TI 2016-054/V
Tinbergen Institute Discussion Paper

How does Market Access for Smallholders affect Export Supply? The Case of Tobacco Marketing in Malawi

Wouter Zant

Faculty of Economics and Business Administration, VU University Amsterdam, and Tinbergen Institute, the Netherlands.

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and VU University Amsterdam.

More TI discussion papers can be downloaded at http://www.tinbergen.nl

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 1600

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

How does Market Access for Smallholders affect Export Supply? The Case of Tobacco Marketing in Malawi

by Wouter Zant*

Abstract

Transaction costs play a key role in the behaviour of smallholders in developing countries. We investigate smallholder tobacco cultivation in Malawi, Malawi's major export crop, and exploit the introduction of an additional tobacco auction floor to measure the impact of a reduction in transaction costs on smallholders' decisions on tobacco crop area and production. Estimations are based on annual data by Extension Planning Area, from the Ministry of Agriculture and Food Security, combined with data from other sources. A 10% reduction in transport cost is shown to lead to an increase in crop area and production of around 4% and 2.5%, respectively. Supply response runs along the extensive margin: both area and production increase, but production slightly less, leading to a decrease in yield, most likely because tobacco cultivation expands to less suitable areas and/or less productive farmers. In view of the non-experimental nature of the data, we confirm impacts by estimating a dose response function using generalised propensity scores. Supply response increases substantially within a distance to auction floor of less than 60km. We find no empirical support for conversion of crop area from maize and other crops into tobacco.

JEL code: D23, O13, O55, Q11, Q13

Key words: transaction costs, market access, subsistence, food & cash crops, Malawi, Africa

***** Wouter Zant is affiliated with the VU University Amsterdam and a research fellow of the Tinbergen Institute, the Netherlands; mailing address: Wouter Zant, 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; fax: +31 20 598 9870.

Introduction

Smallholders in developing countries can choose to produce food crops for home consumption or cash crops for the market[1]. High production costs, high transaction costs, and high risks of output and input prices often make subsistence farming – food production for home consumption – the optimal choice (see e.g. De Janvry et al. 1991; Jayne, 1994; Fafchamps, 1999; Key et al., 2000; De Janvry and Sadoulet, 2006)[2]. Widespread subsistence farming leads to low productivity and low growth in agriculture. And since developing countries have large agricultural sectors with a comparative advantage vis-à-vis non-agricultural sectors, large multiplier effects from agriculture to the remaining sectors of the economy and few alternative growth strategies (de Janvry and Sadoulet, 2010), a stagnant agricultural sector is likely to obstruct the economic growth potential of these countries.

The question arises how can countries overcome this subsistence trap? A possible way out of this trap is to reduce transaction costs for smallholders. Transaction costs – costs incurred in order to sell on the market – include costs of information, collection, loading and transport of goods, bargaining on prices and conditions, monitoring and insurance, with transport costs usually considered the largest component. It is often claimed that transaction costs, which are only partly observed, are large and constitute a major cause of not selling on the market (De Janvry and Sadoulet, 2006). Conversely, improved access to markets – both the mere existence of markets, and also the logistical and marketing infrastructure – decrease transactions costs and should, thereby, trigger smallholders to cultivate crops for the market. Transaction costs, hence, play an important role in explaining the cash crop – food crop decision.

[1] Food crops may also be sold on the market and, hence, are not necessarily or exclusively used for subsistence.
[2] Promotion of either food crops or commercial crops is also at the heart of policy discussions on economic growth and development (see e.g. Harrigan, 2003, 2008).

The current paper aims to contribute to this literature by investigating tobacco production in Malawi. In particular we study the impact of improved market access – caused by the introduction of a new auction floor – on the household decision to grow tobacco. Tobacco in Malawi is by far the most important cash crop and export crop, grown in nearly all Malawi districts and, by regulation, exclusively sold on auctions floors. Moreover, since virtually all production is exported, tobacco is also a key determinant of foreign exchange earnings and government revenues. In 2004 an additional auction floor started operations in Chinkhoma, Kasungu district, on top of the three already existing and operational auction floors (respectively in Limbe (Blantyre), Kanengo (Lilongwe) and Mzuzu (Mzimba). Due to a reduction in transport costs the new auction floor offered farmers in its neighbourhood an opportunity to produce tobacco in a commercially viable way. We exploit the introduction of this new auction floor to quantify the impact of a reduction in transport costs on tobacco area productivity (the intensive margin) and tobacco crop area (the extensive margin). Transaction costs in Malawi tobacco are substantial: estimates (2000/01) are in the range of 14.5% to 22.5% of sales value (see FAO, 2003). Survey based estimates of transport costs are slightly lower (see Section 2). For the empirical measurement of impact we make use of aggregate annual area and production data of smallholders at Extension Planning Area level (EPA), for a period of seven years, from crop seasons 2003/04 to 2009/10, 198 EPAs in total, covering the whole of Malawi.

The paper is organised as follows. In Section 1 we position this study in the literature and highlight its contribution. In Section 2 we describe the Malawi tobacco industry: the importance of tobacco for the Malawi economy, the transition from estate based to smallholder based tobacco production over the past decades and the marketing institutions in the tobacco commodity chain. In Section 3 we show how we measure the impact of improved market access for tobacco smallholders in Malawi. In Section 4 we present and discuss the estimation results. In Section 5 we consider alternative explanations and potential threats to estimated results and run robustness checks. In Section 6 is summarizes and concludes.

1. What does the literature tell?

What causes farmers to grow low yielding food crops for home consumption rather than high return cash crops for the market? And what explains that large groups of farmers prefer not to participate in the market? Various researchers have modelled the decision to grow either subsistence crops or cash crops, and the decision to participate in the market (De Janvry et al.,1991; Goetz, 1992; Jayne, 1994; Omamo, 1998; Key et al., 2000; Renkow et al., 2004; de Janvry and Sadoulet, 2006). Due to transactions costs households have a tendency to get trapped into self-sufficiency and limited participation in the market explains a sluggish supply response (De Janvry et al.,1991). The wedge between producer prices for home produced maize and consumer prices for maize purchased in the market drives the decision to cultivate food crops rather than cash crops, and this wedge is especially large in rural areas, requiring a large decrease of consumption prices to make cash crop production attractive (Jayne, 1994). Transport costs between farms and markets alone are sufficient to account for observed food dominated cropping patterns as optimal responses (Omamo, 1998). The mutual dependence between food crop and cash crop cultivation is similar to non-separable household decisions with incomplete markets (see e.g. De Janvry and Sadoulet, 2006). In a generalization of the model proposed by Goetz (1992), Key et al. (2000) show that both proportional and fixed transaction costs matter: supply response to a price increase is partly due to producers who enter the market (60%), and partly due to those producers who are already sellers on the market (40%). Several researchers take a clear empirical perspective in assessing the implications of high transaction costs (e.g. Fafchamps and Vargas Hill, 2005, and Minten and Kyle, 1999; Jacoby and Minten, 2009). In choosing between selling to an itinerant trader at the farm gate or carrying output to the nearest market town, farm households are more likely to sell to the market when the quantity sold is large and the market is close by, and wealthy farmers are more likely to travel to distant markets (Fafchamps and Vargas Hill, 2005). Differences in food prices between producer regions and urban areas are explained by transportation costs, and road quality is the key determinant of transportation costs (Minten and Kyle, 1999). Large gains in income may be realised from improved road infrastructure for remote households but these gains are small relative to the improved non-farm earning opportunities in town (Jacoby and Minten, 2009). Other empirical work focuses on the impact of search costs – another large component of transaction costs – on behaviour and market prices. Improved market information for households is shown to significantly raise the probability of participating in the market as a seller or a buyer (Goetz, 1992). Various studies exploiting the roll-out of mobile phones have investigated impact on market prices (see for example Aker, 2010; Jensen, 2007). In a particularly relevant study on the soy market in the central Indian state of Madhya Pradesh, Goyal (2010) investigates the impact of a direct marketing channel for farmers, in the form internet kiosks offering price information and warehouses offering quality testing and direct sales to the end-user (a private company), and thereby bypassing intermediary traders. As a result, soybean prices increased, price dispersion decreased and area under soy cultivation increased.

The literature offers persuasive and rigorous evidence, both theoretical and empirical, for the key role that transaction costs play in explaining subsistence farming, on the impact of transaction costs on prices, arbitrage and economic behaviour, on supply, and on the potential welfare improvements that reductions in transaction costs can generate. In the current paper we complement the work of various authors on the choice between food and cash crops, by showing empirically the importance of transport costs in supply response in cash crop cultivation. Contrary to most work (but similar to Goyal, 2010) we investigate the impact of a change in marketing infrastructure. Next, we exclusively look at transport costs rather than search costs, or any other form of transaction costs. And, while most studies consider the price increasing contribution of transaction costs, our analysis exploits a reduction in transport costs (and the consequent but implicit increase in farm gate prices) and investigates its impact on supply responses of farmers.[3] Our estimates support a statistically significant impact of transport costs on cash crop cultivation, that is mainly due to an increase in cultivated area (the extensive margin) rather than enhanced productivity (the intensive margin). A reduction of transport costs is shown to increase welfare and trigger commercial agriculture, and thereby confirms results for other empirical work (Key et al. 2000; Renkow et al., 2004; Jensen, 2007). The empirical investigations also fit the literature that seeks to reveal constraints to export growth strategies and to highlight the potential of export led growth strategies in poverty alleviation (see e.g. Balat et al., 2009).

2. The Malawi tobacco industry

Various articles and publications describe market developments in the Malawian tobacco sector, the evolution of the marketing and regulatory infrastructure over time and the (nearly complete) transformation that took place from the colonial estate based production to smallholder based production, since the end of the 1980s (see Kydd and Christiansen, 1982; Orr, 2000; Diagne and Zeller, 2001; Jaffee, 2003; World Bank, 2004; Poulton et al., 2007; Tchale and Keyser, 2010). We draw extensively on these sources to highlight the key developments and institutional changes which are relevant to the subsequent analysis. Complementary to those descriptions we analyse aggregate historical auction data (1960-2010, source: TCC), the complete 2009 auction transaction data for all Malawian tobacco auction

[3] Unfortunately we do not have sufficiently disaggregated farm gate prices that warrants a useful analysis of the impact of reduced transport costs on farm gate prices.

floors (source: TCC) and tobacco growers information in the Malawi LSMS household data (IHS 2 (2004/05), IHS3 (2009/10) and IHPS (2013); source: NSO).

The role of tobacco in the domestic economy of Malawi

Tobacco is, by far, the most important export product of Malawi accounting for a share of 45% to 65% of total merchandise exports (1994 to 2009, NSO data). The second largest single export products (tea and sugar) account for only a small fraction of total merchandise exports. Tobacco exports also account for about 60% of foreign exchange earnings and, as result of this, the Malawi kwacha - US$ exchange rate tends to fluctuate with export revenues from tobacco (and the tobacco season). All tobacco is exported: Malawi does not have a domestic cigarette industry. The direct contribution of tobacco to GDP, measured as the export value of tobacco in terms of GDP, varies from 9% to 16% (1994 to 2009, NSO data).

Tobacco is cultivated by 19% of the smallholder households, around 375,000 farmers (2004). The bulk of the tobacco growing households – around 65% – are poor or very poor (Economic Council (2000)). In the period from 2003 to 2010 aggregate smallholder crop area allocated to tobacco varied from 141,000 to 184,000 hectares, and smallholder crop production from 95 to 208 thousand tons (source: Agro Economic Survey, Ministry of Agriculture and Food Security). Using a methodology employed by the FAO (FAO, 2003), direct employment in tobacco production and marketing (including processing, transport, auctioning and research) varied from 11% to 19% of total labour supply[4] during 2000-2009.

Tobacco exports generate a major contribution to total government tax revenue in the form of withholding tax levied at the auctions, together with export taxes and export surrender requirements imposed by the Reserve Bank of Malawi. All tobacco taxes and levies add up to an estimated share of 30% in 2000 decreasing to around 20% in 2008 of total

[4] Estimates of direct employment in tobacco production and marketing vary widely with the data used on tobacco and labour supply, even apart from the way employment is related to tobacco area and production.

government tax revenue (Jaffee (2003) reports 23% and FAO (2003) writes: "…tax accounted for more than 20 percent of total national tax revenue")[5].

In summary the figures indicate that tobacco is of extraordinary importance to the Malawi economy. The role of tobacco may extend well beyond these figures, due to indirect effects, backward and forward linkages and dynamics. Some authors claim Malawi's export of tobacco to be the major driver of economic growth (see e.g. Lea and Hammer, 2009).

