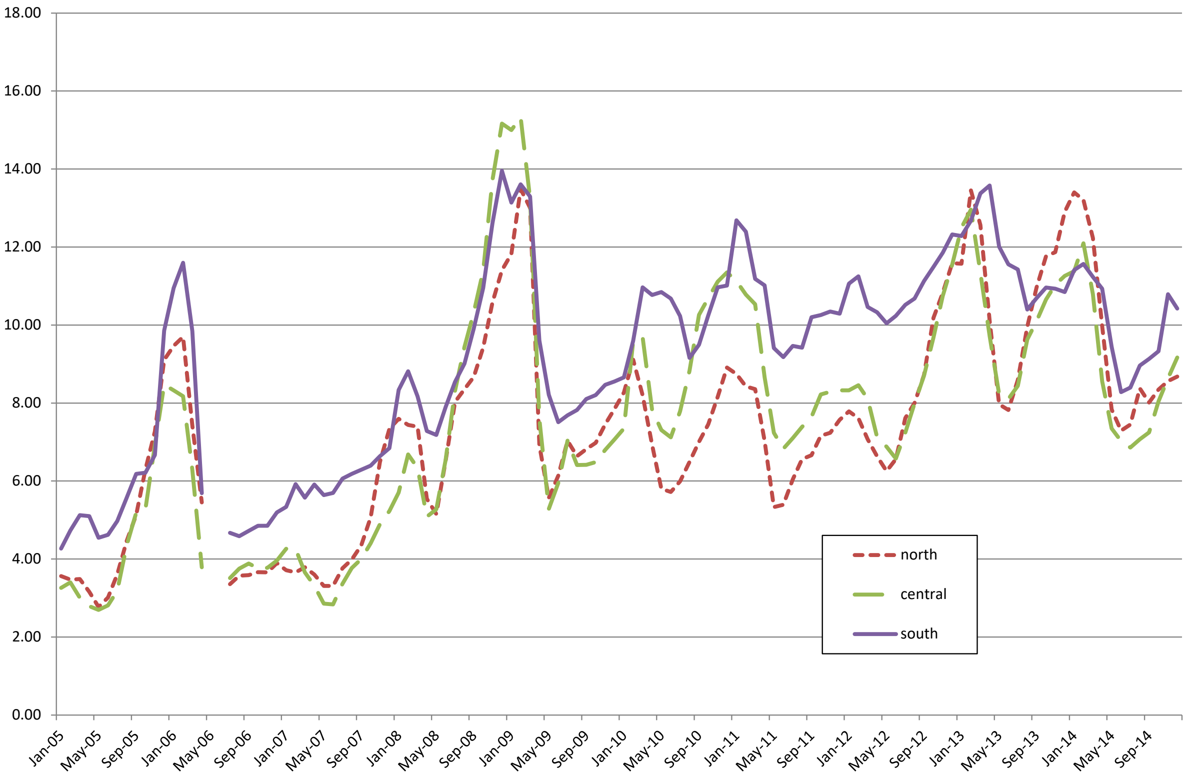
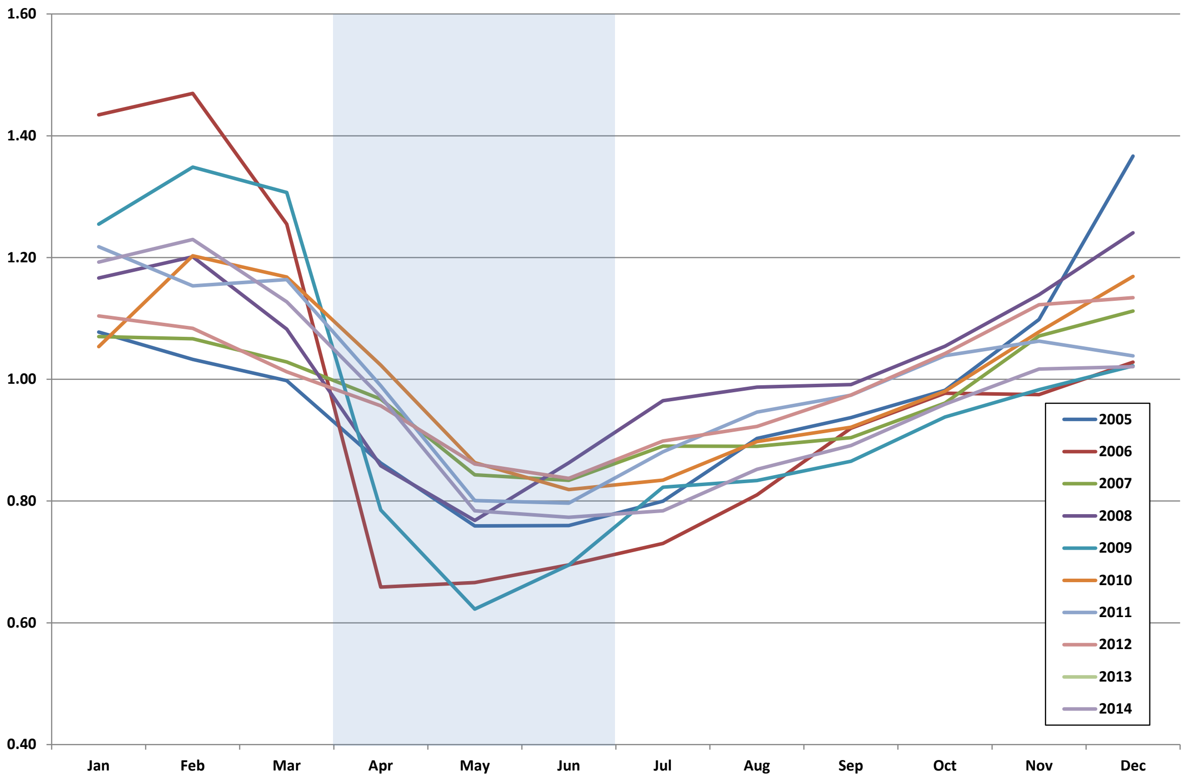
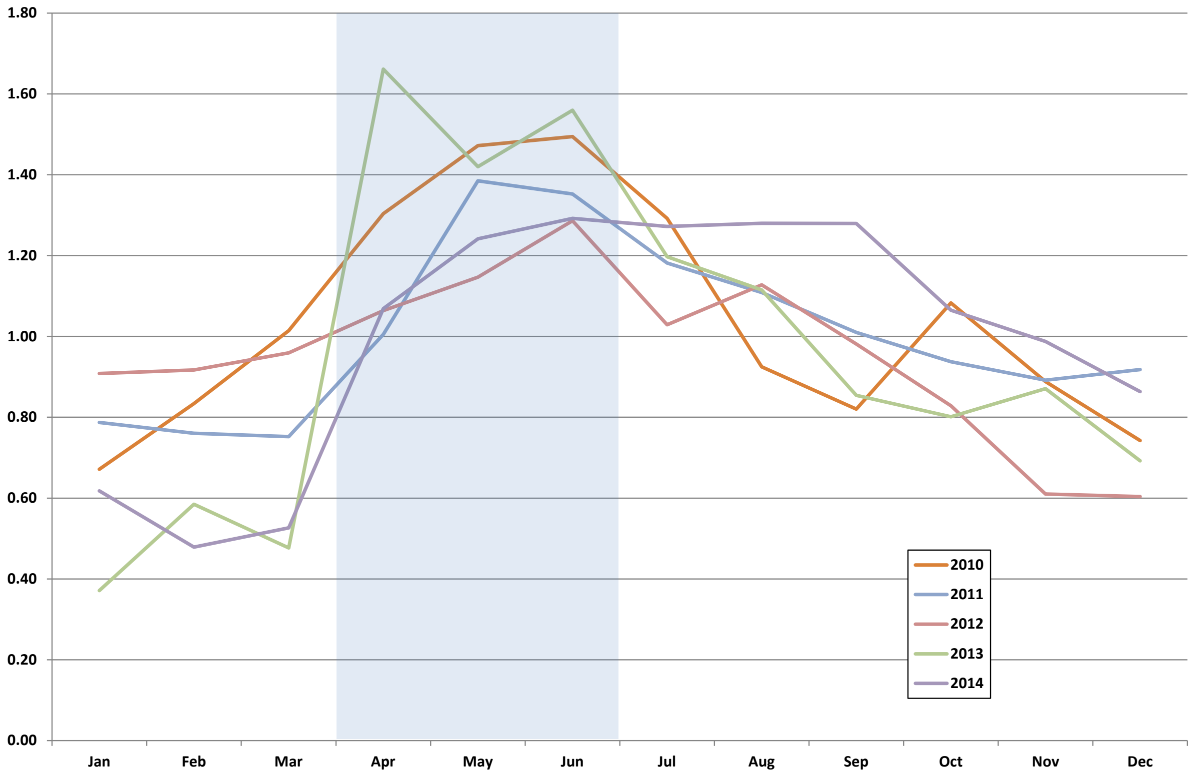


Figure 4 Seasonality in maize prices (index)



Source: authors calculations based on SIMA data; the shaded area indicates the months when trade is most likely.

Figure 5 Seasonality in price difference with Maputo maize prices (index)



Source: authors calculations based on SIMA data; the shaded area indicates the months directly following harvest when trade is most likely.

Maize prices

Maize market prices by region for the period from 2005 to 2014, shown in Figure 3, have a number of characteristics that have implications for the empirical work in this study. The parallel movement in prices of different regions suggests that prices are correlated between regions. Prices in the south are overall higher than in the central and northern region, reflecting the difference between surplus and deficit areas. Prices fluctuate systematically over the season (see Figure 4), with high peaks between January and March, a short but large drop in April-May when the new harvest enters the market, followed by a long gradual increase towards the end of the calendar year. During the lean season – typically characterized by low supply or even shortages – prices rise to levels more than twice as high as directly after harvesting. These movements in prices over the season are common for staple food prices in sub-Saharan countries (see Kaminski et al., 2016; Zant, 2017). Because seasonality between markets is not fully aligned, seasonality also shows up in spatial price differences (see Figure 5): these spatial price differences peak from April to June, directly after harvest when trading activity is high. Jointly with the variation in supply, these fluctuations in prices and spatial price differences point to periods with varying potential benefits from trade. Urban markets typically have higher average price levels (related to the higher level of demand), but lower price variability (related to more storage and alternatives for substitution) and this is confirmed by the data.

Road network and Transport Services

Maize is mainly transported by trucks. Overall, the road infrastructure in Mozambique is not well developed. However, the trunk-road network, connecting main cities and towns (including all maize markets identified in the current study) is in a reasonably good shape during the period of study, particularly after major improvement over the past decades (see Dominguez-Torres and Briceño-Garmendia, 2011). Other roads (secondary roads and feeder roads) are in poor condition and especially during the rainy season many of these roads cannot always be used. Also road density is extremely low, even compared to other sub-Saharan countries (see Dominguez-Torres and Briceño-Garmendia, 2011). The Zambezi river creates a natural barrier to transport by truck: major domestic trade flows of maize are, consequently, from the central area to the south, since Southern Mozambique and most notably the Maputo-Matola area is the major maize deficit area, while the coastal cities north of the Zambezi are supplied by the inland production centers in the north. Angonia in the northeast, also a major production, assembly and wholesale market, supplies Tete, while maize surpluses in both Angonia and Lichinga region are exported to Malawi (USGS / FEWS NET) where prices often are higher (see Appendix, Figure A3).

Maize available for sale in wholesale markets in Maputo (Xiquelene and others) is, amongst other locations, sourced from the central region, Nhamatanda, Chimoio or Manica, around 1100 km by road (Abdula, 2005; SIMA data from 1999-2001), or even as far as Tete, around 1500km by road from Maputo (Tostão and Brorsen, 2005; SIMA trade flow data from 1998-2001). Transport cost data from SIMA – unfortunately only fragmentary available (see *Data and data sources*) – pertain to itineraries between markets as far as 2300km apart (notably Lichinga and Maputo). Southern Mozambique, and the Maputo-Matola area in particular, also rely on South Africa as a major supplier of maize (see Haggblade et al., 2008), where prices often are lower (see Appendix, Figure A2).

The Zambezi river and its bridges

The Zambezi is a large river, that originates in the outer north of Zambia, near the Congo border, runs through Angola, re-enters Zambia, then forms the border between Zambia and Namibia, and Zambia and Zimbabwe, enters Mozambique in the east at the Cahora Bassa reservoir, and runs subsequently southwest to empty in the Strait of Mozambique (Indian Ocean). In the world the Zambezi ranks 31 in terms of length, 16 in terms of drainage area and 37 in terms of average discharge. Relative to the river Rhine in Europe, the Zambezi is more than twice as large in terms of length (2693km versus 1233km) and average discharge (4880m^3s versus 2330 m^3s) and 6.7 times as large in terms of the drainage area (1330000km^2 versus 198735 km^2; source: Wikipedia). The Zambezi is the fourth largest river in Africa, on these accounts. The within Mozambique length of the Zambezi is, however, only around 800km, nearly a third of its total length. Also the width of the river is substantial: in broad valleys the river spreads out over a wide area and is 5km to 8km wide and, not unimportant for transport by river, shallow in many places.

How does the Zambezi river affect the economy of Mozambique? A key contribution of the Zambezi river is power generation at the Cahora Bassa hydropower plant (completed in 1973), that supplies power to both Mozambique and South Africa. However, in terms of transport the river offers much less. Although the river is navigable, commercial long-distance transport by river is not well-developed, mainly due to its unreliability. As indicated above, since crossing the river is a major barrier, the Zambezi river does have a negative impact on within-Mozambique north-south trade and transport by road. This brings us to the, currently, four bridges that cross the Zambezi. The bridge at Tete is a suspension bridge that is in operation since 1973. It is integrated in the highway network and forms a major gateway – in the northern direction – to Zambia, Malawi and the northern part of Mozambique, and – in the southern direction – to Zimbabwe, South-Africa and the southern part of Mozambique. Since the bridge at Tete for a long time has been the only way to cross the Zambezi, the bridge has become a bottleneck in road transport plagued by severe congestion. In the course of the past decade an additional bridge in Tete has been constructed, which was opened for traffic in 2014.