Tobacco cultivation in Malawi: from colonial heritage to smallholder domination

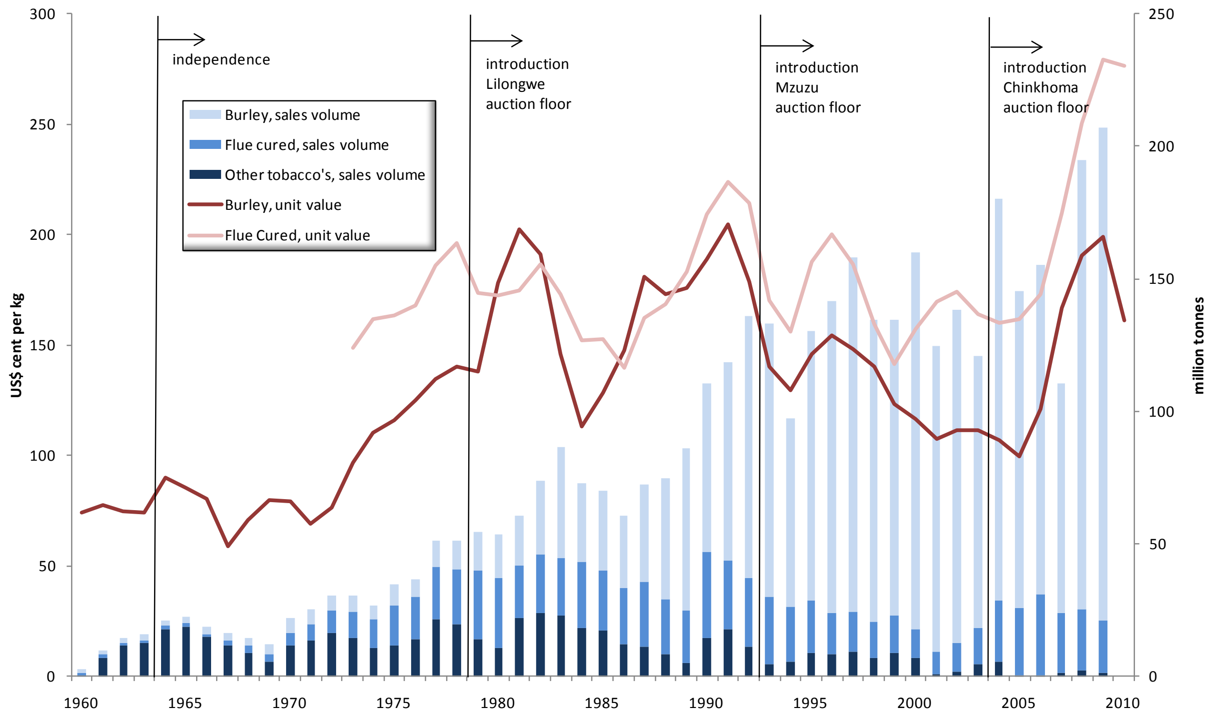
The Special Crop Act of 1964 had created a dual sector, with special privileges for estates and with restrictions for smallholders. Urged by donors to implement liberalisations in the tobacco industry, the 1993 newly elected government introduced amendments to the 1964 Special Crop Act that allowed smallholders to grow burley tobacco, a light air-cured tobacco used primarily for cigarette production (Jaffee, 2003)[6]. The Act was fully repealed in 1996, which included the abolishment of special marketing rights to estates. By 1996/97 all restrictions for smallholders to grow and market tobacco were removed (Diagne and Zeller, 2001). In the course of the 1990s, the change in regulation have given rise to a complete transformation from estate based tobacco cultivation with a high share of western type tobacco's, to a smallholder based tobacco cultivation with a high share of burley tobacco (see Figure 1). High profitability of tobacco as a cash crop – the only really remunerative cash crop available to smallholders – and the wide spread of technical knowledge on tobacco cultivation – since many farmers worked previously on estates as labourers – triggered high growth of smallholder tobacco production. The increase of smallholder production was accompanied by the formation of burley clubs and the introduction of intermediate buyers who provided the

[5] The large share of tobacco proceeds that flows to the Government of Malawi makes the government a major stakeholder in the tobacco industry. Related to this a variety of rents may arise due to lack of competition, lack of transparency and lack of accountability (see e.g. Koester et al., 2004).

[6] Burley tobacco is a light air-cured tobacco used primarily for cigarette production. Western type tobacco's are Flue-cured tobacco (also Virginia), NDDF and SDDF (respectively Northern and Southern Division Dark Fired). These types are smoke and fire dried and aged in curing barns.

logistical link from farmers to auction floors and access to these auction floors (FAO, 2003). Credit for tobacco growers was made available by the Malawi Rural Finance Company (Jaffee, 2003). The transition to a smallholder based tobacco cultivation with a high share of burley tobacco is clearly visible in the data. Over the years, aggregate sales volume at auction floors – a reasonable indicator of aggregate Malawi tobacco production[7] – shows a nearly continuous upward development (see Figure 1), almost entirely due to burley tobacco, with high growth rates of burley tobacco since the end of the 1980s, which slightly levelled off by the end of the 1990s.

Figure 1 Auction Sales Volume and Unit Values of Burley and Other Tobacco[a]



[a] Note to figure – Nominal unit values in US$ cent per kg are on the left axis and sales volume in million tonnes on the right axis. Other tobacco's produced in Malawi are NDDF, SDDF (resp. Northern and Southern Division Dark Fired, so-called western tobacco's) and Sun Air; source: annual aggregate data from the Tobacco Control Commission, Malawi.

[7] Since all tobacco exported from Malawi is required to be sold at auction, unit values and sales volume at auctions are reasonable indicators of average Malawian market prices and aggregate production, despite small quantities of tobacco sourced from Zambia and Mozambique or illegally exported (see e.g. Koester et al., 2004). We assume that these flows are negligible. Note that empirical estimations in this study are based on data from the Ministry of Agriculture and Food Security (Agro-Economic Survey) and not from the Tobacco Control Commission.

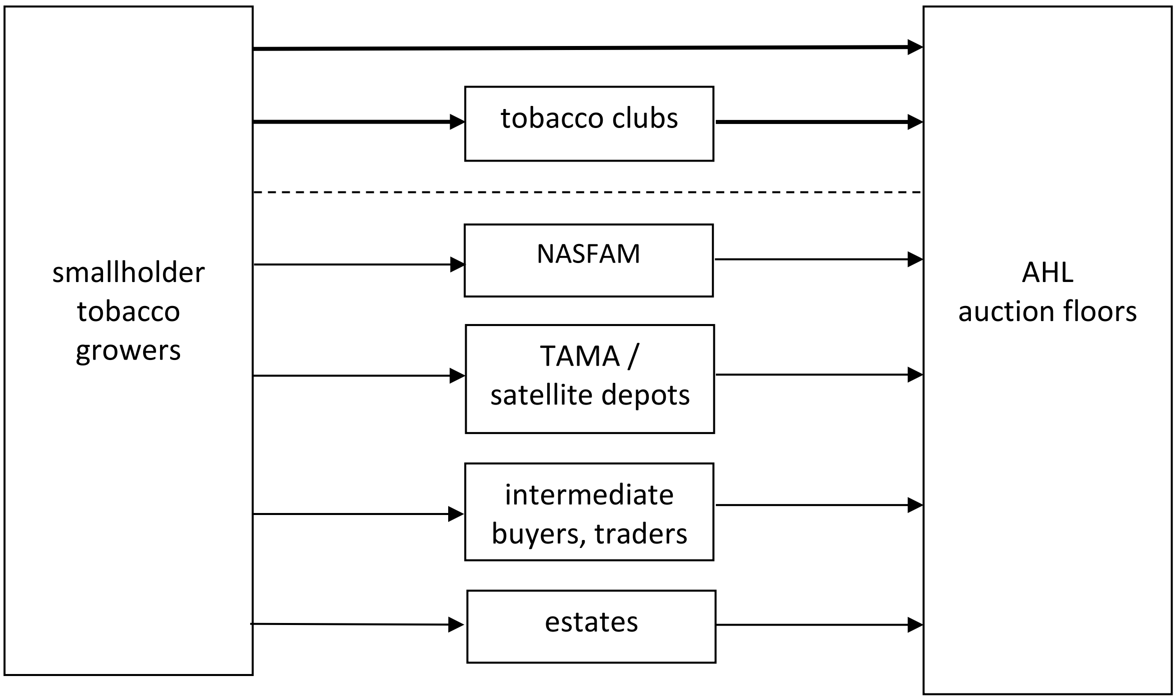
Figure 1 further illustrates the more or less parallel development of burley and flue cured auction unit values, with the latter in most years slightly higher. Visual inspection of the figure suggest that lagged price increases (decreases) coincide with production increases (decreases) in a more or less systematic way, reflecting a positive response of production to (lagged) auction prices (see also Jaffee, 2003 and our own empirical estimates).

Tobacco marketing: auctions, regulations, farmers clubs and other institutions

The tobacco auction system in Malawi has a long history, which dates back to the colonial times, at the start of the 19th century and which was, for a long time, adapted to production and marketing needs of estates. We focus on the period since the 1990s, when the tobacco sector was liberalized. Transport of tobacco to auctions was – both pre and post liberalization – on account of tobacco farmers. Hence, starting in the 1990s a logistical infrastructure for tobacco transport and marketing from rural areas to auctions was put in place to service smallholder farmers. Of key importance in this context are farmer clubs or burley clubs: groups of 10 to 30 farmers that share specific services. Upon registration with the Tobacco Control Commission (TCC) clubs are allocated a quota and are entitled to receive burley seed, fertilizer, advice on cultivation and extension support. From 1991/92 onwards clubs are authorized to sell directly on the auction floors and, since 1994, also to intermediate buyers, introduced in 1994 also to help smallholders to transport their burley tobacco to the auction floors (Orr, 2000). By 1996/97 smallholders were allowed to produce and market tobacco without any restrictions for the first time (Diagne and Zeller, 2001). In general, the process of liberalisation has spurred market access for smallholders: by 1996 83% of smallholder tobacco was marketed directly to the auction floors (Diagne and Zeller, 2001). Access to auctions and thereby access to world market prices, credit facilities and economies of scale in transport are the major incentives for smallholders to join a burley club (Orr, 2000; Negri and Porto, 2008).

Over the period from 2000 to 2010 the number of registered burley clubs nearly tripled from around 20,000 at the start of the 2000s to close to 60,000 in 2010, of which more than half is registered at the Lilongwe auction floor (source: TCC). The existing Tobacco Association of Malawi (TAMA) and the National Association of Smallholders Farmers of Malawi (NASFAM), which was established in the 1990s, also assist in the organisation, collection, storage, transport and sale of smallholder tobacco from rural areas to the auction floors. Shortcomings to the marketing infrastructure – which is continuously developing – were experienced in the area of widely divergent transport rates, storage losses and lack of accountability (see Jaffee, 2003).

Figure 2 Trade channels for smallholder tobacco growers



Source: author

Important for our empirical work is that we assume that transaction costs of tobacco (including transport costs) is on account of smallholders. Evidence supports that most tobacco growing households sell their tobacco directly on auction floors (Diagne and Zeller, 2001: 83% and

IHS2 (2004/05), 85% of smallholder tobacco production sold directly on auction floors. Likewise transaction costs are on account of smallholders. Consequently, net tobacco revenues received by tobacco growers increase if costs of selling tobacco on an auction floor, transactions costs, are reduced.

Figure 3 Tobacco auction floors in Malawi

Source: author

Tobacco marketing is regulated by the Tobacco Control Commission (TCC), a statutory governmental body. The TCC is responsible for market regulation and control, licensing of farmers, quality standards, data & statistics of the tobacco sector, and advising the government on tobacco issues. Operations on all tobacco auction floors are run by a single private sector company, the Auction Holdings Limited (AHL). The establishment of an auction floor requires complementary investments from buyers to properly organize after sales processing, storage and international transport: this makes the auction floor location decision dependent on the support from buyers.

As early as 1939 tobacco was auctioned only at the Limbe auction floor, near Blantyre in the south of Malawi. In more recent years the centre of tobacco production moved in northern direction: auction floors were established in 1979 in Kanengo, near Lilongwe in Central Malawi; in 1993 in Mzuzu in Northern Malawi; and in 2004 in Chinkhoma in the central district Kasungu, between Lilongwe and Mzuzu (see Figure 3). Auction floors normally open from mid-March and close towards the end of October. According to weekly reports from TCC[8] direct trade and contract trade is primarily important for specialty tobacco's (Flue Cured, NDDF and SDDF) and plays a negligible role for the marketing of burley tobacco.

A limited number of companies is active on the demand side. Most companies are subsidiaries of large international traders or international cigarette manufacturers. Over the years the composition of the buying side has changed due to new entrants, mergers and takeovers. However, concentration on the buyer side remains high. The presence of only a limited number of buyers on the auction floors (7 in 2011) raises suspicion of a lack of competitiveness of tobacco pricing and collusion between buyers at the auction floors (see e.g. Koester et al. 2004, Otañez at al., 2007). This is particularly manifest with occasional outbursts of protest from tobacco farmers who complain about the low prices at the auction.

[8] These reports are only available for the period from 2001 to 2006.

Auction transactions: comparison of district composition and unit values

Available auction transaction data for 2009 make it possible to analyse the composition of sales volume and unit values, by auction floor and by district of origin (see Appendix H for an overview of sales volume by auction floor and by district of origin)[9]. More than half of all 2009 tobacco sales originates from the districts Kasungu, Dowa (both in the central region) and Mzimba (in the northern region), comprising the three largest tobacco producing districts. A few districts in the south – notably Mwanza, Chikwawa and Nsanja – have negligible or no tobacco sales. One third of total 2009 sales volume is traded on the Mzuzu auction floor, slightly above 10% on the Chinkhoma auction floor, while the Lilongwe auction floor and the Limbe auction floor have a share of around a quarter each of the total sales volume[10]. The number in Table H1 confirm that the choice of auction is mainly determined by transport costs, as most tobacco is sold on the most nearby auction floor. There are, however, a few exceptions. The Limbe auction floor is more popular among tobacco growers from the districts Salima, Mchinji and Nkhotakota than the more closely located Lilongwe auction floor (see Appendix H). This is likely to be explained by congestion at the Lilongwe auction floor, cheap transport alternatives from Mchinji and Salima (transport by train[11]) and different prices at auctions (see below). Tobacco farmers from other districts also appear to avoid the Lilongwe auction floor (e.g. Ntcheu). Finally, one would expect tobacco sourced from Mzimba to be sold on the Chinkhoma auction floor: these sales were, however, negligible in 2009. This may be caused by the (still) moderate trading volumes at the – only a

[9] We use transaction data of all tobacco auction transactions – a total of around 60,000 transactions – for the year 2009, which are kindly made available by the Tobacco Control Commission. The transaction data pertain to a year that comes five years after the introduction of the Chinkhoma auction floor in 2004.

[10] At the start of the 2000s the Lilongwe auction floor at Kanengo was the largest auction floor in terms of volume of turnover (close to 60% of sales volume, see Jaffee, 2003 and Koester et al. 2004). The available 2009 TCC transaction data indicate that Lilongwe has lost substantial market share.

[11] Both Mchinji, Salima and Ntcheu have railway stations along the Malawi railway line to the south which potentially offers these locations low cost transport services to the Limbe auction floor, near Blantyre in the south (Limbe is along the railway line and also has a railway station). Freight data from the Central and East African Railways (CEAR, www.cear.mw) confirm that – apart from tobacco freight for export – tobacco is transported from rural areas to the Lilongwe and Blantyre auction floors, both located along the railway line.

few years earlier established – Chinkhoma auction floor and the – on average – higher prices at the Mzuzu auction floor.

Since we investigate the impact of the newly established auction floor at Chinkhoma, our primary interest is the districts of origin of tobacco auctioned on this floor. Do tobacco growers in a specific district sell either exclusively on the Chinkhoma floor or use various floors as outlet for their tobacco? Is there a price premium for a specific auction floor? Hence, for these districts we have calculated average unit values of burley tobacco for the different auction floors (see Table 1). The table indicates that average prices for burley realized on the Mzuzu auction floor are highest, while the average prices realized on the Lilongwe auction floor are lowest[12]. In terms of 2009 realized auction prices the Chinkhoma auction floor offers an attractive alternative outlet to the Lilongwe auction floor.

Table 1 Burley tobacco prices by auction floor of sale and district of origin in 2009[a]

auction floor / district	Mzuzu	Chinkhoma	Lilongwe	Limbe
Nkhotakota	1.74 (308)	1.62 (284)		1.51 (370)
Kasungu	1.78 (4576)	1.67 (4453)	1.54 (5040)	1.65 (1146)
Ntchisi	1.70 (366)	1.61 (949)		1.63 (566)
Dowa	1.72 (1387)	1.60 (2740)	1.54 (8201)	1.58 (1649)
Mchinji	1.49 (435)	1.45 (652)		1.56 (1397)
Lilongwe	1.72 (370)	1.48 (864)	1.57 (1661)	1.59 (1602)
Mzimba	1.78 (6812)	1.66 (339)		
Salima				1.54 (509)

[a] Notes – The table reports average transaction prices for 2009 in US$ per kg by district of origin, for districts of the Central region. Source: transaction data of 2009 from the Tobacco Control Commission. Number of transactions are in brackets behind the average price. With a few transactions (<100) numbers are omitted.

[12] Implicitly we assume that seasonality in auction unit values is negligible. This is supported by the evidence. From other data (unit values by week calculated on the basis of tobacco value and volume of all auction floors, from 2001 to 2006, six crop seasons, source: TCC, data not shown here, available on request) we could not detect systematic seasonality in tobacco auction unit values (or tobacco sales volume).

Household survey data

Selected household data on tobacco growers, extracted from the Malawi Integrated Household Surveys (IHS2(2004/05), IHS3(2010) and IHPS(2013)), summarized in Table 2, suggest that unit values have increased over the years, especially in 2013 but show little systematic pattern between auction floors, that per ton-km transport costs also have increased and are particularly high for the Lilongwe auction floor, and transport costs account for 5.0% to 9.8% of sales value and this share depends to a large extent on the tobacco price realized on the auction floor. The household data confirm that nearly all tobacco growers sell their tobacco on auction floors themselves and also bear all costs of this activity. Some other information extracted from the IHS (not shown in the table) is worth mentioning. Household size of tobacco growers is systematically larger (relative to rural non-tobacco growing households) reflecting the high labour requirements of tobacco cultivation and total crop area for tobacco growers is relatively large, but the tobacco share in total crop area is small.

Table 2 Tobacco growers in household survey data (IHS2, IHS3 and IHPS)

auction floor / variable, year	Mzuzu	Lilongwe	Chinkhoma	Limbe	Malawi
Unit value (constant 2013 prices)					
2003/2004	302.1	254.1	234.3	238.3	246.5
2010	331.9	286.5	351.3	337.0	324.6
2013	486.2	474.3	536.2	418.4	484.6
Cost per ton-km (constant 2013 prices)					
2003/2004	171.5	461.2	140.6	118.1	211.8
2010	401.6	697.6	676.7	401.2	557.8
2013	445.5	695.7	765.9	536.5	629.4
Cost share in sales value (in %)					
2003/2004	7.3%	6.4%	7.5%	5.7%	6.6%
2010	9.8%	9.0%	8.2%	8.0%	8.8%
2013	5.2%	5.9%	5.0%	7.1%	5.6%

* The table reports weighted sample averages of unit value, cost per ton-km and cost share in sales value. Unit values are sales values by sales volume (MK/kg). Cost per ton-km is transport costs by sales volume and distance to auction floor (MK/ton/km). Cost share is total transport costs as a share of sales value. All data are household reported. For the IHS2 data we restricted the exercise to tobacco sold at auction floors (comprising 65% of the sales transactions and 85% of sales volume).The rural consumer price index is used to convert current prices into constant 2013 prices (source: NSO).

3. Data and methodology

Data for analysis

The estimations in the empirical section are based on annual data of agricultural production and crop area on the level of Extension Planning Area's (EPAs) from the Agro Economic Survey of the Ministry of Agriculture and Food Security (AES-MoAFS). Extension Planning Areas (EPAs) are subdivisions of districts and have an average size 470 km2 (median: slightly above 400 km2 [13]), an average population of around 65,000 (median: 60,000) and an average of around 20,000 households (median: 19,000). Data on production and area by EPA are available for the crop years from 2003/04 to 2009/10 (seven crop years). The EPA data, consisting of a total of 198 EPAs, cover the whole of Malawi[14]. There are only a few EPAs, located particularly in the southern districts Chikwawa and Nsanje, that have no or negligible tobacco cultivation. We have used these EPA data because they are most complete, comprehensive and detailed about smallholder area and production dynamics. None of the available alternative data (IHS/NSO, TCC) allow insight into the dynamics of tobacco production and area, in a wide range of locations in Malawi, and covering the period that the new auction floor was introduced.[15]

Distance from EPAs to the different auction floors is Great Circle Distances on the basis of latitude and longitude coordinates.[16] Distance measured as the crow flies is different from road distance and road distance is the relevant concept for transport costs. However, the difference between road distance and distance measured as the crow flies is more or less

[13] A few EPAs are larger than two times the mean size, but the bulk of the EPAs (90%) is smaller than 835km2.
[14] The EPA data cover the land area of Malawi that is relevant for agriculture. Some parts of the country (e.g. national parks and lakes) are excluded from the EPA data.
[15] However, the data do have drawbacks. A major drawback is the lack of other information on EPA level. We have overcome this drawback by matching the EPA data with data from other sources (see below).
[16] Great Circle Distance is the shortest path between two points on the surface of a sphere (for calculation see e.g. www.cpearson.com). In these calculations it is assumed that the earth is a perfect sphere: given the small distances on a global scale this entails only a marginal and hence negligible error.

proportional[17,18] and, consequently, using road distance data leads to (nearly) the same intervention locations and similar conclusions. Another potential measurement error may arise because latitude-longitude coordinates for identification of EPAs – usually the main town / village in the EPA – will not necessarily coincide with the tobacco area in the EPA. This generates a potential measurement error that is correlated with the size of the EPA.

Table 3 Summary of data for estimation

Variable	observations	intervention mean	intervention standard deviation	control mean	control standard deviation	| t |
rural / urban						
consumer price index	7	-	-	-	-	-
market level data						
tobacco price (MK)*	140	71.6	34.0	67.5	28.3	2.2
maize price (MK)*	341	22.3	7.9	22.2	8.1	0.2
groundnuts price (MK)*	429	146.2	51.1	156.0	46.3	3.2
weather station level data						
current rainfall (mm)	224	981.6	298.4	929.2	242.3	3.0
EPA level data						
crop cultivation						
tobacco area (1000ha)	1318	2.0	1.4	0.5	0.7	25
maize area (1000ha)	1318	10.8	5.4	7.7	3.8	11
groundnuts area (1000ha)	1318	2.3	1.6	1.1	0.9	16
pulses area (1000ha)	1318	3.1	2.3	3.0	3.1	0.1
total crop area (1000ha)	1318	20.6	8.8	15.5	7.6	9.7
tobacco production (ton)	1318	1698	1407	474	722	20
maize production (ton)	1318	19769	13102	11660	7642	13
groundnuts production (ton)	1318	2195	2098	926	1001	14
pulses production (ton)	1318	1930	1774	1866	2251	0.4
Population						
population density (people/km2)	1372	138.1	59.9	204.2	179.3	6.4
Geography						
distance to town (km)	196	39.2	22.4	36.2	19.7	2.2
distance to city (km)	196	88.6	28.6	66.1	46.5	7.8
agglomeration index	1372	13.7	4.1	17.9	10.6	6.9

Note to table: Observations of data at the rural-urban level, at the market level, and at the weather station level pertain to the number of independent observations in the original data set. Mean, standard deviation and t test for all variables are, however, calculated on the basis of EPA level data or the EPA level variants of market level data and weather station level data. * in constant 2009/10 prices.

[17] A few exceptions apply: distance measured *as the crow flies* would imply that transport from a few EPAs in the district Mangochi to the Lilongwe - , Chinkhoma - and Mzuzu auction floors runs across lake Malawi. In principle transport by ship is feasible, but transport costs differ drastically from transport costs by truck. Hence, in these instances, distance measured *as the crow flies* is adjusted to reflect transport costs by road.

[18] Note that road distance varies over time (and we do not always know how), while distance measured as the crow flies does not change over time. Comparing road distances extracted from Google Maps (retrieved in 2013) showed road distance to be on average 25% to 28% higher than distance measured as the crow flies.

Apart from data on other crops (production, crop area), there are virtually no complementary and publicly accessible data at EPA level that can be used to analyse tobacco production and area. We have resolved this by matching data from other sources to our EPA data. Consequently, the level of aggregation differs between data and measurement error is introduced in attributing data to EPAs and in constructing missing observations. Data on crop prices are, for example, available for respectively 50 to 70 markets and in varying degrees of completeness. Rainfall data are available for around 30 weather stations, but fortunately without missing observations. For some data the approximate nature of variables may entail measurement error: distance to city or town, or distance weighted population of cities and towns are possibly inaccurate measures for transaction costs and agglomeration effects. In summary, we have data on crop prices, production and area of alternative crops, total crop area, rainfall, population, land area and various distances to analyse dynamics in tobacco cultivation.

Statistics of these variables are shown in Table 3: most variables differ between intervention and control group, which is, apart from the price data, merely a reflection of the differences between EPAs. Maize prices are similar, which is an interesting exception. Current rainfall is slightly higher in the intervention EPAs. Spatial and integration variables indicate the intervention EPAs are somewhat more remote, less integrated and with lower population density. Further details on data, data sources and variable construction are in Appendix A.

Intervention locations

Tobacco farmers that benefit from the introduction of the new auction floor in Chinkhoma in 2004 are identified by determining the minimum of the distances from each EPA to the different auction floors. If in 2004 the Chinkhoma auction floor has become the closest auction floor, tobacco growers in those EPAs have realised a reduction transport costs to the auction floor. Practically this implies that all EPAs in the districts Kasungu and Nkhotakota, a large part of locations in the districts Ntchisi, Dowa and Mchinji, and a few in the district Mzimba are intervention locations. In all this concerns 31 EPAs / locations, 15.3% of all locations.

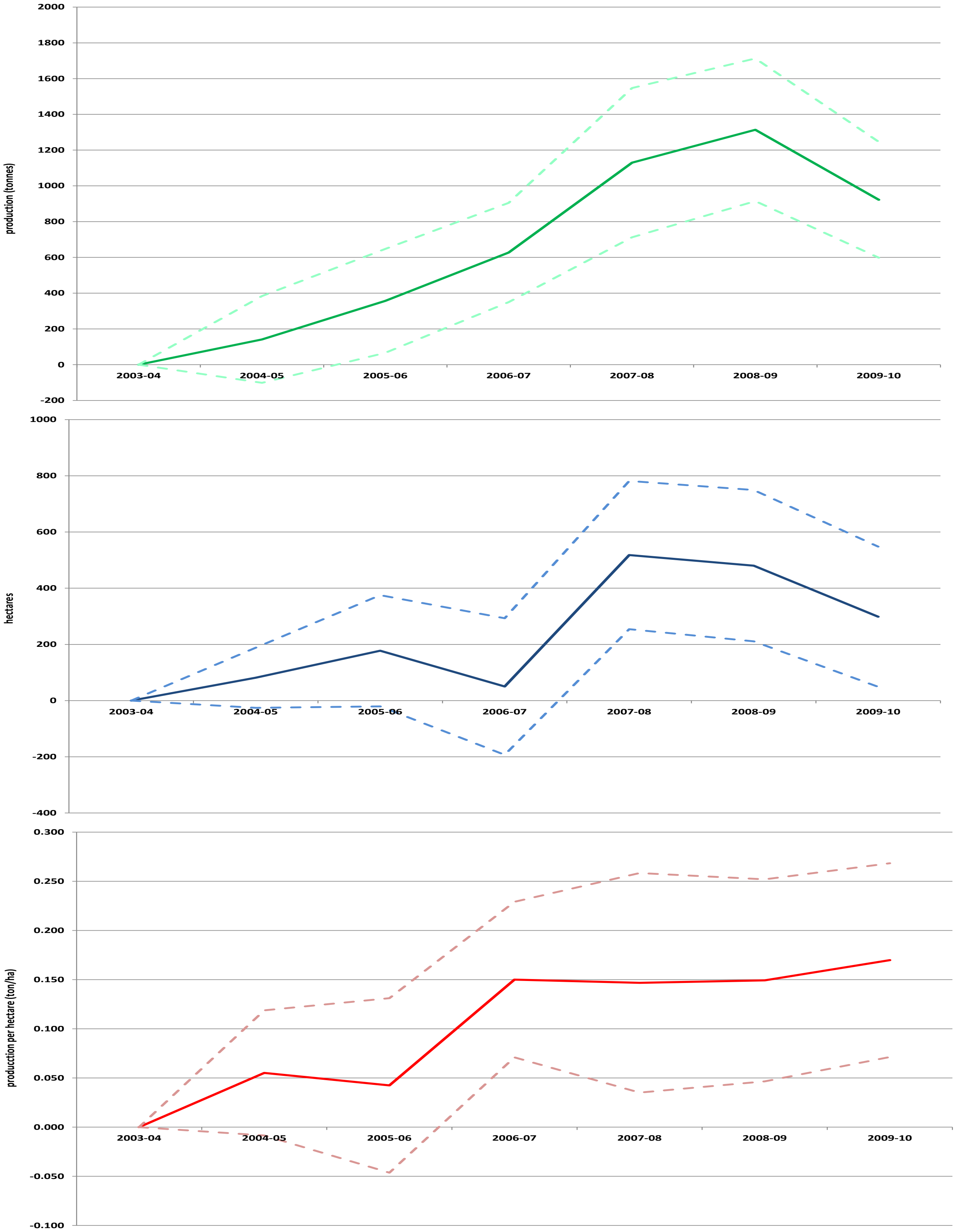
The distribution of sales by district of origin shows that the Chinkhoma auction floor also attracts tobacco outside these districts (e.g. from Lilongwe, Rumphi and Salima district; see Appendix H). Adhering to the rule that "the Chinkhoma auction floor has become the closest auction floor" for the identification for intervention EPAs, is apparently too strict. We assume that this is caused, at least partially, by inaccuracies in the measurement of distance (see *data for analysis* paragraph). Therefore we further consider those EPAs for which the new Chinkhoma auction floor has become the second closest auction floor and where the size of the difference in distance between the closest and the second closest auction floor is less than the potential measurement error in the distance from EPA to auction floor. Potential measurement error in distance to auction is correlated with the size of the EPA and is approximated on the basis of (the root of the) EPA land area. On these grounds we have identified another 14 EPAs, summing to a number of 45 intervention EPAs (22.7%), out of a total of 198 EPAs that potentially benefit from the newly established auction floor.