On August 1, 2009 a new road bridge over the Zambezi River was opened, between Caia and Chimuara, linking Sofala and Zambezia provinces in the centre of the country[2]. The bridge is part of the main north-south highway, and connects major commercial centers in the north (e.g. Nacala, Nampula, Mocuba) and south (e.g. Beira, Chimoio). Tete, where the other road bridges over the Zambezi are located, is around 300km kilometers to the northwest. The construction of the bridge began in March 2006, but already in 1979/80 work had started on the access roads. War and conflict caused major delays in completing the work. The new bridge at Caia has replaced a ferry service between Caia and Chimuara that operated until the completion of the bridge. The 2009 toll for the new bridge is the same as motorists had to pay for using the ferry – 800 meticais (around US$30 dollars) for trucks and 80 meticais for light vehicles. In the period before the introduction of the bridge, the ferry was widely perceived as inefficient due to long waiting times and extensive queues of trucks, causing high risks of spoiling perishable crops in case of food transport. Truck transport often encountered delays

[2] Mozambique has a tradition to tag (presidents) names to bridges. The bridge between Caia and Chimuara is named after president Armando Emilio Guebuza. The original bridge at Tete is known as the Samora Machel bridge and the bridge between Vila de Sena and Muturara as the Dona Ana bridge. In this study we identify bridges by the nearby town or village.

for days or weeks for the ferry trip between Caia and Chimuara. Engine breakdowns and sensitivity to tides made the ferry connection unreliable. Tostão and Brorsen (2005) report: "…in early 2001 the ferry was shut down for nearly two months because the Zambezi river was flooding, and in July 2001 the ferry service was interrupted again because there was not enough water in the river". We conclude that transport by ferry across the Zambezi is very uncertain, thereby obstructing systematic trade. Transport by ferry appears therefore unlikely to have had a major impact on local commodity markets.

Only around 60 km upstream, there is another bridge, a railway bridge, spanning the lower Zambezi River between the towns of Vila de Sena and Mutarara. This bridge was originally built by the Portuguese in 1934 during the Portuguese rule of Mozambique, in order to link Malawi and the nearby Moatize coal fields to the port of Beira. At the time, this 3.7 kilometers long Vila de Sena – Mutarara bridge was the longest railway bridge in Africa. Although not located on a primary highway, it provided an alternative route to pass the Zambezi river, next to the bridge at Tete and the former ferry between Caia and Chimuara. Unfortunately, the bridge was rendered unusable in the 1980s, during the Mozambican Civil War. After the ending of the civil war in 1992, USAID assisted with the repairs and it was converted to a single-lane bridge for vehicle traffic. More recently, in October 2006, the bridge was completely closed to vehicle traffic for rehabilitation and (re-) conversion into a rail bridge: it was re-opened as a rail bridge in August 2009.

It is not exactly clear to what extent the Vila de Sena - Mutarara bridge was effectively used as a road bridge, before it was closed for traffic to be rehabilitated as a railway bridge in October 2006. The Vila de Sena - Mutarara bridge is, however, not integrated in the major trunk roads of Mozambique (see Figure 1), which makes it not useful for regular trade. Whatever is the case, the "closing for rehabilitation date" suggests a clear period of being closed to traffic of 34 months before the re-opening in August 2009. This coincidental correspondence in timing (the introduction of a road bridge between Caia and Chimuara, and the completion of the rehabilitation of the railway bridge between Vila de Sena and Mutarara) allows to mark a period in which absence of these Zambezi bridges overlaps, creating the required variation in shortest trading itineraries by road between markets.

2. Theory, Data and Empirical Strategy

Underlying theory

We propose the following set-up for spatial prices, taken over, with minor changes, from Atkin and Donaldson, 2015. Consider that p_d and p_o denote price at destination and origin locations, then:

$$p_d = p_o + \tau(X_{od}) + \mu_{od} \tag{1}$$

where the price in a destination location (p_d) is the sum of the price at the origin location (p_o), transport costs $\tau(X_{od})$, and a mark-up (μ_{od}). In terms of spatial price gaps and following Atkin and Donaldson (2015), we re-write this equation as:

$$p_d - p_o = \tau(X_d) + \mu(c_d, \phi_d, D_d) \tag{2}$$

indicating that the markup is a function of the traders' marginal cost c_d, the competitive environment faced by traders, summarized by the competitiveness index ϕ_d, and demand conditions D_d. This equation, and this is its key message, expresses that mark-ups $\mu(..)$ and transport costs $\tau(X_d)$ are correlated. Using this expression, the effect of a small change in a cost shifter x_d on the spatial price gap follows:

$$\begin{aligned} \mathrm{d}(p_d - p_o)/\mathrm{d}(x_d) &= (1+\partial\mu/\partial c_d).\partial\tau(X_d)/\partial x_d + \partial\mu/\partial\phi_d.\partial\phi_d/\partial x_d + \partial\mu/\partial D_d.\partial D_d/\partial x_d \\ &= \rho_d\ \partial\tau(X_d)/\partial x_d + \partial\mu/\partial\phi_d.\partial\phi_d/\partial x_d + \partial\mu/\partial D_d.\partial D_d/\partial x_d \end{aligned} \tag{3}$$

where ρ_d is known as the pass-through rate. The pass-through rate is defined as the effect of traders' marginal cost on prices while holding competitiveness (ϕ_d) fixed. Atkin and Donaldson (2015) show that, in general, the pass-through rate is determined by competitiveness and the curvature of the demand curve. Most of the empirical literature refers to a set-up where mark-ups are independent of costs ($\partial\mu/\partial c_d = 0$) resulting in $\rho_d = 1$. In this special case spatial price difference reflect transport costs and may be used directly in estimations. Alternatively, in all imperfectly competitive settings, mark-ups will be correlated with costs ($\partial\mu/\partial c_d > 0$, and $\rho_d \neq 1$), some of the marginal cost will be passed through to prices (hence $\rho_d > 0$), but whether a lower ($\rho_d < 1$) or higher ($\rho_d > 1$) than complete pass-through is optimal is not determined. The second and third term in the derivative expression (equation (3): $\partial\mu/\partial\phi_d.\partial\phi_d/\partial x_d$, and $\partial\mu/\partial D_d.\partial D_d/\partial x_d$) capture that mark-ups vary across locations because of differences in competitive conditions across locations (ϕ_d) and because of differences in preferences across locations (D_d). Atkin and Donaldson (2015) proceed to argue that fairly reasonable assumptions – competitiveness ϕ_d may vary across locations but is fixed within a location, and consumer preferences (D_d) are such that the curvature of the slope of the inverse demand curve is constant (Bulow-Pfleiderer demand) – are sufficient for constant pass through rates ρ_d. The previous price gap equation, jointly with these assumptions can be re-written as follows:

$$p_d - p_o = \rho_d\, \tau(X_d) + (1- \rho_d)(a_d - p_o) \qquad (4)$$

where a_d is a demand shifter. A constant pass through turns out to be particularly useful to estimate trade costs in the presence of varying mark-ups: this expression shows that the pass-through rate ρ_d and the demand shifter a_d are sufficient to control for the bias arising because of unobserved preferences (D_d) and market structure (ϕ_d). Introducing time, and assuming that the pass-through rate ρ_d varies across source-destination combinations, but is fixed over time, and taking account of variable sources of products, we may write, after some reshuffling:

$$p_{dt} = \rho_{od}\, p_{ot} + \rho_{od}\, \tau(X_{odt}) + (1-\rho_{od})\, a_{dt} \qquad (5)$$

In order to extract the pass-through rate from this equation we need to have data on transport costs ($\tau(X_{odt})$) and local demand shifters (a_{dt}). As these variables are not known to the researcher, we assume that both can be approximated with a local time invariant factor, a local trend factor and a residual factor, or formally:

$$\tau(X_{odt}) = \beta_{1od} + \beta_{2od}\, trend + \zeta_{odt} \tag{6}$$

$$a_{dt} = \alpha_{1d} + \alpha_{2d}\, trend + v_{dt} \tag{7}$$

Combining the last three equations yields:

$$p_{dt} = \rho_{od}\, p_{ot} + \gamma_{1od} + \gamma_{2od}\, trend + \varepsilon_{dt} \tag{8}$$

Now we can estimate the pass through rate ρ_{od} for each source-destination combination. This is the first step in adjusting spatial price differences to extract transport costs from these price data. The second step is to use estimated values for pass-through rates by source-destination, to construct adjusted variables and use these adjusted variables to estimate determinants of transport costs. Hence, we estimate:

$$(p_{dt} - \widehat{\rho_{od}}\, p_{ot})/\widehat{\rho_{od}} = \tau(X_{odt}) + \alpha_{1d}\, ((1-\widehat{\rho_{od}})/\widehat{\rho_{od}}) + \alpha_{2d}\, trend\, ((1-\widehat{\rho_{od}})/\widehat{\rho_{od}}) + \varepsilon_{dt} \tag{9}$$

where a hat on the pass-through rates ρ_{od} denotes its estimated value. The presented framework, the resulting two-step procedure and notably equation (8) and (9), are the backbone of the empirical estimations.

The conceptual framework raises various empirical issues: in the first place we need to ascertain that the "source market of the price differential" is the market where the product is actually sourced. We have identified source and destination markets by exploiting information on source and destination in trade cost data, on the availability of producer prices, on the local balance of maize production and maize demand, on the size of the population, and on being located in rural areas or on the coastline. See **Empirical strategy** for more details. Next, we need to find a convincing set of variables that reflects transport costs ($\tau(X_{odt})$). We use road distance and road quality as major determinants of real transport costs. Again, see **Empirical strategy** for more details. Finally, prices should pertain to a homogeneous good with limited quality differences. From the background section we know that white maize grain is produced, consumed and traded throughout Mozambique. Maize prices are recorded without specifications for quality: apparently one kg of white maize grain, in Maputo or Nampula, or in 2005 or 2014, or both, is not different in terms of quality. Therefore, and without denying possible quality differences, we claim that the requirement of a homogenous product is satisfied and potential bias due to unobserved quality differences in the estimations is negligible.

Data and data sources

Market prices for maize are sourced from Sistema de Informação de Mercados Agrícolas de Moçambique (SIMA; www.masa.gov.mz/sima), from the weekly publication Quente-Quente[3]. We use in particular weekly retail prices of white maize grain[4], originally recorded for 27 markets[5], for the period from January 2005 to December 2014. This period covers a timespan of around five years with and without bridge. Price data are collected by interviewing randomly selected traders in each market. Weekly prices are averaged to monthly data. Unfortunately, there are missing observations, even in the monthly averages: missing observations are, however, common in agricultural prices series, due to lack of supply and corresponding absence of transactions, and are not correlated with the presence of the Caia-Chimuara Zambezi bridge. The share of missing observations is also relatively small: after dropping a few

[3] SIMA, which started in the 1990s as a USAID / Michigan State University funded initiative, is responsible for collection and distribution of price information on agricultural commodities, and distributes weekly price bulletins by email (covering amongst others farmer organizations, traders), by SIMA's provincial offices (that further reproduce and distribute this information locally), through the Ministry of Commerce that uses the information in their own bulletins, and through regular broadcasts on the national radio and television news (to whom SIMA contractually offers weekly input to market programs). Traders interviews support the effectiveness of the SIMA price information (see "In Mozambique, Market Information publishes its 500th weekly bulletin, a Cause for Celebration", February 2006 posted on the internet (www.fsg.afre.msu.edu / press / SIMASuccess500.pdf or www.masa.gov.mz/sima/). The rollout of the mobile phone infrastructure that started in the 1990s has further improved the dissemination of price information (see Zant, 2017).

[4] Quadro 3, Preço e Mudança Percentual a Nível de Mercado Retalhista (MT/kg), Grão de Milho Branco (Table 3, Prices and percentage price changes in retail markets (meticais per kg), white maize grain)

[5] Alto Molocue, Angoche, Angonia, Beira, Chimoio, Chokwe, Cuamba, Gorongosa, Lichinga, Manica, Maputo, Massinga, Maxixe, Milange, Mocuba, Monapo, Montepuez, Mutarara, Nacala, Nampula, Nhamatanda, Pemba, Quelimane, Ribaue, Tete, Vilanculos en Xai-Xai. Figure 1 shows the locations of these markets in Mozambique.

markets[6], we have around 88% of the potential number of monthly price observations (2533 / (24 markets x 12 months x 10 years)). The construction of spatial price differences – the price dispersion measure that is used in the estimations – blows up the number of available observations to large numbers[7].

Transport cost data are, for a limited period and a limited number of itineraries, also available from SIMA. Collection of these data is organized similarly to the collection of price data, by asking quotations from randomly selected traders in major source and destination markets. Transport costs are specified by date, itinerary (i.e. source market and destination market), product and weight of the bags transported. We only use transport cost data for maize grain. Transport costs are recorded for the period from August 2001 to December 2010, with nearly three quarter of the observations before 2005. After 2010 – and for undisclosed reasons – the publication of tables with transport cost data becomes less frequent, and also focuses less on long distance trade. Moreover, available transport costs series are also only for a limited number of itineraries (see Appendix, Table A2). In all we have around 800 observations of transport costs, that cover the period of study (January 2005 -December 2014) to a limited extent (see Appendix, Figure A6). The total number of transport cost observations is only a fraction (<10%) of the (adjusted) price differences. Further, and more importantly, there are no observations of transport costs after July 2009, of itineraries that cross the new bridge.

Road distance and travel time are both taken from Google Maps, at the time of writing of this study (2017). Since our study period is from January 2005 to December 2014, this may entail measurement error: we do not incorporate construction of new roads, and road

[6] A few markets of the orginal data (Angoche, Monapo and Vilanculos) are dropped altogether since these markets have no data in the period under consideration.