Event estimates on outcome variables

As a preliminary enquiry we run an event study estimation. Such an estimation is helpful in finding the exact timing of the impact of the intervention, its distribution over the years and possible non-linearities. In this estimation we estimate the outcome variables on a binary intervention variable that is an interaction of years with intervention EPAs, jointly with EPA and year fixed effects, and a polynomial of trends, in formula:

$$Y_{it} = \beta_0 + \sum_t \beta_{1t} I_{jt} + \sum_{km} \beta_{2km}\, trend_{kt}^m + \varphi_i + \omega_t + \varepsilon_{it}.$$

Figure 3 Impact of lower transport costs on production, area and area productivity



Note: we estimated $Y_{it} = \beta_0 + \sum_t \beta_{1t} I_{jt} + \sum_{km} \beta_{2km}\, trend_{kt}^{m} + \varphi_i + \omega_t + \varepsilon_{it}$ and plotted β_1 over t years.

The outcome of this estimation, plotted in Figure 3, suggests a clear impact for all outcome variables, which – not surprisingly – runs more or less parallel. Impacts are delayed and only show up from 2006-2007 (production and production per hectare), and 2007-08 (area) onwards. Since we included a polynomial of trends to find the event estimates, we assume that a log transformation of variables in the impact estimations is likely to work well. Direct evidence through increases in tobacco farm gate prices would also be useful. However, because of its constructed nature we have less confidence in event estimations for tobacco farm gate prices (see, however, Appendix C).

How do transaction costs influence farmers' behaviour?

The theoretical background of the impact of (a reduction in) transaction costs on crop area and production derives from the model of rural household behaviour under market failures (Goetz, 1992; Key et al, 2000; De Janvry and Sadoulet, 2006) which we closely follow. Households maximise the expected present value of current and future utility, where utility is determined by consumption and preferences. In formula: $max\, E \sum_t \beta^t u(c_t, z_t)$ where β is the rate of discount, u is utility, c is consumption, z are preferences and subscripts and power t is time. For each year the households choice variables are consumption, production, inputs, marketed surpluses and savings. Maximization is subject to a budget constraint where the budget constraint, again for each period, is the balance of value of marketed surpluses (both positive and negative) over all goods, fixed transaction costs, transfers and savings. Marketed surpluses, both positive and negative ones, are evaluated at their relevant prices, where purchase and sales prices are assumed to be different, and include variable transaction costs. In formula the budget constraint for a tobacco seller and a (net) maize buyer household is:

$$\left[\left(\left(p^m_{tb,t} - t^p_{tb,t}\right) * m_{tb,t} - t^f_{tb,t}\right) + \left(\left(p^m_{mz,t} + t^p_{mz,t}\right) * m_{mz,t} - t^f_{mz,t}\right) + \cdots\right] + T_t + s_t = 0$$

where p^m is market price, t^p is proportional transaction costs, t^f is fixed transaction costs, m is market surplus (negative in case of maize) and subscripts are crop (tobacco and maize) and time. (The dots $(\ldots)$ stand for the n other crops and goods, which are ignored for the sake of simplicity). A key and extensively studied property of this household model, so-called non separability, is that household responses in one market are affected by failures in other markets. An increase in cash crop prices can only lead to reallocation of land and labour of a household from food crops to cash crops if the increased revenues offsets the decrease in food production and the increased costs of the household on purchased food. This explains low supply response in case of market failures and, hence, also in case of high transaction costs. Alternatively, a productivity increase in food production or a drop in prices of imported food may free up land, labour and other inputs for cash crop production.

With both variable and fixed, household specific transaction costs it is clear that sales and purchase prices for each good are also household specific as well. If households rely for food on own production and produce a cash crop for exchange on the market – reasonably alike tobacco smallholders in Malawi who also cultivate maize – the formal solution to the model implies a shadow price of the home consumed crop, or in formula: $q(p^*_{mz}, p_{tb}, p_q, z_q) = c(\, p^*_{mz}, p_c, y^* z_c)$ where y^* is household income which is written as: $y^* = p^*_{mz} q_{mz} + p_{tb} q_{tb} + p_q q_q + T$. Participation in cash crop market depends on the shadow price of food relative to the market price of the cash crop net of variable transaction costs, in the case of variable and no fixed transaction costs. In the case of fixed transaction costs the gain in utility of cash crops additionally needs to offset the fixed transaction costs. If indirect utility V, a function of prices and income, market participation as a tobacco seller requires (De Janvry and Sadoulet, 2006):

$$V\left(p_{mz}^{m} + t_{mz}^{p}, p_{tb}^{m} - t_{tb}^{p}, \left[\left(\left(p_{mz}^{m} + t_{mz}^{p}\right)m_{mz} - t_{mz}^{f}\right) + \left(\left(p_{tb}^{m} - t_{tb}^{p}\right)q_{tb} - t_{tb}^{f}\right) + T, z_c\right]\right)$$

$$> V(p_{mz}^{*}, y^{*}, z_c)$$

where V is indirect utility.

Estimating the impact of a change in distance to the tobacco auction floor

The objective of this investigation is to measure supply response due to a change in market access. More specifically we measure if and to what extent the new auction floor has given rise to changes in area (the extensive margin), production and yield (the intensive margin). To this end a standard framework is used to estimate the impact of the introduction of a new auction floor in Malawi (the intervention), which reads in formula:

$$\ln Y_{it} = \beta_0 + \beta_1 \ln(distance\ to\ auction\ floor_{it}) +$$

$$\sum_k \beta_{2k}\, lnX_{k,it} + \sum_j \beta_{3j}\, trend_j + \theta_i + \vartheta_t + \varepsilon_{it}$$

where Y_{it} is the outcome indicator and stands resp. for area, production and yield of EPA i in year t, $X_{k,it}$ is a set of k exogenous or predetermined covariates, $trend_j$ is a trend for region j, θ_i are EPA fixed effects, ϑ_t are year fixed effects, and ε_{it} is an error term with zero mean and constant variance. Outcome and intervention variables, and exogenous and predetermined covariates are all transformed into natural logarithms to allow for non-linearities in variables. The coefficient of the variable *distance to auction floor* (β_1), the parameter of interest, measures the average impact of changes in *distance to auction floor* on the dependent variable. We use the above specification as the basic specification of our estimations, with restrictions on coefficients which are apparent from the tables[19].

Covariates ($X_{k,it}$) derive from profit maximizing behaviour of farmers. Expected profits are determined by expected yield, and expected input and output prices. Hence, both tobacco

[19] In comparable work a popular specification also includes interactions of the intervention variable and covariates: since interpretation of coefficients becomes complicated we have omitted estimating such a specification.

area and production are likely to respond to expected prices of tobacco (see e.g. Figure 1 in this study, but also Jaffee, 2003) and expected prices of alternative crops. Therefore, we use previous year farm gate tobacco prices and previous year farm gate prices of other crops (maize, the major alternative crop, but also groundnuts and/or pulses), since tobacco area and production decisions are guided by tobacco prices in the first place, but are also made jointly with decisions on alternative cultivation options. We use farm gate prices since these prices – in contrast with tobacco auction prices – exclude transaction costs and drive smallholder decisions[20]. Given missing markets and non-separability of household behaviour cash crop decisions are also likely to respond to food crop developments like yield (De Janvry and Sadoulet, 2006). This justifies to include lagged yield of food crops as covariates. Additionally, like in most rain-fed agriculture, output of tobacco cultivation is determined by rainfall. Therefore we employ current rainfall as a covariate, particularly in tobacco production. Finally, we experiment with several geographical variables to approximate other transaction costs, access to markets and other agglomeration effects, like population density (for within-EPA transaction costs and agglomeration effects), distance to city or town, and a distance weighted city and town size index (for outside-EPA transaction costs and agglomeration effects). Data sources and data construction are explained in detail in Appendix A.

Using generalised propensity scores and dose response functions for identification of impacts

The choice of location of the new tobacco auction floor is not the result of a random assignment. The auction company will have carefully considered alternatives and investigated the optimal location for doing this investment, basing its eventual choice on an assessment of

[20] Because of its constructed nature (see appendix A) we are reluctant to use the farm gate tobacco prices as direct evidence for the impact of the reduction in transport costs caused by the new auction floor. The results of an event plot are, however, quite supportive (see Appendix C). The use of farm gate prices as covariates (rather than auction prices) is further justified since transaction costs are only observed to a limited extent, and larger than (observed) transport cost (see de Janvry and Sadoulet, 2006).

current and expected turnover of tobacco and long run profit potential of auction services at different (hypothetical) locations. Consequently, causality may not run (only) from market access to decisions of tobacco growers to grow tobacco, but also the other way around, from (expected) tobacco production to the establishment of an auction floor. As a result, OLS estimations are potentially biased: estimates may not reflect the isolated impact of a change in transaction costs.

We propose a propensity score method to reduce the potential bias in the estimations: unfortunately, propensity score matching models are mainly developed for binary treatment variables and hence not suitable for our continuous *distance to auction floor* treatment variable. An extension of the propensity score method for a continuous treatment setting is developed in Hirano and Imbens, 2004, and will, hence, be used for the empirical work in this study. We briefly summarize this method and the associated STATA commands (see Bia and Mattei, 2008), with particular attention to the underlying assumptions of the treatment variable and testing the balancing property. The continuous treatment corollary to binary treatment propensity scores is the Generalised Propensity Score (GPS). Likewise, the GPS can be used to eliminate biases associated with differences in the covariates. The propensity function is defined as the density of actual treatment conditional on covariates ($r(t,x)=f_{T|X}(t,x)$) and the Generalised Propensity Score is R=(r(T,X). The GPS balancing property requires that within strata with the same value of the function r(t,x), the probability of a specific value of the treatment (t) should be independent of the covariates (X). Hence, if we estimate the GPS r(t,x) within a stratum, it should be the same for different values of t.

The function that maps the relationship between outcome and continuous treatment is labelled the (unit-level) dose response function ($Y_i(t)$): we are interested in the average dose response function. The dose response function is estimated at a particular level of the treatment as the average of the conditional expectation over the GPS at that particular level of the treatment, in formula: $\mu(t) = E[\beta\ (t;\ r(t;X))]$. Although Propensity Scoring techniques are only a modest improvement over OLS, GPS and the dose-response estimation that employs GPS, does offer more flexibility in functional form (see also the empirical section).

4. Impact of market access on tobacco area, production and yield

Selected estimation results for our basic specification, using OLS and not controlling for the endogeneity of the intervention variable, are reported in Table 4. The selection is made on the basis of a grid of estimations with varying lags in impact (varying from instantaneous to 3 years), and with various sets of EPAs confronted with reduced transport costs as a result of the introduction of a new auction floor (see Appendix D for grid estimations and alternative definitions of intervention EPAs). The selected estimation results shown in Table 4 assume a two year lag in impact response since this is supported by the grid of estimations. Additionally a relatively broad definition of intervention EPAs is chosen. The estimation results appeared to be hardly sensitive to differences in this definition. In the table the dependent variable is yield (column 1-2), production (column 3-4) and crop area (column 5-6). Each second column shows estimation with a correction for outliers.

Table 4 Market access in tobacco: OLS, only fixed effects and trends

dependent variable	ln(production per hectare)		ln(production)		ln(area)	
	(1)	(2)	(3)	(4)	(5)	(6)
ln(distance to auction floor)	-0.180***	-0.140**	0.078	0.238**	0.259***	0.377***
	(0.0564)	(0.0561)	(0.1186)	(0.1067)	(0.0971)	(0.0976)
$dEPA_i$	yes	yes	yes	yes	yes	yes
$dYEAR_t$	yes	yes	yes	yes	yes	yes
$TREND_k$	yes	yes	yes	yes	yes	yes
number of observations	1239	1227	1239	1227	1239	1227
R^2	0.5337	0.5327	0.9422	0.9426	0.9600	0.9601
RMSE	.30151	.30235	.54485	.54118	.44259	.43999

Notes – The table reports estimates of average impact of distance to auction floor on tobacco area, production and yield. Estimations are based on annual data from 2003-04 to 2009-10 (seven years). All equations are estimated with OLS. Robust standard errors clustered by EPAs are given in parentheses (.) below the coefficient. R^2 = coefficient of determination, and RMSE = Root Mean Squared Error. * means significant at the 10% level ($p < 0.10$), ** at the 5% level ($p < 0.05$), *** at the 1% level ($p < 0.01$).

Estimations, reported in Table 4, on yield (columns 1-2) indicate that the coefficients of the distance to auction floor variable are negative, statistically significant and range in size from –18% to –14%. The coefficients of the distance to auction floor variable in the production equation (columns 3-4) is, as expected, positive and weakly significant, with all observations, but highly significant if outliers are omitted, in which case the coefficient is +24%. In the crop area estimations (columns 5-6) the coefficients of the distance to auction floor variable are, also as expected, positive, but coefficients are now highly significant both estimations, and range in size from +26% to +38%. The estimation results for the different variables are mutually consistent (in all coefficients).