[7] Price data for n markets in a specific month, yields n^2 market pair data, of which $(n^2–n)/2$ are economically relevant, because trade runs from a to b or vv.; hence, with 24 markets one month without missing observations yields 276 market pair data $((24^2-24)/2)$. Note that we further restrict the sample of data for estimations to price differences that connect source markets with destination markets (see also *Further empirical issues*).

maintenance and rehabilitation of existing roads (other than changes due to the introduction and rehabilitation of the bridges). Fortunately, the trunk road network – the road network that connects all major the markets identified in this study – is well maintained and has not changed much during the period of study (see Dominguez-Torres and Briceño-Garmendia, 2011). Whatever change remains – and most importantly – developments in road infrastructure are controlled for through our DiD estimation strategy: common time varying shocks and trends are automatically digested by the time fixed effects.

A few other variables are used, in particular in the estimations with covariates and to estimate the propensity score: data of population by city or town, are from three censuses (1997, 2007, 2016) from the Instituto Nacional de Estatistica Moçambique. Population data for intermediate months and years are constructed by interpolation. Fuel prices are from the International Financial Statistics from the IMF. Maize production by province is from Trabalho de Inquérito Agrícola / Anuario de Estatistica Agararia, Ministry of Agriculture. Maize production data are incomplete: missing years are constructed (see Appendix, Table A1). We have used quotations of the nearest SAFEX white maize futures contract as representative for maize prices in South Africa. These series are taken from the SAFEX web site and converted to Mozambique meticais with the help of IMF/IFS monthly exchange rates (period average, domestic currency per US$). Finally, we use market prices of maize grain from Lilongwe, Malawi, taken from FAO, also converted to Mozambique meticais with the help of IMF/IFS monthly exchange rates. It is assumed that Lilongwe maize prices represent price developments in the Malawi maize market.

3. Empirical strategy

We postulate that real transport costs are empirically determined by road distance and road quality: these two variables jointly represent transport costs ($\tau(X_{odt})$). Following the conceptual framework set out in the *Underlying Theory* section, we further take account of variations in mark-ups across space. In summary, to measure the impact of the Cena and Mutarara bridges on the determinants of transport costs, we start with the following differences-in-differences (DiD) specification:

$$p_{jt}/\widehat{\rho_{jk}} - p_{kt} = \eta_0 + \eta_1\, roaddistance_{jkt} + \eta_2\, roadquality_{jkt}$$

$$+\ \theta_{jk}\left((1-\widehat{\rho_{jk}})/\widehat{\rho_{jk}}\right) + \omega_j\, trend_{jk}\left((1-\widehat{\rho_{jk}})/\widehat{\rho_{jk}}\right)$$

$$+\ \boldsymbol{X}_{jkt}\gamma + \psi_t + \varphi_{jm} + \varepsilon_{jkt} \qquad (10)$$

In this equation $roaddistance_{jkt}$ is the shortest road distance between markets j and k, at time t. We expect that an increase in road distance increases transport costs, and hence $\eta_1>0$. The variable *roadquality* is specified as the number of kilometres realised per hour of travel time or average speed: a high value indicates a good road quality, a low value a poor road quality. We expect that increased road quality decreases transport costs, and hence $\eta_2<0$. The variables road distance and road quality are the real determinants of transport costs. The identification of market pairs that realise a change in shortest road distance due to the bridge (the intervention pairs) is documented below. The RHS variables in the second line of equation (10), jointly with the dependent variable, contain the transformations of spatial price difference, trade pair fixed effect and trend that account for variations in mark-ups across space. In the third line, the vector $\boldsymbol{X}_{jt}$ represents variables that possibly influence the (adjusted) spatial price difference, such as rainfall in source areas, population densities, foreign maize prices, access to imports, domestic transport by sea, etc. Parameters ψ_t are time fixed effects (months), φ_{jm} represents seasonality in (both source and destination) market j, and takes the value 1 for each month (January, February, etc) and zero elsewhere and ε_{jkt} is an error term with zero mean. The parameters of interest are η_1 and η_2 which measure the contribution of road distance and road quality to trade costs.

The first stage regression, required to find trade pair specific estimates of the pass through rate (ρ_{jk}) comes straight from the conceptual framework, and is estimated for each combination of source and destination separately:

$$p_{jt} = \rho_{jk}\, p_{kt} + \gamma_{1jk} + \gamma_{2jk}\, trend + \chi_{jm} + season_t + \varepsilon_{jt} \qquad (11)$$

where χ_{jm} represents seasonality in market j, and takes the value 1 for each month (January, February, etc.) and zero elsewhere and $season_t$ represents between year seasonality. Agricultural crop prices are notorious for seasonality, and maize prices in Mozambique are no exception. Seasonality in prices also translates into seasonality in spatial price differences (see section on *Maize prices*). Moreover, despite the large common component in seasonality in maize prices and price differences, there is considerable variation in timing and amplitude, both within market(pair)s, between years, and between market(pair)s, for the same years. To accommodate for this seasonality, we have included source and destination specific seasonality, in both estimation steps (respectively φ_{km} and χ_{jm}).

Further empirical issues: measuring the shortest road distance / identifying intervention pairs

The key variable that governs impact is the minimum road distance between markets. The pre-intervention period is the period prior to August 2009, the month when the bridge became operational. We have assumed that prior to August 2009 any north-south (and south-north) road transport crossing the Zambezi was directed through Tete and crosses the Zambezi via the bridge in Tete. With this assumption the shortest transport routes for north-south (and south-north) trade in maize in Mozambique are uniquely determined. The intervention market pairs are those market pairs that realise a reduction in shortest transport routes as a result of the introduction of the bridge. By way of example: maize transport from Alto Molocue, north of the Zambezi, to Beira, south of the Zambezi (see Figure 1) involved a 1268km (18 hours, 39 minutes) journey before, and a 749km (11 hours, 5 minutes) journey after the introduction of the bridge, a decrease of more than 40% in both road distance and travel time. Likewise, maize transport from Chimoio (south of the Zambezi) to Quelimane (north of the Zambezi) involved a 1005km journey before and a 564km journey after the introduction of the bridge, again, a decrease of more than 40% in road distance and travel time. The scheme below summarizes the number of observations for the different groups[8].

	before August 2009	August 2009 and later
intervention pair	1947	2640
non-intervention pair	3548	4684

The determination of shortest transport routes implicitly also assumes that freight from north to south (or vice versa) that is transported either through the Vila de Sena-Mutarara bridge – before March 2006, when it operated as a road bridge – or through the ferry that operated between Caia and Chimuara, or through any other informal crossings of the Zambezi river, is of negligible size and has a negligible impact on maize markets, maize market prices and geographical price dispersion. Several sources confirm the poor operation of the ferry (see Section 1), while the Mutarara bridge is not integrated in the road network.

Further empirical issues: how to identify source and destination markets?

To extract transport costs and mark-ups from price differences, we need to identify source and destination markets. We introduce a few simplifying assumptions: we assume that markets are either a source markets or a destination market, and this does not change during the sample period (2005-2014). The latter assumption is straightforward: surplus producer areas and deficit consumer markets are fixed and cannot swiftly transition in the Mozambique context, either from source to destination or from destination to source, neither within season and between seasons[9].

[8] Without restrictions, hence, all road distances and all months.

[9] We are aware that households tend to sell during the months directly after harvest, when prices are low, and often need to purchase in the lean season, when prices are high. However, we are not aware and have no evidence that this effectively leads to long distance reversals of trade flows. In this context it is hard to imagine that maize grain moves from rural areas (e.g. Cuamba) to cities (e.g. Nampula) following the months after harvest, and moves back again during the lean season if prices in rural areas increase faster than in urban areas. During the lean season there is a shortage of supply and, hence, also a shortage of maize to trade. Whatever maize there is fetches a high price in any market. As indicated in a previous section: Urban markets typically have higher average price levels (related to the higher level of demand), but lower price variability (related to more storage and alternatives for substitution).

The first assumption is more difficult to justify as some markets operate both as destination market for a subset of locations, and as source markets for another set of markets[10]. This applies specifically to wholesale assembly markets, like Nampula and Tete and transit markets like Mocuba, Montepuez and Nhamatanda. This is the strategy we follow: we use, in the first place, the source and destination designation in transport cost data because these data, supplied by traders, convey useful information about source and destination markets of actual trade flows. Hence, we decide to identify a market as a source (destination) market if the number of source (destination) market designations in these data dominates (see Appendix, Table A4). This rule determines for most cases if a market is either source or destination market[11]. Complementary to this criteria, a few supply and demand indicators are used: markets with high per capita production, located in rural areas, with a low population and for which producer prices are recorded, are identified as source markets, and markets with low per capita production, without recorded producer prices, with a large population, and located on the coastline are identified as destination markets (see Appendix, Table A4). We end up with 14 source markets and 10 destination[12]. There are a few border cases that warrant further experimentation. Nevertheless, we can be reasonably confident and certain, on the basis of the applied criteria, that the proposed distinction truly identifies source and destination markets[13].

---

[10] Our approach is a crude approximation and cannot account for the myriad of trade flows and trade relationships, that include trade reversals, both between and within season, and markets simultaneously having a supply and demand function. A slightly more flexible way to accommodate for markets that are both source and destination is to determine source and destination for each market pair individually, for example, by using the share of relative spatial price differences by pair: if share(p_j<p_k) larger (smaller) than 75% (25%) than market j is a source (destination) market vis-à-vis market k. This allows markets to play different roles vis-à-vis different markets. The strategy employed in the current paper is more simple but a useful approximation.

[11] The transport cost data suggest that both Nampula and Tete are also major source markets, rather than (only) destination markets (as proposed here). Along these lines we have experimented with variation in the identification of source ad destination markets.

[12] Destination markets are: Beira, Maputo, Massinga, Maxixe, Nacala, Nampula, Pemba, Quelimane, Tete, Xai-Xai; Source markets are: Angonia, Alto Molocue, Cuamba, Chimoio, Chokwe, Gorongosa, Manica, Mocuba, Montepuez, Mutarara, Nhamatanda, Ribaue.

[13] At the same time, it is fair to add that identification of source and destination in so-called barcode level data (see Atkin and Donaldson, 2015; Broda and Weinstein, 2008), where the production factory location of specific domestically manufactured goods and the port of entrance of specific imported goods are recorded, is without doubt more accurate than the approach followed in this study.

Differences-in-differences with a binary impact variable and parallel trend assumption

For several reasons it is useful to re-estimate the differences-in-differences specification with a binary impact variable, rather than with the variables road distance and road quality. We estimate:

$$p_{jt}/\widehat{\rho_{jk}} - p_{kt} = \eta_0 + \eta_1\, bridge_{jkt}$$

$$+\theta_j\, ((1-\widehat{\rho_{jk}})/\widehat{\rho_{jk}}) + \omega_j\, trend_{jk}((1-\widehat{\rho_{jk}})/\widehat{\rho_{jk}})$$

$$+ \boldsymbol{X}_{jkt}\,\gamma + \psi_t + \varphi_{jm} + \varepsilon_{jkt} \qquad (12)$$

where $bridge_{jkt}$ is equal to 1 in period t if the shortest route from j to k runs via the Caia-Chimuara bridge when this bridge was operational, and zero otherwise. All other variables are the same as in equation (10). Impact is now expected to have a negative coefficient reflecting the reduction in transport costs, averaged over itineraries. The differences-in-differences approach requires that that pre-intervention outcomes of intervention and control groups have a parallel trend. Following standard practise (see Autor, 2003) we test the parallel trend assumption by adding a set of interactions of market pair dummies of those market pairs that benefit from the bridge, both before and after the bridge introduction, with time-period dummies. Formally we estimate:

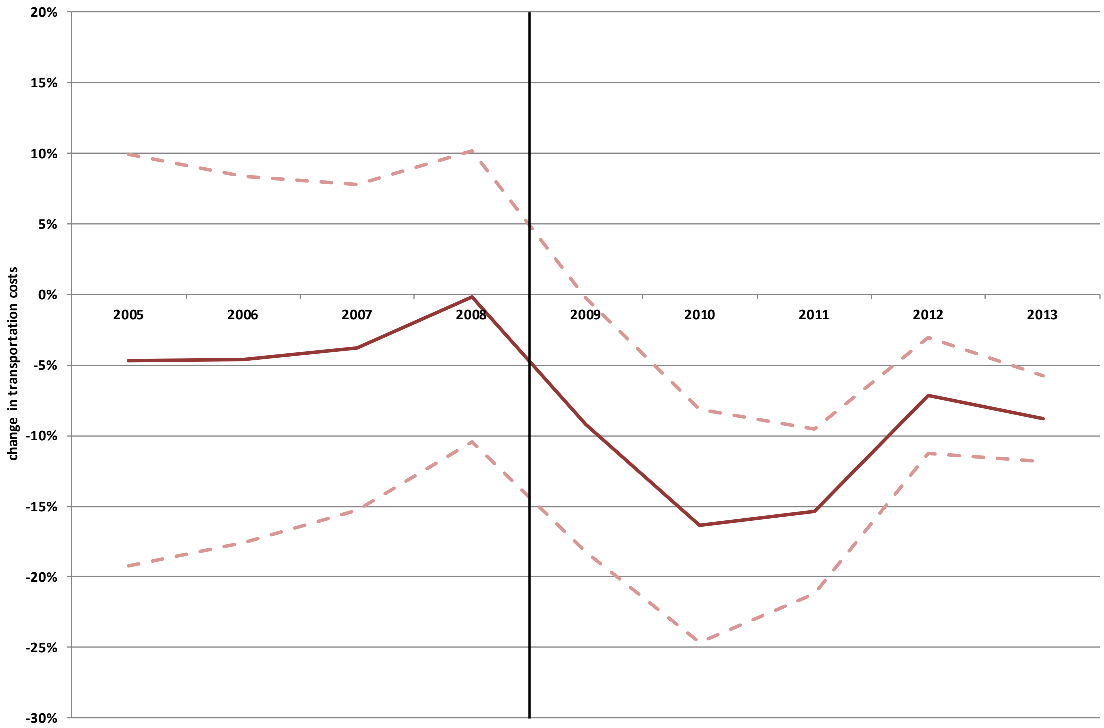
$$p_{jt}/\widehat{\rho_{jk}} - p_{kt} = \eta_0 + \eta_1\, intervention\ pairs_{jkt} + \Sigma\eta_{2time}\, (intervention\ pairs_{jkt}\ x\ time)$$

$$+\theta_j\, ((1-\widehat{\rho_{jk}})/\widehat{\rho_{jk}}) + \omega_j\, trend_{jk}((1-\widehat{\rho_{jk}})/\widehat{\rho_{jk}})$$

$$+ \psi_t + \varphi_{jm} + \varepsilon_{jkt} \qquad (13)$$

where *intervention pairs* are the market pairs, in all time periods, that benefit from the introduction of the bridge, after its introduction, and *time* is a time fixed effect, either month, quarter or year. If coefficients of the interaction terms before the introduction of the bridge are statistically insignificant, all time trends and shocks are absorbed by the time fixed effects, for both intervention and non-intervention pairs, and the estimation result supports the parallel trend assumption. Jointly with the parallel trend, the graphical evidence also shows if coefficients are statistically significant (and negative) after the introduction of the bridge, and offers information on the dynamic path of impact (stable? decreasing or increasing over time?).