On the basis of these estimations we infer – and this is a key outcome of this work – that a reduction in transport cost leads to an increase in crop area and production. Coefficients may be interpreted as elasticities: under the assumption that the distance to the auction floor is a good approximation of transport costs, a 10% reduction in transport cost is likely to lead to an increase in crop area of around 4% and an increase in production of around 2.5%. The impact of a reduction in transport costs is larger on crop area than on production, leading to a negative impact on aggregate area productivity. This is likely to come about as a result of cultivating less productive crop land or as a result of less experienced farmers engaging in tobacco cultivation.

We are particularly interested to know if the impact of a reduction of transport costs runs through the intensive margin (an increase in production per hectare), through the extensive margin (an increase in area) or both, and which impact dominates? If the increase in production is the same or smaller than the increase in area, supply response through the extensive margin dominates. Conversely, if the increase in production is smaller than the increase in area supply response through the intensive margin dominates. With a larger response in area, the estimations in Table 4 indicate that the supply response rolls out mainly along the extensive margin. Increase in production is due to a change in cultivated area, rather than a change in area productivity, and, in fact area productivity even decreases. Apparently increasingly less productive crop area is used in tobacco cultivation and less productive farmers engage in tobacco cultivation.

Although the data used for estimation are aggregate data and do not allow to infer how individual farmer household behave, it is likely that the increase in area also extends to additional and poor smallholders households, rather than pertaining only to an extension of crop area of existing tobacco growers. Existing tobacco growers are less likely to realize substantially lower yields on expanded crop area, while new tobacco growers are likely to use less suitable soils and need to learn about cultivation practices. Conversely, a supply response along the intensive would, much more, be associated with making steps on the learning curve, increasing inputs like fertilizer and irrigation, stepping up crop maintenance, or, more in general, intensification of tobacco cultivation. Such a development is probable to take place on well-to-do farm households that improve on their already existing tobacco cultivation. Increases in area productivity are costly and therefore often beyond the opportunities of poor households. Hence, evidence of a supply response along the extensive margin is good news: it is more compatible with responses of poor smallholders and suggests that poor smallholders are triggered to allocate (more) area to commercial agriculture. This re-allocation is thus likely to lead to poverty alleviation and improved welfare.

In order to investigate if the results obtained in Table 4 are robust we have re-run the estimations for identical sample periods while including covariates as controls. We selected rainfall, lagged tobacco and maize prices, lagged area productivity in tobacco and agglomeration as covariates: these variables are all predetermined, not correlated with the intervention variable but correlated with the dependent variables. We expect rainfall to affect production and production per hectare (and much less crop area), while the other covariates can be expected to affect both area and production (and therefore possibly much less production per hectare).

Table 5 Market access in tobacco: OLS, fixed effects, trends and covariates

Dependent variables	ln(production per hectare)		ln(production)		ln(area)	
	(1)	(2)	(3)	(4)	(5)	(6)
ln(distance to auction floor)	-0.116**	-0.148**	0.247**	0.304**	0.361***	0.453***
	(0.0539)	(0.0609)	(0.1163)	(0.1284)	(0.1083)	(0.1137)
ln(rainfall)		0.114*		0.236**		0.122
		(0.0585)		(0.1002)		(0.0853)
ln(tobacco price$_{t-1}$)		0.044		0.203*		0.159**
		(0.0986)		0.1221		(0.0704)
ln(maize price$_{t-1}$)		0.145*		-0.108		-0.253***
		(0.0750)		(0.1255)		(0.0913)
ln(production/area)$_{tb,t-1}$		0.044		0.268**		0.224***
		(0.0665)		(0.1179)		(0.0711)
ln(agglomeration index)		4.59		-9.03		-12.62
		(4.764)		(9.322)		(7.691)
number of observations	1032	1032	1032	1032	1032	1032
R^2	0.5334	0.5403	0.9480	0.9502	0.9636	0.9657
RMSE	0.29714	0.29581	0.51578	0.50637	0.42033	0.40883

Notes – The table reports estimates of average impact of distance to auction floor on tobacco area, production and yield. Estimations are based on annual data from 2003-04 to 2009-10 (seven years). All equations are estimated with OLS. Robust standard errors clustered by EPAs are given in parentheses (.) below the coefficient. All estimations include a constant term, EPA and year dummies ($dEPA_i$, $dYEAR_t$) and trends ($trend_k$) of which we do not report coefficients and standard errors. R^2 = coefficient of determination, and RMSE = Root Mean Squared Error. * means significant at the 10% level ($p < 0.10$), ** at the 5% level ($p < 0.05$), and *** at the 1% level ($p < 0.01$). Reported estimations are cleaned for outliers. See appendix E for the estimations results of Table 5 without outlier cleaning.

To a large extent covariates have expected signs and are statistically significant. Rainfall is statistically significant in the production and yield estimations, while lagged tobacco price and tobacco yield are statistically significant in the production an area equations. The agglomeration index performs poorly due to interaction with trend and EPA dummies: sign and significance improve if either of these or both are omitted. The positive coefficient of lagged maize prices in the area productivity equation (column 1) is difficult to explain. However, the key message of Table 5 concerns the impact variable. Coefficients of distance to auction floor are statistically significant and with the same sign as estimations in Table 4.

Estimations of Table 4 are concluded to be reasonably robust for inclusion of covariates as controls since the coefficients of the distance to auction floor variable in Table 5 remain reasonably near values estimated in Table 4.

Using generalised propensity scores in the estimation of dose response functions

In order to reduce the potential bias that arises with OLS estimation, we employed a generalised propensity score method (GPS), which is specifically developed to generate unbiased estimates of the population Average Treatment Effect (ATE) for the case of continuous treatment (see Hirano and Imbens, 2004)[21].

Table 6 Market access in tobacco: dose response estimations

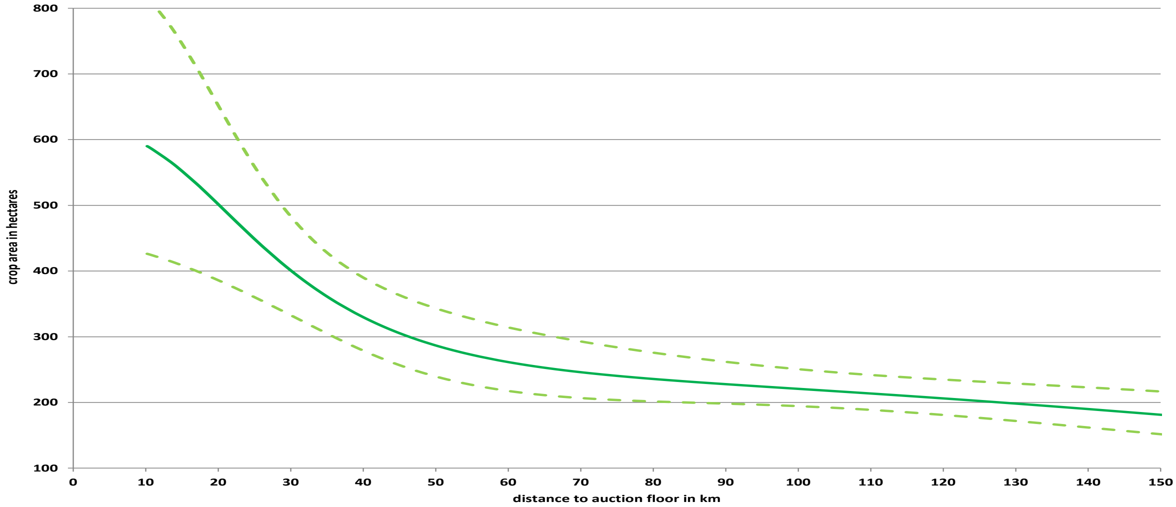
Dependent variable:	ln(production per hectare)	ln(production)	ln(area)
	(1)	(2)	(36)
ln(distance to auction floor)	0.0001	-0.007***	-.007***
	(0.0003)	(0.0014)	(0.0014)
generalized propensity score	-0.246***	-0.990***	-0.744**
(GPS)	(0.070)	(0.3591)	(0.3489)
number of observations	1038	1038	1038
F (.)	(2,1035)	(2,1035)	(2,1035)
	6.80	14.43	14.58
Prob > F	0.0012	0.0000	0.0000
Adjusted R^2	0.0111	0.0253	0.0255
RMSE	0.38879	2.0035	1.9464

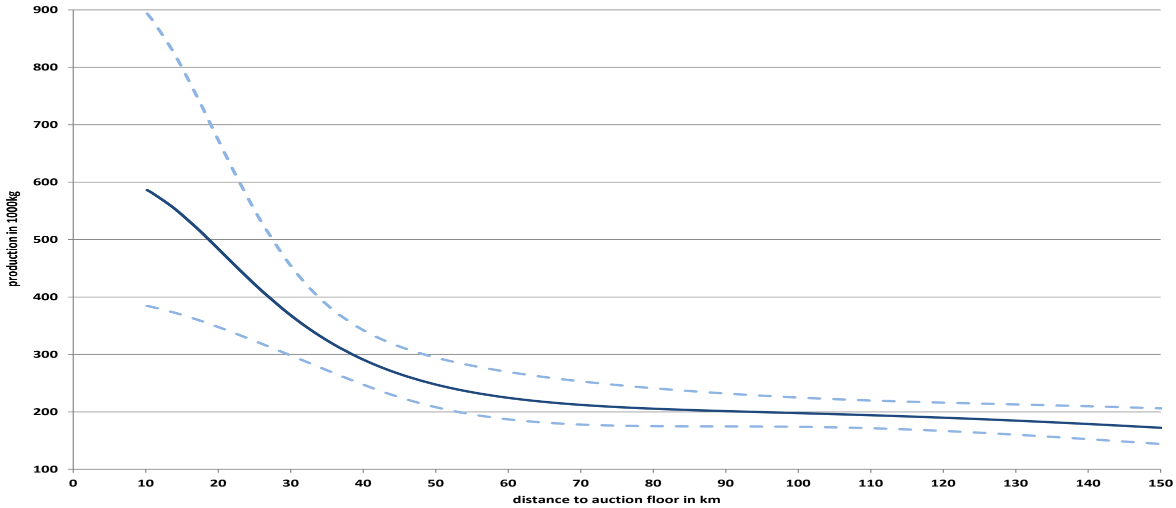
Notes – The table reports dose response estimations of the introduction of a tobacco auction floor on tobacco production per hectare, using generalized propensity scores, the continuous treatment variant of propensity scores. Covariates used in constructing the generalized propensity score are: rainfall, real lagged tobacco and maize price, lagged area productivity in maize and tobacco and the lagged share of maize in total crop area. The selection of variables underlying the estimation of the propensity score support the balancing property. Estimations are based on annual data from 2003-04 to 2009-10 (seven years). * means significant at the 10% level ($p < 0.10$), ** at the 5% level ($p < 0.05$), *** at the 1% level ($p < 0.01$).

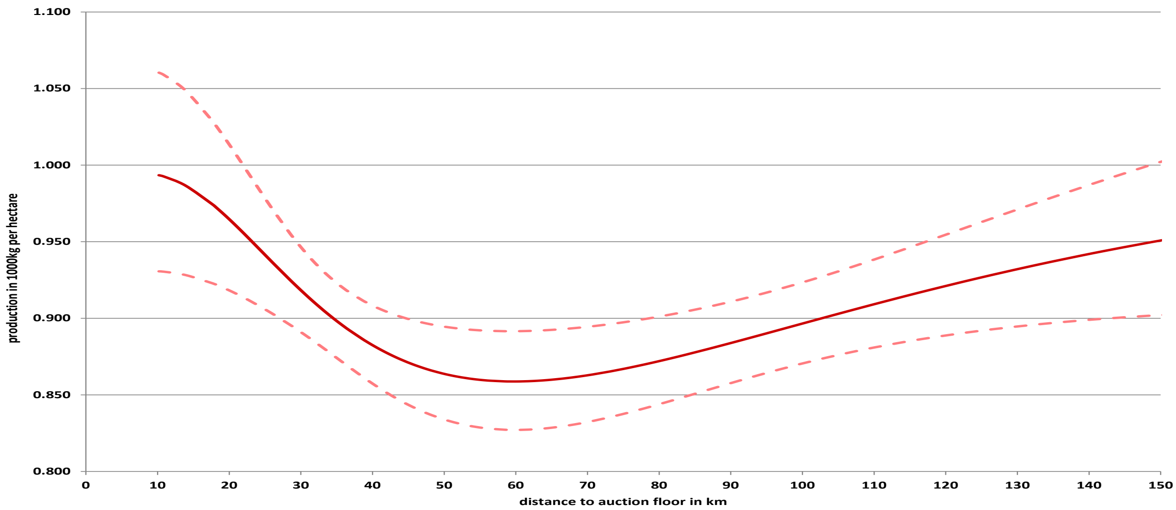
In the estimation of the GPS we have divided the intervention variable – distance to auction floor – into three intervals, equally sized in terms of observations. The balance of the covariates is investigated by testing whether the mean in one of the treatment groups differs from one of the others or both. For this purpose each interval is blocked into four blocks and

[21] We employed the STATA commands *gpscore* and *doseresponse*. Output reported in this section is based on these commands.

Figure 4 Dose-response function: impact of distance to auction floor on area, production and area productivity







GPS scores are compared. Covariates used in constructing the generalised propensity score are: rainfall, real lagged tobacco and maize price, lagged area productivity in maize and tobacco and the lagged share of maize in total crop area. The selection of variables underlying the estimation of the propensity score (and the selection of intervals and blocks within intervals) support the balancing property at the 20% level. Bootstrap methods are used to derive standard errors and confidence intervals, using 50 bootstrap replications.

The estimations, reported in Table 6, document the results of the estimation of the dose-response function based on the generalised propensity score. The interpretation of the coefficients in Table 6 is complicated due to reduced form nature of propensity scores and the use of a flexible specification of the dose-response function. However, the figures of the dose response function are easier to digest. These figures, shown in Figure 4, suggest that both crop area and production are negatively correlated with distance to auction floor: a higher distance to the auction floor is associated with lower crop area and production. Moreover, the figures suggest that crop area and production in EPAs with a short distance to auction floor (less than 60 km) have a much larger response than EPAs are farther away from auction floors. In fact, the figures indicate that a change in distance to auction floor beyond 60 km triggers very little response. Within a distance of 60 km from an auction floor, a 10 km reduction induce a 80 hectare increase (a 90 ton increase in tobacco production), while a similar reduction beyond 60 km only leads to a crop area increase of around 10 hectare (an increase in tobacco production of around 10 ton). The non-linear impact is due to the GPS method. The elasticities estimated with OLS correspond roughly with the elasticities of the dose response estimations (Figure 4), for distances to auction floor beyond 60km.

5. Alternative explanations and potential threats

Impact on other crops

The statistically significant impact on tobacco area, production and yield in the EPAs that are benefitting from the newly established auction may be a coincidental outcome that applies to all crops in these EPAs. For this reason we have repeated the impact estimations using area, production and yield of alternative crops, notably maize, groundnuts and pulses. Maize is the key food crop and produced by virtually all households. Maize accounts for more than 50% of total crop area and around 60% of the Malawi food consumption diet (MoAFS and FAO). Groundnut is (partly) a cash crop like tobacco but also a food crop. Groundnut cultivation has a country wide distribution similar to tobacco and groundnuts are also an important source of income for farm households, although less important than tobacco. Cultivation of pulses is less widely distributed. With the exception of maize, these alternative crops are, like tobacco, high value crops with per kg price of 4.5 to 8 times the price of maize. Since endogeneity of the intervention locations cannot be an issue for the alternative crops, we do not need to apply estimation methods that adjust for the associated bias and therefore we may estimate with OLS. The OLS results of these estimations are reported in Table 7 (and an extended version including covariates in Appendix F).

The table shows that the coefficients of distance to auction floor are small and statistically insignificant, or, in one case, only weakly significant. Hence, the estimation results do not support a systematic and statistically significant impact on area, production and yield, of maize, groundnuts and pulses. The hypothesis that the estimated impact on tobacco area, tobacco production and tobacco yield applies to other crops as well, is not confirmed by impact estimations for other crops. This result further confirms the statistically significant impact of improved market access – by the introduction of a new auction floor – on production per hectare, production and area of smallholder tobacco farmers.

Table 7 Impact estimations with placebo crops

Dependent variable:	ln(production per hectare)	ln(production)	ln(area)
Crop: maize	**(1)**	**(2)**	**(3)**
ln(distance to auction floor)	-0.076 (0.0635)	-0.037 (0.0653)	0.039 (0.0507)
Number of observations	1108	1108	1108
R^2	0.7997	0.8757	0.9452
RMSE	0.24426	0.29186	0.16229
Crop: groundnuts	**(1)**	**(2)**	**(3)**
ln(distance to auction floor)	-0.107* (0.0613)	-0.008 (0.0646)	0.098 (0.0611)
Number of observations	1108	1108	1108
R^2	0.6101	0.9150	0.9407
RMSE	0.35118	0.43817	0.30064
Crop: pulses	**(1)**	**(2)**	**(3)**
ln(distance to auction floor)	-0.0859 (0.1277)	0.0352 (0.2024)	0.121 (0.1090)
Number of observations	1108	1108	1108
R^2	0.6183	0.9348	0.9618
RMSE	0.27257	0.38579	0.28106

Notes – The table reports estimates of average impact of distance to auction floor on maize, groundnut and pulses production per hectare, production and area. Estimations are based on annual data from 2003-04 to 2009-10 (seven years). All equations are estimated with OLS. All estimations include a constant term, EPA and year dummies ($dEPA_i$, $dYEAR_t$) and trends ($trend_k$) of which we do not report coefficients and standard errors. Robust standard errors clustered by EPAs are given in parentheses (.) below the coefficient. R^2 = coefficient of determination and RMSE = Root Mean Squared Error. * means significant at the 10% level ($p < 0.10$), ** at the 5% level ($p < 0.05$), *** at the 1% level ($p < 0.01$).

Apart from showing a statistically insignificant impact, of negligible size, of distance to auction floor on alternative crops, the economics behind crop choices is interesting. Has the positive impact on tobacco crop area (and production) has given rise to a negative impact on crop area (and production) of alternative crops? Was crop area of other crops (maize, groundnuts, pulses, etc.) substituted into tobacco area? For this purpose, we employ the event study specification, shown at the start of this paper: this specification shows the average change in crop area (crop production) for each year in the EPAs that benefitted from the new auction floor and is more flexible for variation in responses between years and responses that are not related to distance to auction floor. The results of this exercise, again, implemented for maize, groundnuts and pulses (not shown here, available on request from the author), suggest no substitution of crop area between these crops and tobacco, in response to the introduction of the new auction floor.

We further corroborated this observation by estimating how increases in total crop area and decrease in crop area of alternative crops are associated with changes in tobacco area (see Appendix G). We find that increases in tobacco area are mostly associated with increases in total crop area, and not with decreases in area of alternative crops, and, hence, the evidence does not support substitution of area of other crops into tobacco area.

Trend developments

The measured impact could be the result of differences between the intervention and non intervention EPAs that existed before the intervention took place. Hence, we need to verify if the intervention and non-intervention EPAs were on a similar time path during the pre-intervention period. From the event study plots (see Section 2) and the grid estimations it is evident that the measured impact is delayed a few years. This gives us just sufficient years of observations to investigate if the intervention and non-intervention EPAs were on a similar time path during the pre-intervention period, from 2003-04 to 2005-2006.

Table 8 Comparison of variables of intervention and non–intervention locations in the pre-intervention period

	intervention	non-intervention	F test
crop area	0.024	0.026	F(1,173): 0.00
	(0.0453)	(0.0272)	0.9701
Production	0.140**	0.108***	F(1,173): 0.16
	(0.0717)	(0.0351)	0.6858
Yield	0.117**	0.082***	F(1,173): 0.47
	(0.0461)	(0.0190)	0.4918

Notes – The table reports coefficients of a trend of the group of intervention EPAs and the group of non-intervention EPAs, both during the pre-intervention period (2003/04 to 2005/6), an F test and its p-value on the significance of the difference. Estimated equation is: $lnY_{it} = \beta_0 + \sum_j \beta_{1j} I_j * trend + \varphi_i + \omega_t + \varepsilon_{it}$. Robust standard errors clustered by EPAs are given in parentheses (.) below the coefficient. * means significant at the 10% level ($p < 0.10$), ** at the 5% level ($p < 0.05$), *** at the 1% level ($p < 0.01$).

Table 8 shows test results on the coefficients of a trend of intervention and non-intervention EPAs, regressed on the outcome variables, all during the pre-intervention period. Trend developments in tobacco area, production and yield of both the intervention EPAs and the non-intervention EPAs are consistently shown to be similar in the pre-intervention period. Hence, we conclude that intervention and non-intervention EPAs are on a similar time path before the intervention.

Ceilings to expansion

Another issue concerns the presence of ceilings to expand: if all land suitable for tobacco cultivation is exhausted, there are no possibilities for further growth of tobacco production. EPAs that meet these conditions cannot be used as controls. Potential availability of crop area can be investigated: EPA data on crop area are available for all major crops and for this exercise we distinguish crop area for tobacco, maize and other crops, where “other crops” is an aggregate of rice, groundnuts, pulses, cassava, sweet potatoes, cotton, sorghum and millet. The data underscore the overwhelming importance of maize, with an average share of around 50% of total crop area. Note that the average tobacco share in total crop area is less than 5% (median less than 2.5%). Expansion of tobacco crop area is assumed to be realized either through expansion of total crop area or through substitution with other crops. We assume that the potential for expansion of total crop area may be approximated with the gap between the maximum total crop area realized over the years 2003/04 to 2009/10 and actual total crop area, both by EPA. Hence, $pa_{it,expansion} = MAX_t(AREA_{it}) - AREA_{it}$ where $pa_{it,expansion}$ is potential crop area by expansion and $AREA_{it}$ is crop area of all crops, both in EPA i and in period t.

Next, we assume that all crop area potentially available through substitution with crop area of other crops originates from the “other crops” sector and only to the extent that crop area allocated to other crops exceeds a minimum level, reflecting minimum requirements for food security. By restricting are available for potential substitution to a limited part of the “other crops” area, we acknowledge the importance of maize and other food crops on food security grounds. Hence, crop area potentially available through substitution with other crops is computed as the difference between current “other crop” area and the minimum area allocated to other crops during the period 2003/04 to 2009/10. Hence, we have $pa_{it,substitution} = area_{other\ crops,it} - MIN_t(area_{other\ crops,it})$ where $pa_{it,substition}$ is potential crop area by substitution and $area_{othercrops,it}$ is crop area allocated to other crops, both in EPA i and in period t).

The sum of potential expansion and substitution area ($pa_{it} = pa_{it,expansion} + pa_{it,substitution}$) expressed in terms of tobacco area ($pa_{it}/area_{tobacco,it}$), should be high in order not to be a restriction for growth of tobacco area. On the basis of this numerical exercise we find a few of the control EPAs to have potential expansion opportunities for tobacco cultivation less than 100% of existing tobacco area in the period 2005/06 to 2009/10 (EPAs in the districts Lilongwe (3), Machinga (1), Mzimba (1) and Rumphi (1). Even these EPAs have a minimum opportunity for expanding tobacco area of 20%. Hence, the average expansion opportunities of non-intervention EPAs, expressed in terms of existing tobacco area, are high and we should conclude that there are no effective restrictions in this respect[22].

Quality of the data

Researchers occasionally point at the poor quality of Malawi administrative data, mostly, however, in relation to maize production data. For this reason we have compared the EPA data from the Agro Economic Survey of the Ministry of Agriculture and Food Security (AES-MoAFS) – the data that we use for the empirical estimations – with the auction sales volume data from Tobacco Control Commission (TCC) and with tobacco information extracted from the Integrated Household Survey 2 (IHS-2) from the National Statistical Office (see Appendix B). Most discrepancies between tobacco data from different sources, however, have reasonable

[22] Strictly we should also analyse if availability of labour and tobacco cultivation expertise is a restriction to growth of tobacco production in the control EPAs. This is particularly relevant since tobacco cultivation is considered labour intensive. Unfortunately we are unable to implement such an analysis. Hence, we have implicitly assumed no restrictions on these grounds.

explanations[23]. However, a number of discrepancies merit further investigation. Since the EPAs in the districts with the largest discrepancies (Dowa, Lilongwe, Kasungu and Mchinji) partly belong to the group of EPAs that is likely to benefit from the newly introduced auction floor, this observation points at the possibility of having estimated a statistical artefact in the impact estimations, that reflects the data collection process rather than a real response of tobacco growers. To investigate if the estimated impacts are statistical artefacts, we have checked the robustness of the results by omitting data from these districts.

Table 9 Market access in tobacco: omitting observations of specific districts

Dependent variable:	**ln(production per hectare)**	**ln(production)**	**ln(area)**
District omitted: Kasungu	**(1)**	**(2)**	**(3)**
ln(distance to auction floor)	-0.142	0.273	0.416**
	(0.1100)	(0.2143)	(0.1757)
Number of observations	1185	1185	1185
R^2	0.5169	0.9393	0.9585
RMSE	0.31138	0.55193	0.44725
District omitted: Lilongwe	**(1)**	**(2)**	**(3)**
ln(distance to auction floor)	-0.111*	0.394***	0.505***
	(0.0570)	(0.1163)	(0.1162)
Number of observations	1094	1094	1094
R^2	0.5260	0.9397	0.9585
RMSE	0.32007	0.56099	0.45692
District omitted: Mzimba	**(1)**	**(2)**	**(3)**
ln(distance to auction floor)	-0.160**	0.158	0.318***
	(0.0645)	(0.1029)	(0.0928)
Number of observations	1095	1095	1095
R^2	0.5239	0.9449	0.9638
RMSE	0.32131	0.54211	0.43346
District omitted: Mchinji	**(1)**	**(2)**	**(3)**
ln(distance to auction floor)	-0.139**	0.236**	0.375***
	(0.0549)	(0.1066)	(0.0965)
Number of observations	1185	1185	1185
R^2	0.5266	0.9402	0.9588
RMSE	0.30581	0.55015	0.44849

Notes – The table reports estimates of average impact of distance to auction floor on tobacco production per hectare, production and area. The robustness of the basic estimation of Table 4 is tested by omitting observations from specific districts. Estimations are based on annual data from 2003-04 to 2009-10 (seven years). All equations are estimated with OLS. All estimations include a constant term, EPA and year dummies ($dEPA_i$, $dYEAR_t$) and trends ($trend_k$) of which we do not report coefficients and standard errors. Robust standard errors clustered by EPAs are given in parentheses (.) below the coefficient. R^2 = coefficient of determination and RMSE = Root Mean Squared Error.

[23] Explanations are, for example: the distinction between smallholders and estates, and burley tobacco and other tobacco's, storage by farmers and lags in sales, measurement errors in recording, (illegal) cross border trade, etc.

The results reported in Table 9 show that estimates of impacts do change (which is not surprising given the key role these districts play in the identification of the impact of distance to auction floor) but remain to a large extent statistically significant and of similar size. These robustness checks confirm that previous estimation of impacts results can be maintained.

6. Summary and conclusion

We have investigated the impact of improved market access for a typical developing country cash crop / export crop on the smallholder's decisions on cultivated area and production. For this purpose we have exploited tobacco area and production data for the period around the introduction of an additional tobacco auction floor in Malawi. Tobacco is the most important cash crop in Malawi, grown in all district of Malawi, exclusively sold on auction floors, and subsequently entirely exported. There are four tobacco auction floors (Limbe (close to Blantyre), Kanengo (close to Lilongwe), Mzuzu and Chinkhoma), of which the one in Chinkhoma has started operations in 2004. Estimations are based on annual data by Extension Planning Area (EPAs), 198 in total, covering the whole of Malawi, for a period of seven years, from 2003 to 2009. The estimation results support a statistically significant positive impact of the introduction of the new auction floor and the related decrease in transport costs, on tobacco area and production. As the increase in production is less than the increase in area, area productivity decreases. The decrease in area productivity arises most likely because tobacco cultivation expands to less suitable areas and/or less productive farmers. The impact of the introduction of the Chinkhoma auction floor is confirmed with generalised propensity matching techniques that are especially designed for the case of continuous treatment. Alternative explanations for the estimated impact (estimated impact applies to all crops, intervention and non-intervention on a different time path, restrictions to expansion in non-intervention locations and the measured impact is the result of poor quality of the data) could all be rejected.