Figure 6 Testing for a common trend in the pre-treatment period



Note: the dotted lines indicate 95% confidence intervals

The outcome of this exercise, shown in Figure 6, confirms positive and statistically significant impacts after, and non-significant impacts before the introduction of the new bridge. Consequently, on the basis of the figure we cannot reject the hypothesis of a parallel trend in the pre-treatment period for intervention and non-intervention observations. The figure further confirms the consistency of impacts of a reduction of around 10% over the years after 2009. Figure 6 is a major result of this study.

4. Empirical Estimation, Covariates and Verification

Estimation of a basic specification

The estimation results of the first stage regression – the estimations needed to construct a the pass-through rate by market pair – are not reported (since we have 14 source and 10 destination markets this means presenting estimation output of 140 regressions; estimation output is available from the author on request). In the Appendix, Figure A4, we do show the estimates of the pass-through rate plotted against road distance, for all combinations of source and destination. Similar to the estimates of pass-through rates in Atkin and Donaldson (2015), our estimates are lower for further away source markets. The restriction that the pass-through rate needs to be positive is never violated. Also, only with two exceptions pass-through rates are below one. The average pass-through rate is 0.544, and, hence, given the incomplete pass-through there is clear evidence of imperfect competition. Atkin and Donaldson (2015) indicate that, if anything, estimates of the pass-through rate are likely to be biased upward, when shocks at source are correlated with shocks at destination. Especially in the case of agricultural commodities, and more so than in the case of imported goods, common shocks are likely (e.g. timing of harvest, weather shocks). Bias arising from common shocks is mitigated by including monthly seasonality (χ_{jm}) and annual seasonality (*season*$_t$). Remaining upward bias further strengthens the claim that the estimated pass-through rate is evidence of imperfect competition.

Next, we transform variables with estimated pass-through rates ($\widehat{\rho_{jk}}$) and estimate equation (10). Equations are estimated with OLS and include market pair and time fixed effects, and source and destination specific seasonality. Following standard practise (see Bertrand et al., 2004) standard errors in the estimation are clustered at the level of the intervention, in our case this concerns four groups: intervention and non-intervention pairs, both before and after the introduction of the bridge. Than we make a few sample adjustments: we drop observations that are more than 1800km apart (road distance): this avoids estimation results driven by trade pairs that are extremely far apart, and thereby less relevant[14]. Two additional sample adjustments are made: we focus on the marketing season, the months following harvest, since most trade takes place during these months[15]. In particular we estimate from April onwards with the first eight months (Table 1, column 2 and 4) and the first four months (Table 1, column 3). Next, we further restrict the sample period before the introduction of the new and rehabilitated bridges to the period that is fully "without bridge" facilities (Table 1, column 4). This restriction aims to realise a clean identification strategy (see Section 1).

Table 1 Impact of distance and road quality: basic estimation

Dependent variable: $\ln[(p_j - \widehat{\rho_{jk}} p_k)/\widehat{\rho_{jk}}]$

variables	Sample (1) within 1800km road distance[a]	(2), as (1) excluding Dec to Mar[b]	(3), as (1) excluding Aug to Mar[b]	(4), as (2) only 34 months before[c]
ln(road distance)	0.180** (0.050)	0.222** (0.059)	0.243** (0.061)	0.323** (0.087)
ln(road quality)	-0.401 (0.195)	-0.629** (0.193)	-1.122** (0.270)	-0.706* (0.269)
adj R^2	0.831	0.828	0.842	0.824
observations	11008	7392	3630	6374

All equations are estimated with OLS and include pass-through corrected trends, market pair and time fixed effects, and source and destination specific seasonality. Robust standard errors, clustered by group (before/after and (non)intervention), are in brackets next to the coefficient. *$p<0.10$, **$p<0.05$, ***$p<0.01$.
[a] Restricted to road distances that are common in long distance trade, less than 1800 km.
[b] Restricted to marketing months April to November (2), and restricted to marketing months April to July (3) .
[c] The period "without bridges" is from October 2006, the closing of the Mutarara bridge for rehabilitation works, to July 2009, 34 months before August 1, 2009, the opening of the Caia bridge (which also coincided with the completion of the rehabilitation of the Mutarara bridge).

Selected estimation output of the basic specification (equation 10) – reported in Table 1 – shows several interesting outcomes. In the first place the estimations confirm the importance of the coefficients of road distance and road quality: these coefficients are both statistically significant (or close to significant) and have the expected signs in all samples. The estimated coefficients are relatively stable with substantially different samples. The trend terms interacted

[14] A meaningful cut-off distance is found by searching for consistency of outcomes with different estimation models (notably with distance and quality as impact variables, with a binary impact variable, and in the parallel trend test).
[15] Alternatively, during the lean season trade is minimal, markets are thin, and prices fluctuate wildly and unsystematically. Under these circumstances spatial price differences tend to be less informative.

with the pass-through rate (not shown) are also highly significant in all estimations. The logarithmic transformation allows an interpretation in terms of elasticities: a 10% reduction in road distance leads to a 2.2-3.2% reduction in transport costs, and a 10% improvement in road quality leads to a 6.3-11.2% reduction in transport costs. It should be noted that changes in road distance are – in our case – large while changes in road quality are usually modest. It is tempting to compare the size of the coefficients, for both road distance and road quality, with the size of the coefficients if we do not take account of spatially varying mark-ups: although a direct comparison is difficult, previous work suggests that taking account of spatially varying mark-ups yields substantially lower road distance and road quality elasticities[16].

Estimations with covariates

Following the specification formulated in equation (10), we next include ***X*** variables in order to investigate if the previous estimation results are robust for the inclusion of covariates. We have considered covariates associated with domestic trade and (nominal) trade costs, reflecting developments on the demand side and on the supply side, and associated with imports or exports of food. We briefly discuss mechanisms underlying potential influences of covariates. Fuel prices and wages of truck drivers are the major nominal costs of transport, and both are, hence, expected to influence transport costs positively. Developments on the demand side are approximated with the size of the population in markets. We expect a larger population in destination market, ceteris paribus, to increase prices in destination markets and thereby to increase price differences between source and destination markets. For population in source markets we expect the reverse. The sum of the population of a market pair affects the size of

[16] A direct comparison is difficult because the transformation on the variables change the estimation samples in a way that obstructs exact matching. Estimates in previous versions of this paper, where we did not take account of spatially varying mark-ups (available from the author on request), are in the range of +0.86 to +0.92 for road distance, and -1.3 and -2.3 for road quality.

the bilateral trade flows: higher total numbers will intensify trade and trade opportunities, and thereby decrease transport costs.

Developments on the supply side are less easy and straightforward to incorporate: we employ rainfall (by district) of source markets in the previous production season, and per capita production by province of source markets, also of the previous production season. Both variables have their drawbacks: rainfall by district may be a reasonable predictor of rain-fed maize production but is less suitable to indicate maize available for trade to spatially dispersed markets. Per capita production more adequately reflects the availability for trade, but maize production data are only available by province and for a limited set of years[17]. We expect higher per capita maize production in source markets, ceteris paribus, to decrease prices in source markets and thereby, ceteris paribus, to increase price differences between source and destination markets.

We have further experimented with foreign prices, since prices in neighbouring countries, but especially in South Africa and Malawi, are likely to affect supply and demand in Mozambique. Note that South Africa and Malawi play a different role: South Africa is a major source of maize grain imports, while Malawi is a major destination of maize grain exports (see Zavale, 2014)). The influence of foreign prices is therefore ambiguous: it depends on whether maize is imported from foreign markets and foreign prices are lower (South-Africa), or maize is exported to foreign markets and foreign prices are higher (Malawi). South African maize prices are mostly well below Maputo maize prices, even if import tariffs are accounted for (see Appendix, Figure A2)[18]. In the southern Mozambique terminal maize markets these foreign prices are likely to exert a downward pressure on prices and we expect a negative impact of

[17] For the period of our study (2004-2014) annual maize production data are available for the years 2002-2003, 2005-2008 and 2012. The remaining years are predictions on the basis of a simple model with province and year fixed effects, province trends and rainfall shocks (see Appendix, Table A1).
[18] Maize imports are subject to a 2.5% import tariff and a 17% Value Added Tax, which is not levied on domestic production (see Zavale, 2014). These import duties, however, do not offset the price difference with Maputo.

South-African maize prices on price differences between Mozambique markets. For Malawi maize prices the reverse applies (see Appendix, Figure A3): Malawi is an outlet for Mozambique surplus maize, and generally higher Malawi maize prices exert an upward pressure on Mozambique price differences. In the case of South African prices we have used quotations of the nearest SAFEX white maize contract as an indicator of South-African maize prices. Specifically we have calculated monthly averages of daily quotations and converted these to Mozambique meticais. For Malawi we assume that Lilongwe market prices, also converted to Mozambique meticais, are representative for Malawi[19]. A final issue concerns domestic transport of freight by sea: unfortunately we lack (sufficient variation in) the data to adequately control for possible influences along these lines[20].

Selected estimations including covariates are reported in Table 2. Note that estimations in Table 2 are without time fixed effect (since this would absorb variation in national variables, like national consumer price index, diesel prices and SAFEX white maize spot prices). Estimations confirm previous estimates of impact variables: coefficients of road distance and road quality are both statistically significant and have a similar size as estimations without covariates. The performance of covariates is mixed: we find consumer prices and Lilongwe maize prices to be significant and with the expected signs. Trade pair population consistently has the expected sign, but is only occasionally statistically significant. Diesel and gasoline

[19] See Appendix, Figure A2 and A3 for the development of domestic maize prices vis-à-vis SAFEX white maize spot and Lilongwe average maize prices. For Mozambique exports one needs to compare post-harvest prices (April-July) with foreign prices (Malawi), while for Mozambique imports one needs to compare lean season prices (January-March) with foreign prices (South Africa).

[20] Trade between different destination markets is not likely. However, there is an exception: most destination markets are located along the coast (Pemba, Nacala, Quelimane, Beira, Massinga, Maxixe, Xai-Xai and Maputo) and this offers a low-cost alternative to transport freight by sea. Especially markets well connected with inland surplus areas have opportunities for profitable trading transactions with deficit destination market on the coast in the south, Maputo in the first place. Distance of inland markets to the nearest seaport approximates the potential impact of access to other markets, either domestic or foreign. Unfortunately, we do not have variables expressing influences through these channels and as a result potential impacts need to be captured by market pair fixed effects.

prices, and SAFEX prices white maize spot prices do not generate consistent results, most likely because of interactions, mutually and with the consumer price index.

Table 2 Impact of distance and road quality: including covariates

Dependent variable: $\ln[(p_j - \widehat{\rho_{jk}} p_k)/\widehat{\rho_{jk}}]$

	sample							
variables	(1) within 1800km road distance[a]		(2), as (1) excluding Dec to Mar[b]		(3), as (1) excluding Aug to Mar[b]		(4), as (2) only 34 months before[c]	
ln(road distance)	0.198**	(0.050)	0.227***	(0.034)	0.222**	(0.068)	0.273**	(0.068)
ln(road quality)	-0.430	(0.188)	-0.678**	(0.201)	-1.099**	(0.267)	-0.707*	(0.262)
ln(cpi)	1.600***	(0.254)	1.379**	(0.392)	1.143**	(0.287)	2.278***	(0.207)
ln(diesel price)	0.432**	(0.093)	-0.087	(0.444)	0.214	(0.476)	0.006	(0.198)
ln(gasoline price)	-0.537**	(0.100)	-0.150	(0.285)	-0.999*	(0.3751)	0.352	(0.252)
ln(population pair)	-1.071	(0.465)	-0.769	(0.482)	-0.153	(0.610)	-1.684**	(0.496)
ln(SAFEX white maize)	-0.021	(0.023)	0.077***	(0.011)	0.276***	(0.049)	0.063**	(0.017)
ln(Lilongwe price)	0.380**	(0.115)	0.358**	(0.086)	0.488***	(0.078)	0.288***	(0.044)
adj R^2	0.816		0.818		0.834		0.816	
observations	11008		7392		3630		6374	

All equations are estimated with OLS and include pass-through corrected trends, market pair fixed effects, and source and destination specific seasonality. Robust standard errors, clustered by group (before/after and (non)intervention), are in brackets next to the coefficient. *p<0.10, **p<0.05, ***p<0.01.

[a,b, c] See Table 1

Verification of impact with observed transport costs data

The ultimate test of the estimations reported in the previous tables, and, indeed, the ultimate test of the Atkin and Donaldson methodological framework, is to estimate a similar relationship with observed transport costs, rather than (adjusted) price differences. If the applied framework is an adequate technique to control for spatially varying mark-ups, and thereby a justification to use (adjusted) spatial price differences for the estimation of transport costs, estimation with observed transport costs should generate similar results. The obvious problem is that observed transport costs data generally are not, or only fragmentary, available and the Mozambique case is no exception.

However, as indicated in the documentation on the data, we do have some observed transport costs data from SIMA, for a limited period (August 2001 to December 2010, most of these data from before 2005) and for a limited number of itineraries. There are no observations of transport costs after July 2009, of itineraries that cross the new bridge. In order to verify the estimations based on spatial price differences, we estimate the following differences-in-differences equation, a simplified version of equation (10):

$$tc_{jkt} = \eta_0 + \eta_1\ roaddistance_{jkt} + \eta_2\ roadquality_{jkt} + \psi_t + \varepsilon_{jkt} \tag{14}$$

In this equation all variables with pass-through rate transformations are omitted. Additionally, since we do not have post-intervention transport cost observations, there is no variation of road distance and road quality over the years and by implication impacts are only identified by cross-sectional variation. Therefore, we need to drop market pair fixed effects. Moreover, we cannot cluster standard errors, as done in the case of the price differences estimation: hence, we cluster errors by market pair. Finally, market specific seasonality is omitted to preserve statistical power: this seasonality is absorbed by time fixed effects (and, hence, we also drop φ_{jm}).

Table 3 Verifying impact of distance and road quality with transport cost data

Dependent variable: $\ln(tc_{jk})$

	sample (see also note to Table)			
Variables	(1) all observations	(2), as (1) within 1800km road distance	(3), as (2) excluding Dec to Mar	(4), as (2) excluding Aug to Mar
ln(road distance)	0.474^{***} (0.062)	0.444^{***} (0.077)	0.456^{***} (0.072)	0.448^{***} (0.096)
ln(road quality)	-0.693^{**} (0.311)	-0.663^{**} (0.309)	-0.726^{**} (0.306)	-0.540^{*} (0.315)
adj R^2	0.571	0.536	0.561	0.531
Observations	634	596	419	224

Note to table: The selection of source and destination markets in the transport cost data correspond to the selection explained in detail in Section 2. All equations are estimated with OLS and include time fixed effects. Robust standard errors are clustered by market pair $^{*}p<0.10$, $^{**}p<0.05$, $^{***}p<0.01$.

Selected estimation results, reported in Table 3, show impacts of road distance and road quality that are statistically significant and have the expected sign. Estimation results are also robust to large variations in samples. On the whole, the size of the coefficients is slightly larger (in absolute terms) relative to the adjusted price gap estimations in Table 1 and 2, but not far off-the-mark: at standard levels of confidence coefficients of Table 3, and those of Table 1 and 2 are statistically in the same range. Difference may be due to differences in the samples (see

Appendix, Figure A6). Restricting road distance further brings the coefficients still closer to the coefficients in Table 1 and 2. We conclude that the available transport cost data offer reasonable support for the applied framework to use (adjusted) spatial price differences for the estimation of transport costs.

5. Potential threats and robustness checks

Estimating Differences-in-differences with a binary impact variable

Estimation results using a specification with a binary impact variable are reported in Table 4a and 4b. Table 4a repeats the estimations of Table 1, and Table 4b repeats estimations of Table 2, with the only difference that a binary impact variable is substituted for the two impact variables road distance and road quality. Note again that estimations in Table 4b, similar to estimations in Table 2, are without time fixed effect. Coefficients of the impact variable, the bridge dummy, are mostly statistically significant and vary in size from a reduction in transport costs of 6% to 10%. With month-year dummies (Table 4a) the fit of the estimation is slightly better: apparently there are time fixed effects that are not captured by the price variables. Again, coefficients of consumer price index and Lilongwe maize prices are statistically significant, have expected signs and relatively stable coefficients. Other price variables perform less. Including covariates improves accuracy of impact variables and pass-through adjusted trend.

Table 4a Impact of distance and road quality: binary impact variable, basic

Dependent variable: $\ln[(p_j - \widehat{\rho_{jk}} p_k)/\widehat{\rho_{jk}}]$				
	Sample			
Variables	(1) within 1800km road distance[a]	(2), as (1) excluding Dec to Mar[b]	(3), as (1) excluding Aug to Mar[b]	(4), as (2) only 34 months before[c]
bridge (binary)	-0.029 (0.016)	-0.049* (0.018)	-0.058** (0.018)	-0.072* (0.027)
adj R^2	0.831	0.827	0.841	0.824
Observations	11008	7392	3630	6374

All equations are estimated with OLS and include pass-through corrected trends, market pair and time fixed effects, and source and destination specific seasonality. Robust standard errors, clustered by group (before/after and (non)intervention), are in brackets next to the coefficient. *p<0.10, **p<0.05, ***p<0.01.

[a,b] See Table 1

Table 4b Impact of distance and road quality: binary impact variable, covariates

Dependent variable: $\ln[(p_j - \widehat{\rho_{jk}} p_k)/\widehat{\rho_{jk}}]$

	Sample			
variables	(1) within 1800km road distance[a]	(2), as (1) excluding Dec to Mar[b]	(3), as (1) excluding Aug to Mar[b]	(4), as (2) only 34 months before[c]
bridge(binary)	-0.039* (0.017)	-0.054*** (0.009)	-0.050** (0.016)	-0.060** (0.018)
ln(cpi)	1.612*** (0.265)	1.392** (0.404)	1.159** (0.305)	2.305*** (0.216)
ln(diesel price)	0.453** (0.104)	-0.063 (0.457)	0.249 (0.497)	0.041 (0.203)
ln(gasoline price)	-0.542** (0.099)	-0.159 (0.290)	-1.002* (0.370)	0.341 (0.250)
ln(population pair)	-1.084 (0.468)	-0.781 (0.492)	-0.170 (0.403)	-1.707** (0.500)
ln(SAFEX white maize)	-0.024 (0.025)	0.075 (0.014)	0.271** (0.053)	0.059** (0.014)
ln(Lilongwe price)	0.380** (0.115)	0.359** (0.087)	0.490*** (0.078)	0.288*** (0.045)
adj R^2	0.816	0.817	0.834	0.816
Observations	11008	7392	3630	6374

All equations are estimated with OLS and include pass-through corrected trends, market pair fixed effects, and source and destination specific seasonality. Robust standard errors, clustered by group (before/after and (non)intervention), are in brackets next to the coefficient. *$p<0.10$, **$p<0.05$, ***$p<0.01$.

[a,b] See Table 1

Propensity Score Matching estimation

Since the placement of the bridge is non-random, and, hence, since the selection of market pairs that realize a reduction in road distance and an improvement in road quality is also non-random, a differences-in-differences estimation comparing this selection with all other market pairs is potentially suffering from selection bias in time-varying observables. As a result estimated causal impacts of the bridge on transport costs may be biased. In the spatial economics literature the major strategy to address this is to develop plausible instruments that meet the requirements[21]. In empirical studies this has resulted in the so-called planned route IV, the historical route IV and the inconsequential place approach (Redding and Turner, 2014). Since we do not have data on planned or historical routes and inconsequential places are not common in the Mozambique trunk road network, these IV strategies cannot be implemented. Instead, we have implemented propensity score matching (PSM).

[21] Instruments have to satisfy the exclusion restriction, meaning that the excluded exogenous variables – the instrument – is correlated with the change in infrastructure, but only affects transport costs through this channel.

The matching strategy builds on the Conditional Independence Assumption (CIA) meaning that the outcome in both intervention and control group, is independent of treatment assignment given the propensity score. In order to implement Propensity Score Matching estimations, we first estimate the propensity score i.e. the probability of treatment, in the current analysis the probability to make use of the Zambezi bridge. For this purpose we employ a logit model[22]. We experiment with a range of variables, including the covariates of the previous estimation, for the estimation of the propensity score. In the selected estimation[23] the propensity score is assumed to be determined by a trend, whether source and destination market are on different side of the Zambezi interacted with trend, population size in source and destination markets, the sum of population at source and destination market, and the spatial difference in (seasonal) rainfall. We assume that the selected variables meet the requirements since they both influence assignment into intervention or control group, while assignment into intervention does not affect these variables. Also, the larger part of the influence of the selected variables on outcome is likely to run through the treatment variable. Results of the propensity score estimation are reported in the Appendix, Table A5. Coefficients of the covariates in the propensity score estimation have expected signs. The pseudo R2 indicates how well variables explain the probability to make use of the Zambezi bridge and is thereby a formal test of the model. These statistics are comfortably high (R^2=0.72, see Appendix, Table A5).

In order to match treatment and control observations, we use Kernel Matching as a matching algorithm. This is motivated by the availability of a large number of control observations. Kernel Matching is a non-parametric estimator that uses a weighted average of all control group observations to construct the counterfactual outcome. Weights depend on the

[22] Probit or logit are likely to give similar outcomes. However, the logit distribution has more density mass in the bounds and this corresponds with our empirical setting (see also Caliendo et al., 2005)

[23] In selecting variables for the propensity score estimation, we aimed at maintaining the maximum number of observations (to improve power) and focusing on spatial and climate variables to guarantee exogeneity.

distance between each observation from the control group and the treatment observation for which the counterfactual is estimated. Higher weights are placed on observations close in terms of propensity score and vv. As more information is used compared to, for example, Nearest Neighbour matching, Kernel Matching results in a lower variance, and, thus, higher precision estimates. Kernel Matching is also time consuming since for each treatment observation an appropriate set of weighted controls is constructed. The Kernel function is the Epanechnikov kernel. Following accepted practise we use a bandwidth of 0.06[24].

We have tested the robustness of the matching algorithm by also implementing Nearest Neighbour (NN) as a matching algorithm, using 3-10 of the nearest controls, with replacement, combined with a caliper threshold, where the caliper takes values 0.002-0.010. Replacement is justified because the distribution of the propensity score is different in the treatment and control group: not allowing replacement may include using observations with very different propensity scores, and make estimates less accurate. The diverging distributions of the propensity score are also apparent from the common support figures shown in the Appendix, Figure A5. Restricting matches to those within the caliper threshold – a maximum distance of the propensity score of treatments and matched control observations – further decreases the possibility of bad matches and hence bias. Unfortunately, the literature does not give a clue which values for the tolerance level are appropriate. Finally, ordering is done randomly since estimations with NN matching are dependent on the ordering of the data.

The overlap and region of common support between treatment and comparison group is shown graphically in the Appendix (see Figure A5). The cut-off is the straightforward and standard "minima and maxima criterion": treatment (control) observations whose propensity score is higher than the maximum or less than the minimum propensity score of the controls

[24] This bandwidth value is the default value in the STATA routine psmatch2 (E. Leuven and B. Sianesi, 2003, 'PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing'.)

(treatments) are dropped. Visual inspection of the figures confirm that the range of values of the matched propensity score have both treatment and control observations with probabilities between 0 and 1. Hence, the overlap condition is satisfied. In order to assess the quality of the matching procedure we use the standardised bias, before and after matching, as suggested by Rosenbaum and Rubin (1985)[25]. The results of this exercise, reported in the Appendix, Table A7, indicate that matching on the estimated propensity score balances the covariates in the matched samples reasonably well.

Table 5 Impact of bridges: Propensity Score Matching (PSM), Kernel Matching

Dependent variable: $\ln[(p_j - \widehat{\rho_{jk}} p_k)/\widehat{\rho_{jk}}]$

		sample			
variables		(1) within 2000km road distance[a]	(2), as (1) excluding Dec to Mar[b]	(3), as (1) excluding Oct to Mar[b]	(4), as (2) only 34 months before[c]
est. technique		PSM / KM	PSM / KM	PSM / KM	PSM / KM
bridge (ATT)		-0.095** (0.043)	-0.131** (0.051)	-0.132** (0.063)	-0.122** (0.051)
bridge (ATU)		-0.108	-0.127	-0.116	-0.125
bridge (ATE)		-0.101	-0.129	-0.125	-0.123
on support:	treated	1216	835	501	832
	untreated	932	663	408	669
off support	treated	311	245	166	248
	untreated	7053	4702	2953	3586
observations		9512	6445	3210	5335

Equations are estimated with propensity score matching (PSM), using kernel matching as matching algorithm. Estimates of the propensity scores are in the Appendix, Table A5. Matching algorithm: Kernel Matching, Epanechnikov kernel and bandwidth 0.06. Standard errors are in brackets next to the coefficient. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.
[a,b, c] See Table 1.

Selected estimation results of the PSM (KM) estimations are reported in Table 5. The PSM results generate a statistically significant average treatment effect on the treated (ATT) which ranges from -0.09 to -0.13. Estimations with NN matching generate similar results (see Appendix, Table A6) as with Kernel Matching and give confidence about the robustness of the

[25] $B = \frac{(\bar{X}_1 - \bar{X}_0)}{\sqrt{(V_1(X) + V_0(X))/2}}$ where $\bar{X}_1$ ($\bar{X}_0$) and $V_1(X)$ ($V_0(X)$) are, respectively, the average and variance of covariate X in the treatment (control) group. The standardised bias, B, is calculated before and after matching, for each covariate X.

matching procedure. PSM results further indicate an average treatment (ATE) that is reasonably in line with estimates based on the differences-in-differences specification (see Table 5 vis-à-vis Table 4a and 4b). We conclude that the PSM results further confirm the impact of bridges on transport costs. In the following section we assess if the estimated road distance and road quality elasticities, estimated with the differences-in-differences specification translate into transport cost reductions for specific itineraries, and how these cost reductions can be decomposed to road distance and road quality.

6. Discussion of outcomes and conclusion

The estimation outcome achieved so far offers insight into the impact of bridges, averaged over itineraries, and how the reduction in transport costs can be attributed to the key determinants of transport costs. These averages are interesting and useful[26] but less informative about realised cost reductions of particular itineraries that benefit from the new bridge. However, with the estimated elasticities for road distance and road quality, we are now in the position to also measure benefits for specific itineraries.