Since the data used for estimations are not household data, we cannot identify subsistence households, and, hence, we cannot answer the question if access to market is going to help smallholders to move out of subsistence farming. However, the evidence does support a significant and sizable impact on tobacco area and production. Especially because the supply response is mainly on the extensive margin, it appears likely that lower transaction costs trigger smallholder farmers to shift to commercial agriculture. The bulk of the tobacco growing smallholders is poor or ultra-poor (around 65%, Economic Council (2000)). Also since 81% of all households in rural areas are classified as subsistence farmers (Integrated Household Survey-2, 2004-05) it is unlikely that the measured changes are fully outside the domain of subsistence farmers[24].

[24] Technically it is possible that the entire increase in yield, production and area is due to the 19% non-subsistence farm households. Only household or farm level data can offer a conclusive answer.

Acknowledgements

I would like to thank Eric Bartelsman, Hans Binswanger, Frank Bruinsma, Chris Elbers, Christopher Gilbert, Peter Lanjouw, Thea Nielsen, Menno Pradhan, Asger Moll Wingender, Bram Thuysbaert, and conference participants in Götenborg, Sweden (Nordic Conference in Development Economics, 2012), Dakar, Senegal (PEGNet, 2012), Université Dauphine, Paris, France (DIAL, 2013), Milan (ICAE, 2015), colleagues of VU-University Amsterdam and various anonymous journal reviewers for comments and suggestions on previous versions of this paper. Errors and omissions in the current paper are the responsibility of the author. I also would like to thank Hans Quene for skilful assistance in constructing the data and variables.

References

Aker, J.C., 2010, 'Information for Markets Near and Far: Mobile Phones and Agricultural Markets in Niger', *American Economic Journal: Applied Economics*, 2 (July), 46-59.

Balat, J., I.Brambilla, and G.Porto, 2009, 'Realizing the gains from trade: Export crops, marketing costs, and poverty', *Journal of International Economics*, 78, 1, 21-31.

Bia, M. and A.Mattei, 2008, 'A Stata Package for the Estimation of the Dose-Response Function through adjustment for the generalized Propensity Score', *The Stata Journal*, 8, 3, 354-373.

De Janvry, A., M.Fafchamps and E.Sadoulet, 1991, 'Peasant Household Behaviour with Missing Markets: Some Paradoxes Explained', *The Economic Journal*, 101, November, 1400-1417.

De Janvry, A. and R.Kanbur (eds.), 2006, *Poverty, Inequality and Development, Essays in Honor of Erik Thorbecke,* Kluwer Publishing.

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.

De Janvry, A. and E.Sadoulet, 2010, 'Agriculture for Development in Africa: Business-as Usual or New Departures', *Journal of African Economies*, 19, AERC, Sup. 2, ii7-ii39.

Diagne, A. and M. Zeller, 2001, Access to Credit and its Impact on Welfare in Malawi, Research Report 116, IFPRI, Washington.

Fafchamps, M., 1999, Rural poverty, Risk and Development, Rome, FAO.

Fafchamps, M., and R.Vargas Hill, 2005, 'Selling at the Farm-gate or Travelling to the Market, *American Journal of Agricultural Economics*, 87, 3, August, 717-734.

Food and Agricultural Organization of the United Nations (FAO), 2003, 'Issues in the Global Tobacco Economy: Selected case studies', Rome.

Goetz, S.J., 1992, 'A Selectivity Model of Household Food Marketing Behavior in Sub-

Saharan Africa’, *American Journal of Agricultural Economics*, 74, 444-452.

Goyal, A., 2010, ‘Information, Direct Access to Farmers and Rural Market Performance in Central India’, *American Economic Journal: Applied Economics*, 2 (July), 22-45.

Harrigan, J., 2008, ‘Food Insecurity, Poverty and the Malawian Starter Pack: Fresh Start or False Start?’, *Food Policy*, 33, 237-249.

Harrigan, J., 2003, ‘U-Turns and Full Circles: Two Decades of Agricultural Reform in Malawi 1981-2000’, *World Development*, 31, 5, 847-863.

Hirano, K. and G. Imbens, 2004, ‘The Propensity score with continuous treatment’, in Gelman A. and Xiao-Li Meng (eds.), *Applied Bayesian Modeling an Causal Inference from Incomplete-Data Perspectives*, John Wiley and Sons, Chapter 7, 73-84.

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

Jaffee, S., 2003, ‘Malawi’s Tobacco Sector: Standing on One Strong Leg is Better Than on None’, World Bank, Africa Region Working Paper Series, nr. 55.

Jayne, T.S, 1994, ‘Do High Food Marketing Costs Constrain Cash Crop Production? Evidence from Zimbabwe’, *Economic Development and Cultural Change*, 42, 2 January, 387-402.

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.

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.

Kydd, J. and R.Christiansen, 1982, ‘Structural Change in Malawi since Independence: Consequences of a Development Strategy Based on Large-Scale Agriculture’, *World Development*, 10, 5, 355-375.

Koester, U., G.Olney, C.Mataya, and T.Chidzanja, 2004, 'Status and Prospect of Malawi's Tobacco Industry: A Value Chain Analysis', report for The Emergency Drought Recovery Project, Ministry of Agriculture, Malawi.

Lea, N. and L.Hammer, 2009, 'Constraints to Growth in Malawi', The World Bank, Policy Research Working Paper 5097.

Negri, M and G.G.Porto, 2015, 'Burley Tobacco Clubs in Malawi: Nonmarket Institutions for Exports', *International Economics*, forthcoming.

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.

Orr, A., 2000, ' 'Green Gold'?: Burley Tobacco, Smallholder Agriculture, and Poverty Alleviation in Malawi', *World Development*, 28, 2, 347-363.

Otañez, M.G., H.Mamudu and S.A.Glantz, 2007, 'Global Leaf Companies Control the Tobacco Market in Malawi', *Tobacco Control*, 16, 261-269.

Poulton, C., J.Kydd and D.Kabame, 2007, 'All Africa Review of Experiences with Commercial Agriculture: Case Study on Malawi Tobacco', background paper for the Competitive Commercial Agriculture in Sub-Saharan Africa Study (CCAA study).

Renkow, M., D.Hallstrom and D.Karanja, 2004, 'Rural Infrastructure, Transaction Costs and Market Participation in Kenya', *Journal of Development Economics*, 73, 349-367.

Tchale, H. and J.Keyser, 2010, 'Quantitative Value Chain Analysis, An Application to Malawi', The World Bank, Policy Research Working Paper 5242.

World Bank, 2004, Diagnostic Trade Integration Study, Volume 2, Malawi, Chapter 4, Tobacco Sector.

Appendix A Data used, data sources and variable construction

Annual data of smallholder agricultural production and crop area on the level of Extension Planning Area's (EPAs), for the years from 2003/04 to 2009/10, are from the Agro-Economic Survey of the Ministry of Agriculture and Food Security (AES-MoAFS). All production and area data pertain to smallholders and exclude estates. An EPA reclassification in Salima and Nkothakotha district has made a number of before-reclasssification EPAs different from their equally named after-reclassification EPAs. Therefore, after reclassification observations – the shortest series – have been removed.

Distances between locations are calculated using standard Great Circle Method (for calculation see e.g. www.cpearson.com / Excel / LatLong.aspx). Data on latitude and longitude coordinates, required for this calculation, are from www.geocom.com, www.geonames.org, www.mapcrow.com and GOOGLE Earth. We have checked if road distance would generate different outcomes: road distances is between 20 and 40% higher compared to distance measured as the crow flies. However, with a few exceptions (e.g. locations that are connected through Lake Malawi) the differences are more or less proportional within acceptable error margins. Hence, for convenience we have maintained distance measured as the crow flies calculations.

Monthly farm gate prices for tobacco are observed for close to 50 locations, scattered over Malawi (Agro-Economic-Survey, Ministry of Agriculture and Food Security). However, these series are not complete. Around 43% of the tobacco farm gate price data used (annuals, seasonal averages) are directly taken from the data sources. The remaining observations are constructed by calculating a location specific (average) share of farm gate prices in national auction prices (Tobacco Control Commission (TCC)) and imputing these values to fill up the missing observations. Time series auction price data are unfortunately only available at the national level. Tobacco farm gate prices expressed as a share of the auction prices are 35.2% on average (median: 33.2%). This compares reasonably well with other sources (see Koester et al., 2004). All price data are attributed from physical market locations to EPAs on the basis of proximity. In some cases this involved averaging over various locations (triangulation). Data on farm gate maize prices are also from the Agro-Economic-Survey. These monthly series are available for close to 60 locations, but in contrast with burley tobacco farm gate prices the maize price series are much more complete: around 84% of the maize farm gate price data are directly taken from the data source. The remaining observations – the missings – are constructed using nearby prices. Like tobacco prices the maize price series are attributed to EPAs on the basis of the minimum distance of the geographical location of farm gate prices to the EPA. Groundnut prices are market prices – due to limited availability of farm gate prices and in contrast with tobacco and maize prices – and available for over 70 markets. All prices are deflated with the Malawi consumer price index for rural areas.

Annual data on rainfall in mm are from around 30 meteorological stations and supplied by the Department of Climate Change and Meteorological Services, Blantyre. Again we exploit the distance between meteorological stations and EPAs to find the rainfall series that is relevant for a specific EPA. The distance to the nearest weather station is, in most cases, less than 20km. If more than one weather station is nearest to an EPA we calculated the average between the nearest weather stations (triangulation).

Data on the number of households by EPA for one year (2007/08) are from the Ministry of Agriculture and Food Security. Combined with district data on average household size and population, we have constructed EPA population series for 2003/04 to 2009/10. Population by district data are census based and from the National Statistical Office (NSO). The EPA population series is used to construct EPA population density (EPA population in numbers by EPA land area in km2) or, alternatively, per capita area. EPA land area is constructed on the basis of a map of EPAs and made consistent with data on district area (source:

www.geohive.com). The size of EPAs in km2, pertains to land area (and hence excludes large lakes, like for example Lake Chilwa). The population density series varies both over time and between EPAs (but, naturally, the variation over time is limited). For the construction of an agglomeration index and distance to cities and towns we use a 1998 and 2008 listing of population of Malawi cities and towns, taken from National Statistical Office of Malawi.

For spatial integration of EPAs with the rest of the country we have constructed an agglomeration index. The agglomeration index is population size of city/town over distance to city/town, summed over all cities and towns, or in formula:

$$agglomeration\ index_{it} = \sum_{j} [population_{jt}\ /\ distance\ to\ city\ or\ town_{ij}]$$

$$where\ i = \ EPA, j = town\ or\ city\ \ and\ t = year$$

The larger the population of the city or town and the shorter the distance to the city or town, the higher the index. The agglomeration index reflects the degree of embedding of an EPA in the network of cities and towns. The higher this degree the lower transaction costs, and hence, we expect a positive relationship with tobacco production per hectare, production and area. It also represents the requirement to have access to market, since cash crop growers need to rely more on markets for their food purchases, for purchases of manufactured goods and for purchases of inputs (see de Janvry and Sadoulet, 2006)).

For descriptive purposes we use one year of auction transaction data (2009, a total of around 62,000 transactions), which were kindly made available by the Tobacco Control Commission (TCC). Each transaction contains information on type of tobacco, number of bales, volume in kg, value in US$, district of origin and club. Also for descriptive purposes we use annual aggregate times series data on the tobacco sector from TCC, posted on TCC website (www.tccmw.com).