For a selection of itineraries this is shown in Table 6. Total reduction in transport costs due to the bridges range from 6% to 21%. In most instances the cost reduction is for the larger part due to the shorter distance: with a few exceptions (in particular itineraries to Nacala) change in quality mostly contributes only modestly to transport cost reduction. Overall, roughly two-third of the cost reduction is on account of road distance and one-third on account of road quality.

[26] To measure the benefit of the bridge for the Mozambique maize market, we ideally need to weigh reductions in transport cost per itinerary with the size of freight on transported through these routes. Unfortunately such trade flow data are not available.

Table 6 Reduction in trade costs by itinerary due to bridge

from north to south		road distance	trnsport cost	road quality	trnsport cost			
From	**to**	**%Δ**	**%Δ**	**%Δ**	**%Δ**	**total**	**road distance**	**road quality**
AltoMolocue	Maputo	-16.2%	-4.9%	1.4%	-1.1%	-5.9%	82.2%	17.8%
AltoMolocue	Beira	-40.9%	-12.3%	-0.6%	0.5%	-11.8%		
Mocuba	Maputo	-17.7%	-5.3%	2.4%	-1.8%	-7.1%	74.7%	25.3%
Mocuba	Beira	-48.1%	-14.4%	-4.4%	3.3%	-11.1%		
Ribaue	Maputo	-15.4%	-4.6%	2.5%	-1.9%	-6.5%	71.1%	28.9%
Ribaue	Beira	-37.4%	-11.2%	-0.1%	0.1%	-11.1%		
Nampula	Maputo	-14.8%	-4.4%	3.2%	-2.4%	-6.8%	64.9%	35.1%
Nampula	Beira	-35.2%	-10.6%	1.0%	-0.8%	-11.3%	93.4%	6.6%
from south to north								
Chimoio	Nampula	-18.5%	-5.6%	3.4%	-2.6%	-8.1%	68.5%	31.5%
Chimoio	Nacala	-17.1%	-5.1%	12.8%	-9.6%	-14.7%	34.8%	65.2%
Gorongosa	Nampula	-29.7%	-8.9%	8.5%	-6.4%	-15.3%	58.3%	41.7%
Gorongosa	Nacala	-26.8%	-8.0%	17.5%	-13.1%	-21.2%	38.0%	62.0%
Manica	Nampula	-14.5%	-4.4%	3.2%	-2.4%	-6.8%	64.4%	35.6%
Manica	Nacala	-7.1%	-2.1%	12.5%	-9.4%	-11.5%	18.5%	81.5%

Source: authors' calculations (elasticity of transport costs for road distance: +0.30; elasticity of transport costs for road quality: -0.75)

In this study we have investigated the impact of bridges in Mozambique on transport costs. The applied methodology allows for potentially oligopolistic traders with spatially varying mark-ups. For the identification we exploited the introduction of the bridge between Caia and Chimuara. This event generated the required variation in trading distances between markets, needed to attribute impact to road distance and road quality. The key finding is that, averaged over itineraries, the bridge has caused a 6% to 10% reduction in transport costs. For specific itineraries this reduction in costs is as large as 21%. Roughly two-third of the cost reduction is due to the reduction in road distance. Results are robust for inclusion of covariates and for controlling for non-random assignment of bridge placement. The applied methodology estimated on the basis of (adjusted) spatial price differences is supported by similar estimations on the basis of observed transport costs.

References

Abdula, D.C., 2005, 'Improving Maize Marketing And Trade Policies To Promote Household Food Security In Southern Mozambique', thesis submitted for the degree of Master of science, Department of Agricultural Economics, Michigan State University.

Atkin, D. and D. Donaldson, 2015, 'Who's Getting Globalized? The Size and Implications of Intra-national Trade Costs', National Bureau of Economic Research, w21439.

Autor, 2003, 'Outsourcing at Will: The Contribution of Unjust Dismissal Doctrine to the Growth of Employment Outsourcing', *Journal of Labor Economics*, 2003, 21, 1.

Azam, J.P., 2007, *Trade, Exchange Rate and Growth in Sub-Saharan Africa,* Cambridge Cambridge University Press, UK.

Banerjee, A., E. Duflo and N.Qian, 2012, 'On the Road: Access to Transportation Infrastructure and Economic Growth in China', working paper.

Baulch, B. 1997a, 'Transfer Costs, Spatial Arbitrage, and Testing for Food Market Integration', *American Journal of Agricultural Economics*, 79, 2, 477–487.

Baulch, B. 1997b, 'Testing for Food Market Integration Revisited', *The Journal of Development Studies*, 33, 4, 512–34.

Bertrand, M., E. Duflo and S. Mullainathan, 2004, 'How much should we trust difference-in-differences estimates?', *Quarterly Journal Of Economics,* 119, 1, 249-275.

Broda, C.M. and D.E. Weinstein, 2008, 'Understanding International Price Differences Using Barcode Data', NBER Working Paper 14017.

Brooks, W. and K. Donovan, 2017, 'Eliminating Uncertainty in Market Access: The Impact of New Bridges in Rural Nicaragua', working paper.

Casaburi, L., Glennerster R. and T. Suri, 2013, 'Rural Roads and Intermediated Trade: Regression Discontinuity Evidence from Sierra Leone', mimeo.

Cirera, X. and C. Arndt, 2008, 'Measuring the impact of road rehabilitation on spatial market efficiency in maize markets in Mozambique', *Agricultural Economics*, 39, 17-28.

De Janvry, A. and E. Sadoulet, 2006, 'Progress in the Modelling of Rural Households' Behaviour under Market Failures', Chapter 8 in De Janvry, A. and R.Kanbur, 2006.

Dominguez-Torres, C. and C.Briceño-Garmendia, 2011, 'Mozambique's Infrastructure: A Continental Perspective', AICD report, World Bank, Washington.

Donaldson, D., 2010, 'Railroads of the Raj: Estimating the Impact of Transportation Infrastructure', Asia Research Centre, Working Paper 41 (also forthcoming in *American Economic Review*).

Eaton, J. and S. Kortum, 2002, 'Technology, Geography and Trade', *Econometrica*, 70, 5, 1741-1780.

Fackler, P.L. and B.K. Goodwin, 2001, 'Spatial Price Analysis', *Handbook of Agricultural Economics*, 1, 971-1024.

Fafchamps, M. and B. Minten, 2012, 'Impact of SMS-Based Agricultural Information on Indian Farmers', *World Bank Economic Review*, 27, February, 1-32.

Gopinath, G., P.-O. Gourinchas, C-T. Hsieh, N. Li, 2011, 'International prices, costs and mark-up differences', *American Economic Review*, 101, 6, 2450–2486.

Jacoby, H.G., 2000, 'Access to Markets and the Benefits of Rural Roads', *The Economic Journal*, 110, 713-737.

Jacoby, H.G. and B. Minten, 2009, 'On measuring the Benefits of Lower Transport Costs', *Journal of Development Economics*, 89, 1 , 28-38.

Jensen, R., 2007, 'The Digital Provide: Information (Technology), Market Performance, and Welfare in the South Indian Fisheries Sector', *Quarterly Journal of Economics*, 72, 3, 879-924.

Kaminski, J., L. Christiaensen and C.L.Gilbert, 2016, 'Seasonality in Local Food Prices and Consumption: Evidence from Tanzania', *Oxford Economic Papers*, 68, 3, 736-757.

Kerem Coşar, A., P.L.E. Grieco and F. Tintelnot, 2015, ‘Bias in estimating border- and distance-related trade costs: Insights from an oligopoly model’, Economics Letters, 126, 147-149.

Key, N., E.Sadoulet and A. de Janvry, 2000, ‘Transactions Costs and Agricultural Household Supply Response’, *American Journal of Agricultural Economics*, 82, May, 245-259.

Minten, B. and S. Kyle, 1999, ‘The effect of Distance and Road Quality on Food Collection, Marketing Margins, and Traders’ Wages: Evidence from the Former Zaire’, *Journal of Development Economics*, 60, 467-495.

Omamo, S.W., 1998, ‘Transport Costs and Smallholder Cropping Choices: An Application to Siaya District, Kenya’, *American Journal of Agricultural Economics*, 80, 116-123.

Redding, S.J. and M.A Turner, 2014, ‘Transportation Costs and the Spatial Organization of Economic Activity’, Working Paper 20235, National Bureau of Economic Research, Cambridge, United States.

Renkow, M., D.G. Hallstrom and D. D. Karanja, 2004, ‘Rural infrastructure, transactions *costs* and market participation in Kenya’, *Journal* of *Development Economics*, 73,1, 2004, 349-367.

Rosenbaum, P.R. and D. B. Rubin, 1985, ‘Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score’, *The American Statistician*, 39, 1, 33-38.

Tostão, E. and B.W. Brorsen, 2005, ‘Spatial price efficiency in Mozambiques’s post-reform maize markets’, *Agricultural Economics*, 33, 205-214.

Tschirley, D., D. Abdula and M.T.Weber, 2006, ‘Toward Improved Maize Marketing and Trade Policies to Promote Household Food Security in Central and Southern Mozambique’, Research Report 60E, Ministry of Agriculture, Mozambique.

De Vletter, F and E. Polana, 2001, ‘Female Itinerant Maize Traders in Southern

Mozambique: A Study of a Higher-End Informal Sector Activity and Its Potential for Poverty Reduction', ILO-SAMAT discussion paper 17.

Volpe Martincus, C., J. Carballo and A. Cusolito, 2017, 'Roads, exports and employment: Evidence from a developing country', *Journal of Development Economics,*125, 21-39.

Zant, W., 2013, 'How Is the Liberalization of Food Markets Progressing? Market Integration and Transaction Costs in Subsistence Economies', *World Bank Economic Review, 27, 1*, 28-54. DOI: 10.1093/wber/lhs017.

Zant, W., 2017, 'Impact of Mobile Phones on Staple Food Markets in Mozambique: Improved Arbitrage or Increased Rent Extraction?', Tinbergen Discussion Paper 17-021/V, Tinbergen Institute, Amsterdam.

Zant, W., 2018, 'Trains, Trade and Transaction Costs: How does Domestic Trade by Rail affect Market Prices of Malawi Agricultural Commodities?', *World Bank Economic Review*, 32, 2, 334-356.

Zavale, H., 2014. 'Analysis of price incentives and disincentives for maize in the Republic of Mozambique for the time period 2005-2013', MAFAP, FAO, Rome.

Appendix

Table A1 Maize production by province (x1000 tonnes)

Year	CabDel	Gaza	Inhamb	Manica	Maputo	Nampul	Niassa	Sofala	Tete	Zambez
2002	85651	66921	18455	162822	21769	117435	175233	76091	205199	185198
2003	93071	56453	16691	172190	7622	89112	159660	104119	183433	298928
2004	*93743*	*89445*	*20724*	*184375*	*20376*	*109505*	*165788*	*87553*	*213827*	*230428*
2005	80363	40818	18013	162180	10400	102544	121748	52651	173989	178811
2006	104987	102091	32456	204026	29265	124000	222590	102489	260331	213241
2007	85655	60941	29049	211935	10891	93911	103820	96837	211826	229045
2008	76120	63815	36890	187079	26556	99623	170402	105093	238901	209090
2009	*78292*	*55351*	*33686*	*207662*	*21358*	*106359*	*146640*	*102677*	*226168*	*199351*
2010	*75522*	*54044*	*34541*	*212661*	*33227*	*105741*	*143084*	*106001*	*228720*	*193658*
2011	*72850*	*52768*	*24699*	*217780*	*36049*	*105126*	*139615*	*109434*	*231301*	*188127*
2012	68410	48675	20625	227748	31570	112494	143761	118346	226912	178848

Source: Trabalho de Inquérito Agrícola (TIA) / Anuario de Estatistica Agararia, Ministry of Agriculture; numbers in italics are constructed using a simple specification with year and province fixed effects, province specific trends and seasonal rainfall. Some province names are abbreviated: CabDel=Capo Delgado; Inhamb=Inhambane; Nampul=Nampula; Zambez=Zambezia).

Table A2 Availability of transport cost data: by itinerary

itinerary	n	in %	by source	itinerary	n	in %	by source
Alto Molocue – Nampula	62	7.5%		Montepuez – Nacala	25	3.0%	
Alto Molocue-Quelimane	47	5.7%		Montepuez – Pemba	24	2.9%	
Alto Molocue – Maputo	18	2.2%		all from Montepuez			6.1%
Alto Molocue – Maxixe	16	1.9%		Nampula – Maxixe	21	2.5%	
Alto Molocue – Beira	10	1.2%		Nampula – Maputo	18	2.2%	
total from Alto Molocue			19.9%	Nampula – Xaixai	15	1.8%	
Angonia - Tete	48	5.8%		Nampula – Beira	9	1.1%	
total f rom Angonia			7.6%	total from Nampula			10.0%
Chimoio – Tete	21	2.5%		Nhamatanda – Maputo	53	6.4%	
Chinmoio – Xaixai	19	2.3%		Nhamatanda – Beira	16	1.9%	
Chimoio – Massinga	12	1.5%		Nhamatanda – Xaixai	15	1.8%	
Chimoio – Maputo	11	1.3%		total from Nhamatanda			12.1%
total from Chimoio			8.4%	Ribaue – Beira	12	1.5%	
Gorongosa – Beira	39	4.7%		total from Ribaue			2.5%
Gorongosa – Maputo	14	1.7%		Tete – Chimoio	21	2.5%	
total from Gorongosa			8.6%	Tete – Maputo	16	1.9%	
Lichinga – Maputo	12	1.5%		Tete – Maxixe	16	1.9%	
total from Lichinga			2.3%	Tete – Massinga	11	1.3%	
Manica – Beira	25	3.0%		total from Tete			8.5%
total from Manica			4.7%				
Mocuba – Quelimane	16	1.9%					
Mocuba – Nampula	13	1.6%					
total from Mocuba			3.9%				

Source: calculations based on SIMA data

Table A3 Availability of transport cost data: by year

year	n	in %	year	n	in %
2001	39	4.7%	2006	68	8.3%
2002	213	25.8%	2007	51	6.2%
2003	184	22.3%	2008	63	7.6%
2004	54	6.6%	2009	45	5.5%
2005	87	10.6%	2010	20	2.4%
all years	824				

Source: calculations based on SIMA data

Table A4 What are source and destination markets in maize trade?

markets	1	2	3	4	5	6
Pemba	**73.0**	*0%*	0.2%	*2.9%*	*139*	*yes*
Montepuez	**73.0**	**21.4%**	**5.6%**	0.0%	**76**	**no**
Lichinga	**186.5**	*0%*	**2.4%**	0.4%	*142*	**no**
Nacala	37.6	0.3%	0.0%	*4.4%*	*206*	*yes*
Angonia	**117.6**	**40.0%**	**8.2%**	0.0%	14	**no**
Cuamba	**186.5**	**29.9%**	0.4%	0.2%	79	**no**
Ribaue	37.6	**21.4%**	**2.1%**	0.2%	26	**no**
Nampula	37.6	*0%*	10.4%	10.8%	*472*	**no**
Alto Molocue	**68.7**	**5.5%**	**22.5%**	0.0%	42	**no**
Milange	**68.7**	*0%*	0.5%	0.0%	30	**no**
Tete	**117.6**	*0%*	7.7%	10.5%	156	**no**
Mocuba	**68.7**	**34.5%**	**3.4%**	0.5%	169	**no**
Mutarara	**68.7**	**35.3%**	0.5%	0.1%	9	**no**
Quelimane	**68.7**	*0%*	0.2%	0.0%	193	*yes*
Gorongosa	**175.8**	**36.7%**	**7.9%**	0.4%	19	**no**
Manica	**175.8**	**72.9%**	**3.8%**	0.0%	36	**no**
Chimoio	**175.8**	**84.4%**	**8.0%**	3.4%	237	**no**
Nhamatanda	53.7	*0%*	**12.4%**	0.1%	26	**no**
Beira	53.7	0.3%	0.6%	*13.5%*	432	yes
Massinga	45.7	**13.2%**	0.4%	*4.2%*	21	*yes*
Maxixe	45.7	0.3%	*0.0%*	*10.2%*	*109*	*yes*
Chokwe	47.0	**1.4%**	0.9%	0.7%	53	**no**
XaiXai	47.0	*0%*	0.7%	*8.4%*	*116*	*yes*
Maputo	15.3	*0%*	0.3%	*20.1%*	*1095*	*yes*

Note to Table: Column 1: per capita production in kg, 1999-2007, by province; 2: availability of weekly producer price data by market, 1999-2009 (source: SIMA); 3: source markets in weekly transport cost data by market pair, 2001-2010 (source: SIMA); 4: destination markets in weekly transport cost data by market pair, 2001-2010 (source: SIMA); 5: population size in 2007, x1000, by market (source: Instituto Nacional de Estatistica Moçambique); 6: located on the coastline. Markets are ordered from north to south. Characteristics that indicate source (destination) markets are printed in bold (italics).

Table A5 First stage logistic estimation of propensity score: probability of a route crossing the new Zambezi bridge

Dependent variable: probability of using bridge (d_bridge (binary))				
Variables	(1) within 2000km road distance[a]	(2), as (1) excluding Dec to Mar[b]	(3), as (1) excluding Oct to Mar[b]	(4), as (2) only 34 months before[c]
estimation technique	logit	logit	logit	logit
Variables				
trend	-0.300*** (0.110)	-0.763*** (0.259)	-1.920** (0.950)	-0.761*** (0.252)
across Zambezi x trend	0.136*** (0.009)	0.169*** (0.022)	0.275*** (0.080)	0.166*** (0.021)
ln(population of pair)	0.491*** (0.059)	0.489*** (0.070)	0.502*** (0.102)	0.487*** (0.069)
ln(pc production at source)	-2.322*** (0.118)	-2.315*** (0.140)	-2.345*** (0.207)	-2.293*** (0.139)
ln(pc production at dest.)	0.703*** (0.119)	0.679*** (0.139)	0.777*** (0.209)	0.670*** (0.139)
pseudo R2	0.717	0.715	0.728	0.693
Observations	9512	6445	3210	5335

Standard errors are in brackets next to the coefficient. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. [a,b] See Table 1.

Table A6 Impact of bridges: Propensity Score Matching (PSM), Nearest Neighbour

Dependent variable: $\ln[(p_j - \widehat{\rho_{jk}} p_k)/\widehat{\rho_{jk}}]$					
Variables		(1) within 2000km road distance[a]	(2), as (1) excluding Dec to Mar[b]	(3), as (1) excluding Sep to Apr[b]	(4), as (2) only 34 months before[c]
est. technique		PSM / NN	PSM / NN	PSM / NN	PSM / NN
bridge (ATT)		-0.093* (0.051)	-0.111** (0.055)	-0.133* (0.078)	-0.133** (0.055)
bridge (ATU)		-0.120	-0.186	-0.133	-0.175
bridge (ATE)		-0.105	-0.145	-0.133	-0.152
on support:	treated	815	682	214	688
	untreated	636	564	196	580
off support	treated	712	398	305	392
	untreated	7349	4801	2495	3675
observations		9512	6445	3210	5335

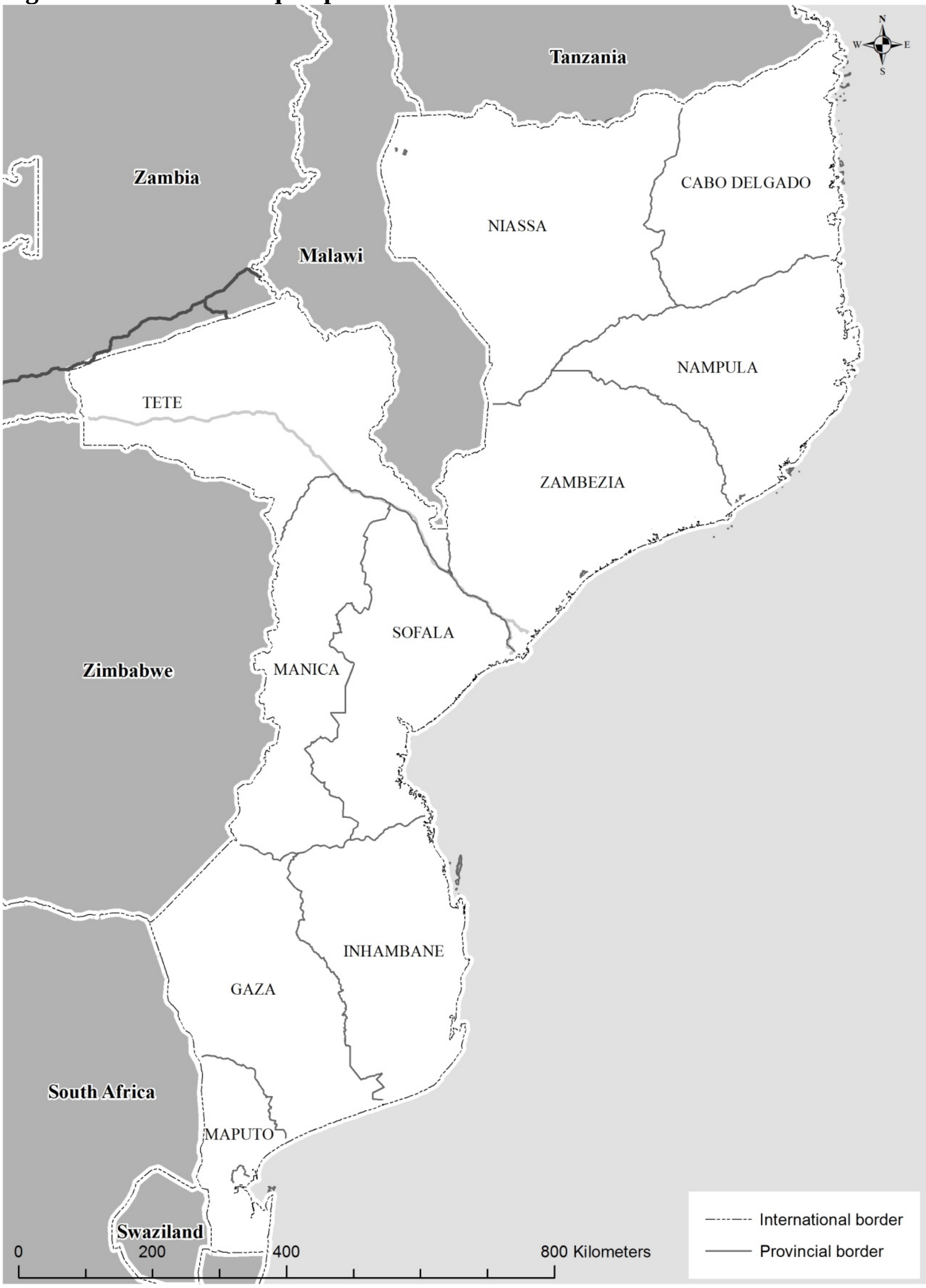
Equations are estimated with propensity score matching (PSM), using Nearest Neighbor (n=2-10), with replication, combined with Caliper threshold (0.001-0.01). Standard errors are in brackets next to the coefficient. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. [a,b] See Table 1.

Table A7 Standardized Bias of Covariates, before and after Matching

	(1) within 2000km road distance[a]		(2), as (1) excluding Dec to Mar[b]		(3), as (1) excluding Oct to Mar[b]		(4), as (2) only 34 months before[c]	
before and after matching	before	after	before	after	before	after	before	after
trend	1.166	1.169	1.153	1.123	0.917	1.123	0.909	1.067
across Zambezi x trend	3.242	1.177	3.253	1.130	3.001	1.131	3.018	1.073
ln(population of pair)	0.186	0.132	0.219	0.137	0.171	0.145	0.208	0.148
ln(pc production at source)	0.068	0.020	-0.005	0.040	0.061	-0.017	-0.016	-0.005
ln(pc production at dest.)	0.026	0.011	0.009	0.014	0.007	-0.002	-0.010	-0.004

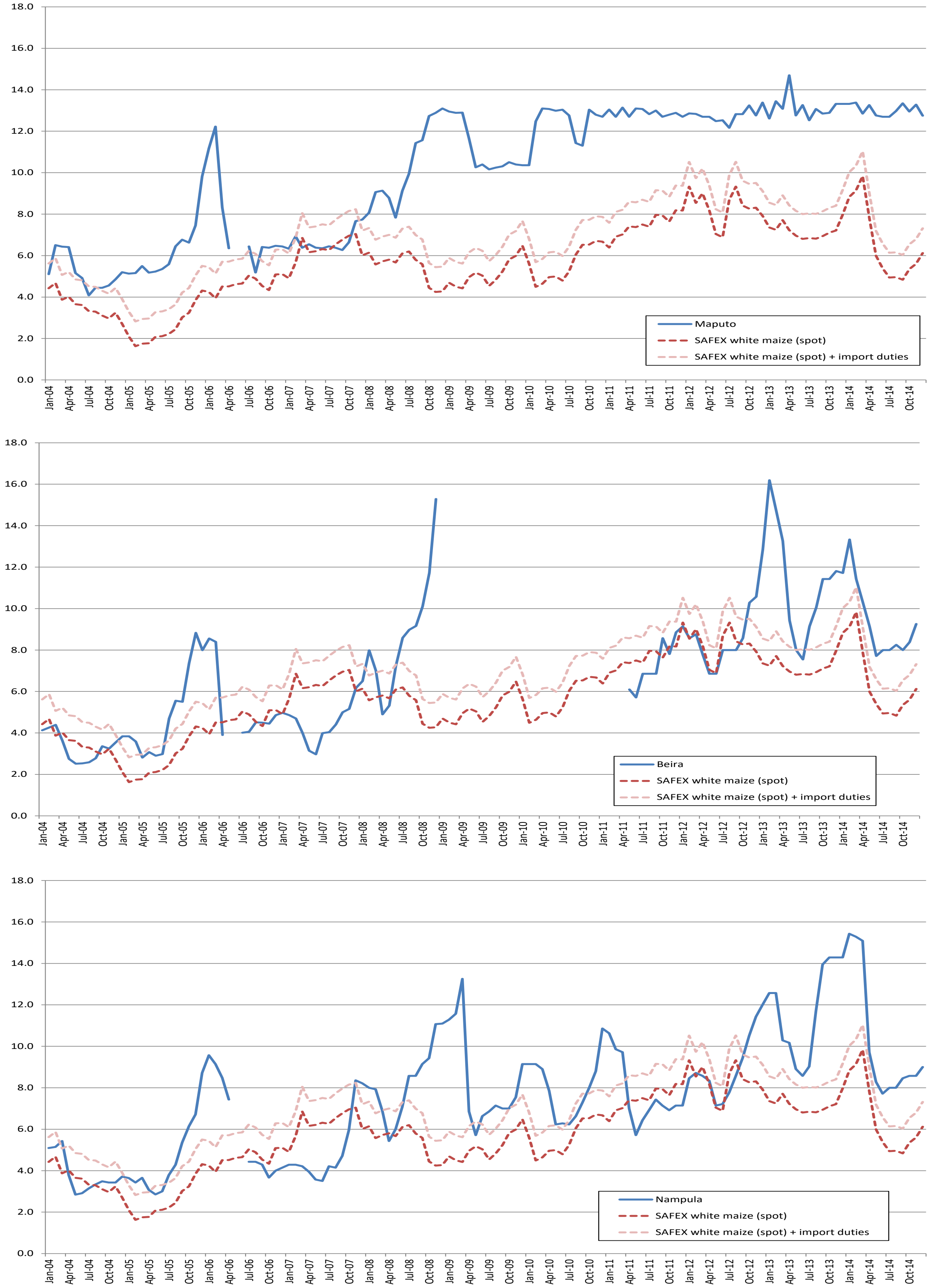
Note to Table: $B = \frac{(\bar{X}_1 - \bar{X}_0)}{\sqrt{(V_1(X) + V_0(X))/2}}$ where $\bar{X}_1$ ($\bar{X}_0$) and $V_1(X)$ ($V_0(X)$) are, respectively, the average and variance of covariate X in the treatment (control) group. The standardised bias, B, is calculated before and after matching, for each covariate X. The statistics in the table correspond with the Propensity Score Matching estimates with Kernel Matching reported in the main text.

Figure A1 Mozambique: provinces



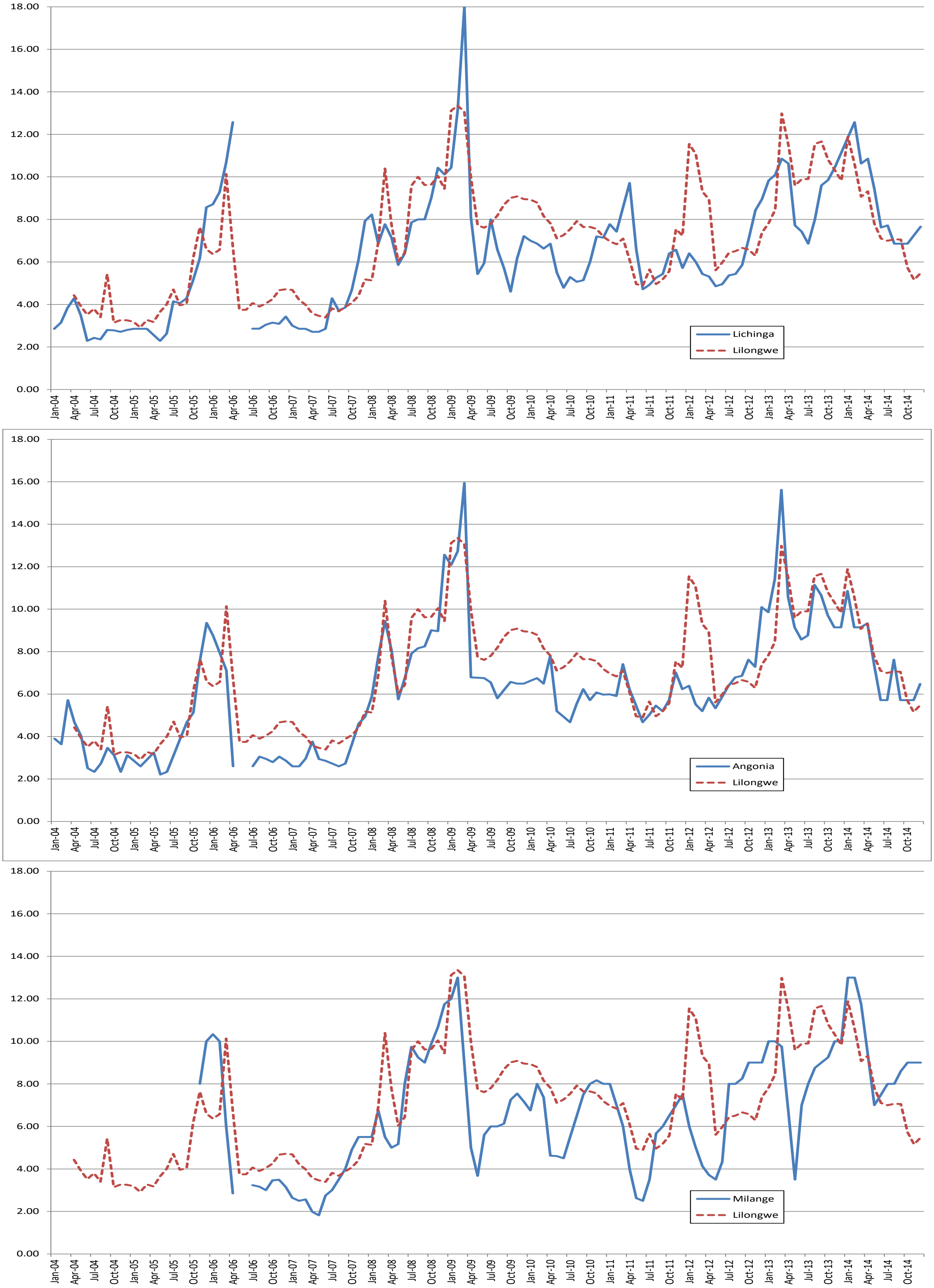
Source: VU SPINlab

Figure A2 Domestic maize prices vis-à-vis SAFEX white maize spot (meticais per kg)



Source: SIMA, IFS (IMF), SAFEX (SAFEX white maize (spot) are monthly averages of daily quotations of the nearest contract), all prices in Mozambique meticais per kg. The light dotted line includes 2.5% import tax and 17% VAT (Tschirley et al., 2006; Zavale, 2014).

Figure A3 Domestic maize prices vis-à-vis Lilongwe maize prices (meticais per kg)



Source: SIMA, IFS (IMF) and FAO, all prices in Mozambique meticais per kg.

Figure A4 Estimated pass-through rates and road distance

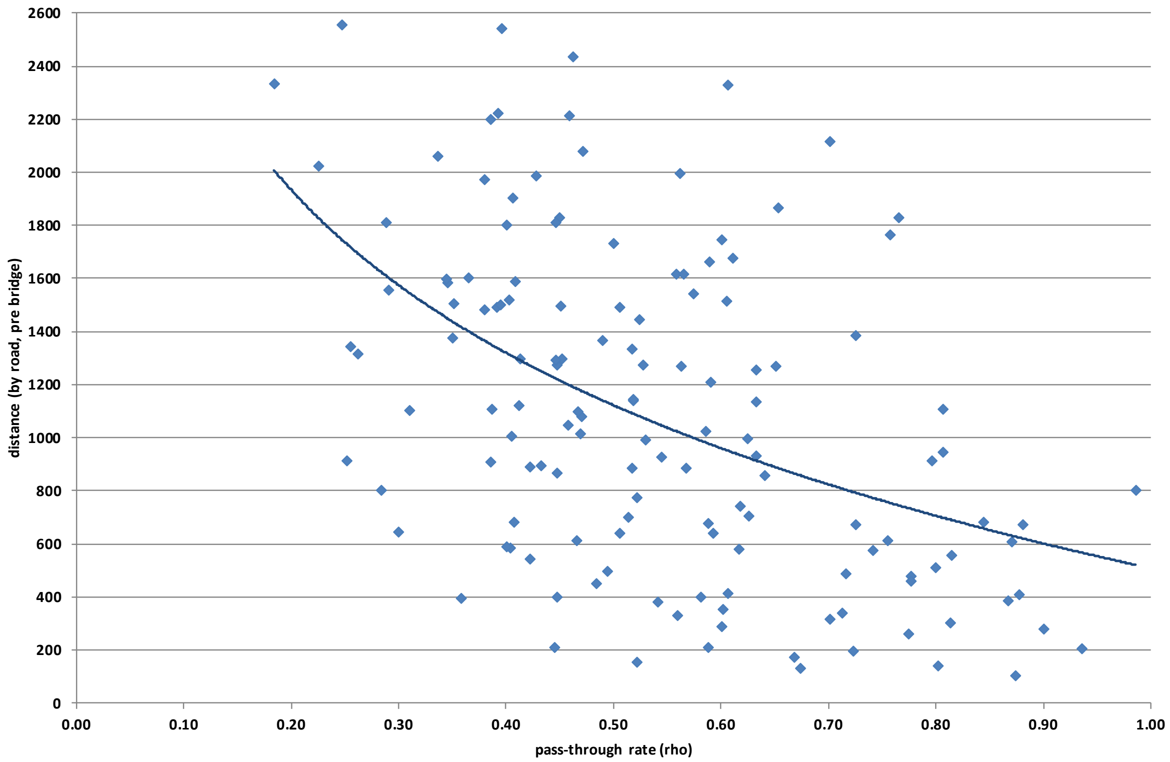
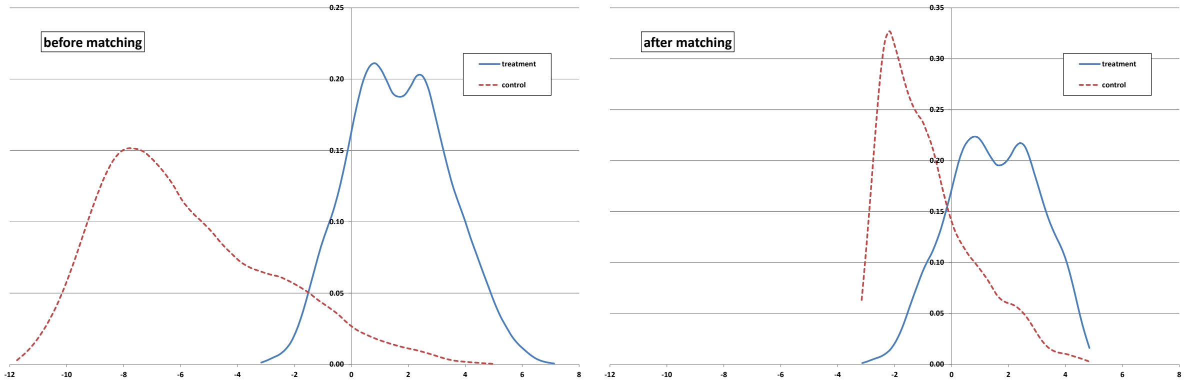
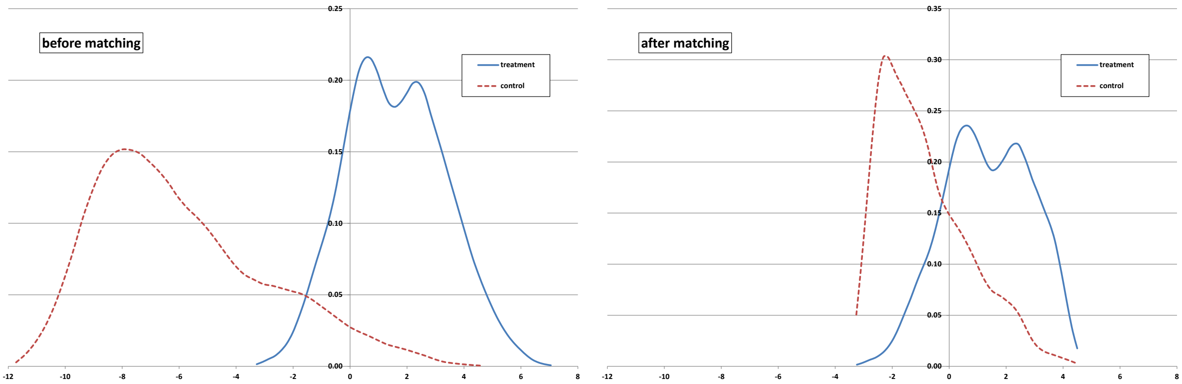


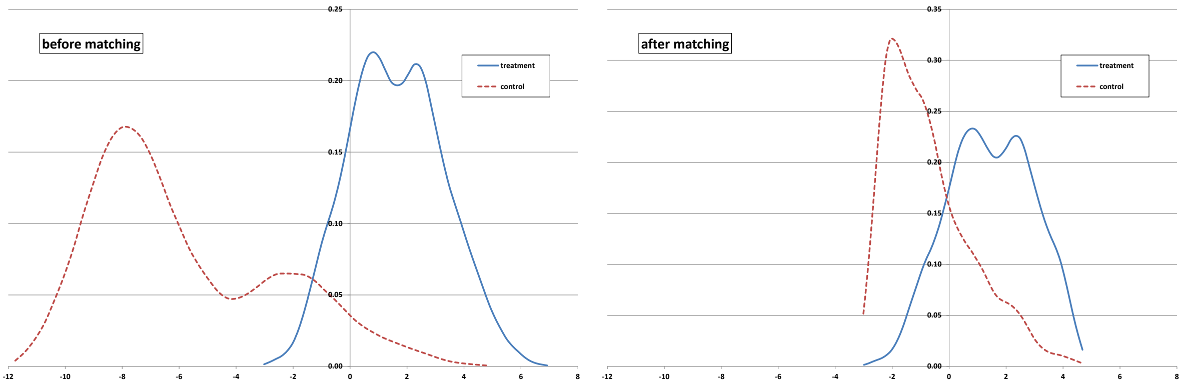
Figure A5 Common support between treatment and control group



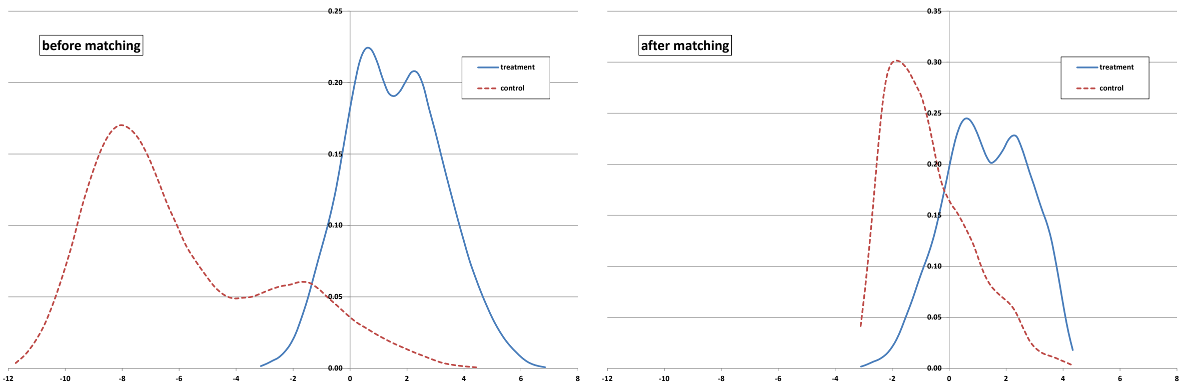
Note: PSM, Kernel Matching, (Table 5, column (1))



Note: PSM, Kernel Matching (Table 5, column (2))



Note: PSM, Kernel Matching (Table 5, column (3))



Note: PSM, Kernel Matching (Table 5, column (4))

Figure A6 Available data over the years: spatial price difference versus transport cost data