Appendix B Comparison of tobacco data by source: AES-MoAFS, IHS (NSO) and TCC*

Figure B1 Aggregate annual production (AES-MoAFS) versus sales volume (TCC)

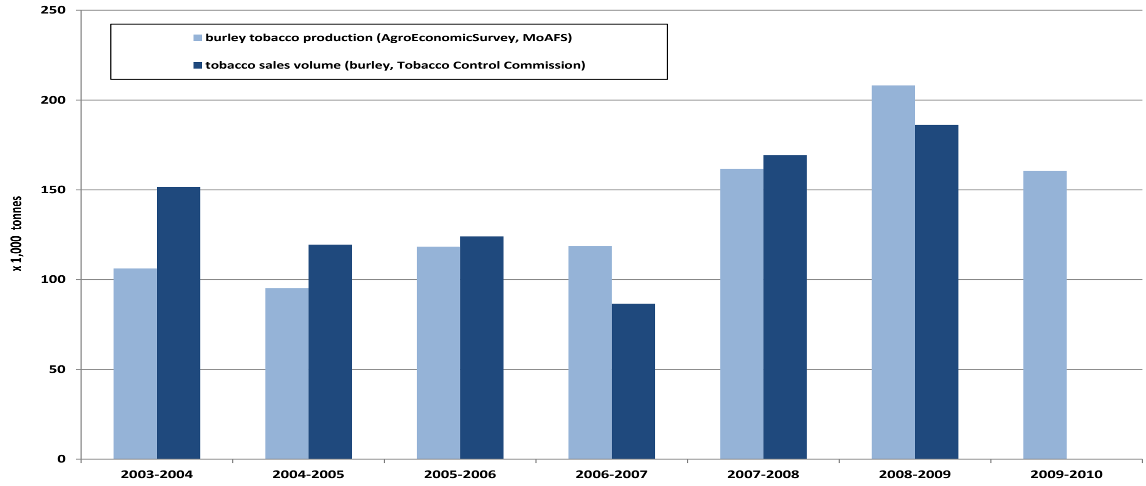


Figure B2 Production (AES-MoAFS) versus sales volume (TCC) by district, 2009

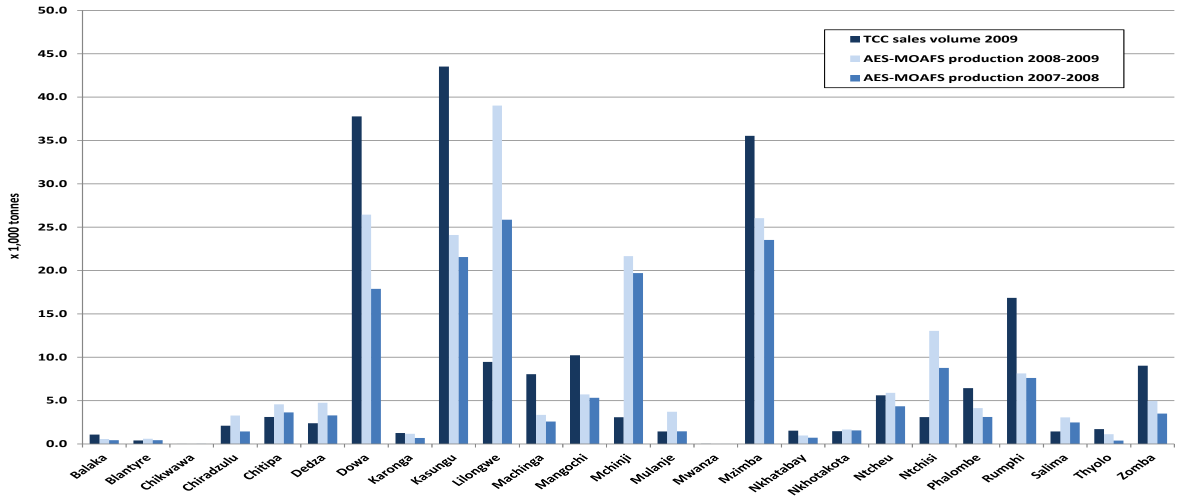
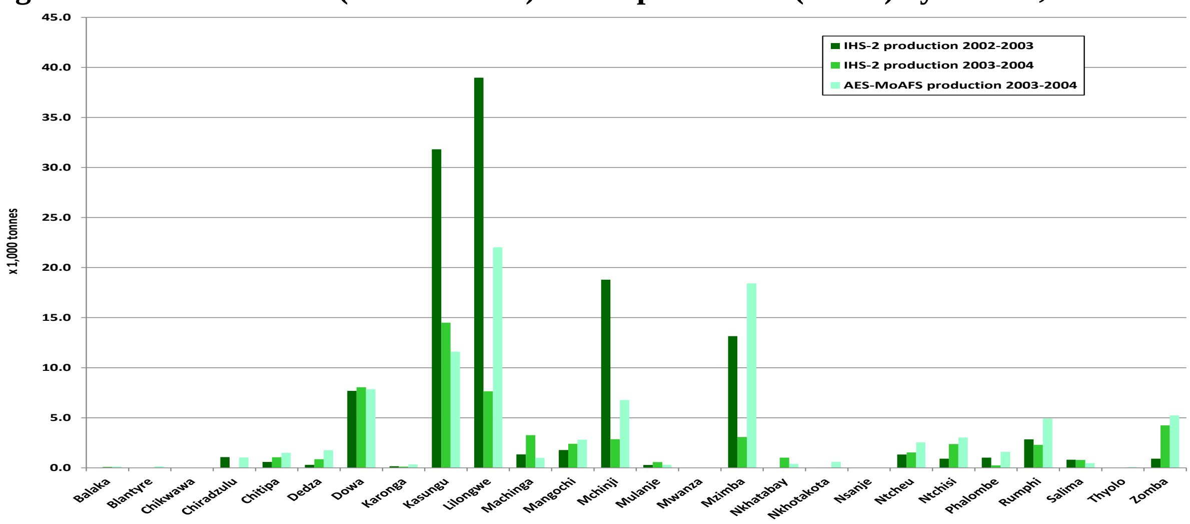


Figure B3 Production (AES-MoAFS) versus production (IHS-2) by district, 2003/04

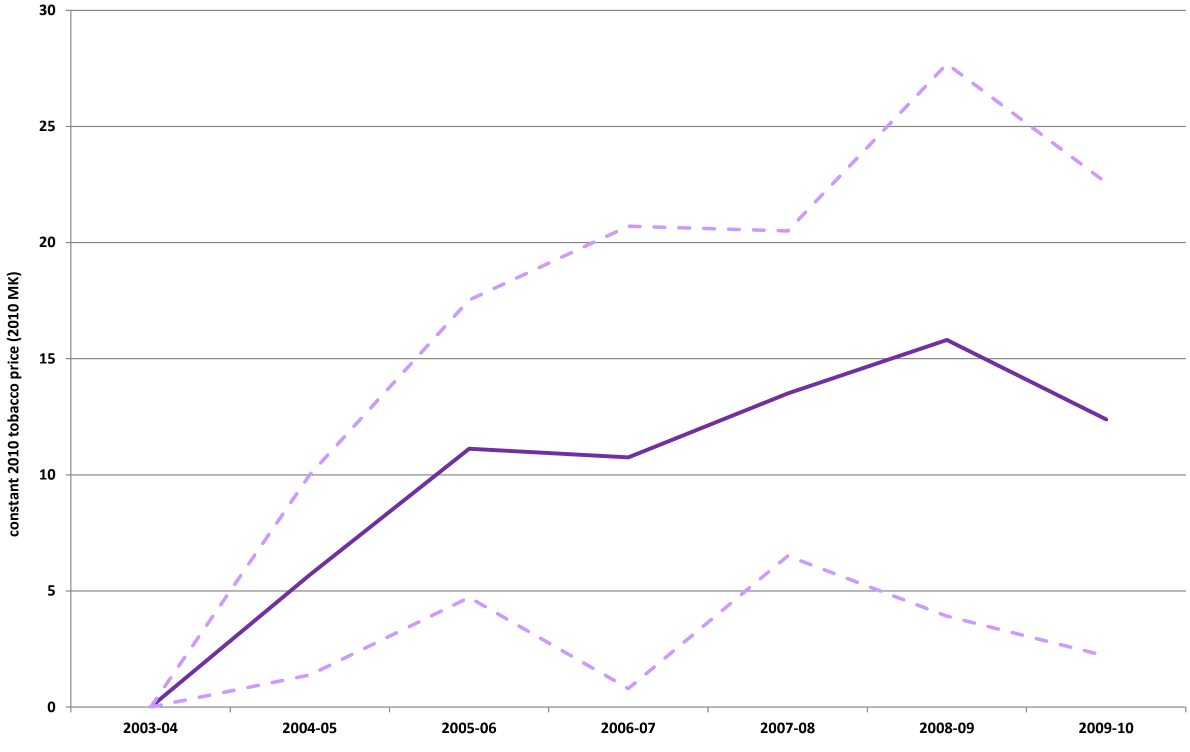


* AES-MoAFS: Agro-Economic Survey, Ministry of Agriculture and Food Security; TCC: Tobacco Control Commission; IHS: Integrated Household Survey; NSO: National Statistical Office.

Appendix C Event study plots

The Figure below shows the plot of the event estimation of tobacco farm gate prices and is similar to the plots in the main text (Figure 3). Likewise, we estimated $Y_{it} = \beta_0 + \sum_t \beta_{1t} I_{jt} + \sum_{km} \beta_{2km}\, trend_{kt}^{m} + \varphi_i + \omega_t + \varepsilon_{it}$ and plotted β_1 over t years. We are hesitant to put much trust in the plotted event figure, because data on tobacco farm gate prices are to large extent constructed (see Appendix A). But, apart from the wide spread, the plot is quite supportive for our investigations: It confirms statistically significant average higher tobacco farm gate prices due to the introduction of a new auction floor, and corresponds with around 20% higher farm gate prices.

Figure C1 Impact of lower transport costs on farm gate tobacco prices



Appendix D Varying lags in impact and alternative sets of intervention EPAs

Table D1 production per hectare by lag in impact and definition of intervention EPAs

production per hectare	**impact starts in (years delay):**			
definition of intervention EPAs	**2004-05 (0)**	**2005-06(1)**	**2006-07(2)**	**2007-08 (3)**
md1	-0.017 (.063); 0.522	-0.042 (.050); 0.523	-0.167*** (.057); 0.526	-0.145*** (.050); 0.525
md#	-0.032 (.066); 0.522	-0.049 (.051); 0.523	-0.180*** (.059); 0.526	-0.165*** (.051); 0.526
md$	-0.091 (.082); 0.523	-0.084 (.055); 0.523	-0.198*** (.061); 0.527	-0.179*** (.053); 0.526
md#+	-0.071 (.067); 0.523	-0.073 (.052); 0.523	-0.189*** (.061); 0.528	-0.185*** (.051); 0.527
md$+	-0.118 (.075); 0.523	-0.101* (.055); 0.524	-0.204***(.063); 0.528	-0.196*** (.052); 0.528
md$++	-0.199** (.100); 0.560	-0.152** (.063); 0.526	-0.211*** (.057); 0.531	-0.198*** (.048); 0.530

Note to table: Estimated specification: $lnY_{it} = \beta_0 + \beta_1 \ln(distance\ to\ auction\ floor_{it}) + \varphi_i + \omega_t + \varepsilon_{it}$, where Y_{it} is production, area and production per hectare of tobacco in EPA i in period t; φ_i and ω_t are EPA and year fixed effects and ε_{it} is an error term. Number of observations: 1239 (all observations); md1 = minimum distance to auction floor; md2 = second minimum distance to auction floor and laEPA = landarea EPA. The table report the coefficient of interest (β_1), its robust standard error and coefficient of determination (R2).

md# = md1 or md# = md1 – **1***laEPA$^{1/2}$ if landarea$^{1/2}$ > md2-md1;
md$ = md# or md$ = md# – **2***laEPA$^{1/2}$ if 2*laEPA$^{1/2}$ > md2-md1 > laEPA$^{1/2}$;
md#+ = md1 or md#+ = md1 – **2***laEPA$^{1/2}$ if laEPA$^{1/2}$ > md2-md1;
md$+ = md#+ or md$+ = md#+ – **2***laEPA$^{1/2}$ if 2*laEPA$^{1/2}$ > md2-md1 > laEPA$^{1/2}$;
md$++ = md#+ or md$++ = md#+ – **3***laEPA$^{1/2}$ if 2*laEPA$^{1/2}$ > md2-md1 > laEPA$^{1/2}$

Table D2 production by lag in impact and definition of intervention EPAs

production	**impact starts in (years delay):**			
definition of intervention EPAs	**2004-05 (0)**	**2005-06(1)**	**2006-07(2)**	**2007-08 (3)**
md1	0.180 (.128); 0.942	0.293** (.124); 0.942	0.158 (.116); 0.942	0.198** (.092); 0.942
md#	0.162 (.130); 0.942	0.283** (.125); 0.942	0.149 (.117); 0.942	0.186** (.093); 0.942
md$	0.100 (.138); 0.942	0.244** (.124); 0.942	0.131 (.116); 0.942	0.178 (.094); 0.942
md#+	0.079 (.126); 0.942	0.188 (.117); 0.942	0.083 (.113); 0.942	0.118 (.092); 0.942
md$+	0.029 (.132); 0.942	0.157 (.118); 0.942	0.069 (.113); 0.942	0.113 (.093); 0.942
md$++	-0.111(.148); 0.942	0.023 (.129); 0.942	-0.020 (.116); 0.942	0.038 (.101); 0.942

Note to table: See previous table

Table D3 crop area by lag in impact and definition of intervention EPAs

crop area	**impact starts in (years delay):**			
definition of intervention EPAs	**2004-05 (0)**	**2005-06(1)**	**2006-07(2)**	**2007-08 (3)**
md1	0.197** (.094); 0.959	0.335*** (.095); 0.960	0.325*** (.090); 0.960	0.343*** (.080); 0.960
md#	0.195** (.094); 0.960	0.332*** (.095); 0.960	0.329*** (.091); 0.960	0.351*** (.082); 0.960
md$	0.190** (.094); 0.959	0.328*** (.096); 0.960	0.329*** (.091); 0.960	0.357*** (.084); 0.960
md#+	0.149* (.084); 0.959	0.261*** (.087); 0.960	0.271*** (.079); 0.960	0.303*** (.071); 0.960
md$+	0.147* (.085); 0.959	0.259*** (.088); 0.960	0.273*** (.080); 0.960	0.309*** (.072); 0.960
md$++	0.088 (.085); 0.959	0.175* (.093); 0.959	0.190** (.088); 0.959	0.236*** (.080); 0.960

Note to table: See previous table

Table D4 production per hectare by lag in impact and definition of intervention EPAs (with trends)

production per hectare	impact starts in (years delay):			
definition of intervention EPAs	**2004-05 (0)**	**2005-06(1)**	**2006-07(2)**	**2007-08 (3)**
md1	0.010 (.067); 0.530	-0.016 (.049); 0.530	-0.148^{***} (.054); 0.533	-0.125^{***}(.048); 0.532
md$^{\#}$	-0.002 (.070); 0.530	-0.020 (.050); 0.530	-0.161^{***} (.056); 0.533	-0.145^{***}(.049); 0.533
md$^{\$}$	-0.061 (.082); 0.530	-0.056 (.051); 0.531	-0.181^{***}(.056); 0.534	-0.160^{***} (.050); 0.533
md$^{\#+}$	-0.039 (.069); 0.530	-0.042 (.051); 0.530	-0.171^{***} (.061); 0.534	-0.166^{***} (.051); 0.534
md$^{\$+}$	-0.087 (.075); 0.531	-0.072 (.051); 0.531	-0.188^{***}(.061); 0.535	-0.180^{***} (.051); 0.534
md$^{\$++}$	-0.175^{*} (.099); 0.533	-0.130 (.059); 0.533	-0.200^{***} (.053); 0.537	-0.186^{***} (.045); 0.536

Note to table: see table 1a; Estimated specification: $lnY_{it} = \beta_0 + \beta_1 \ln(distance\ to\ auction\ floor_{it}) + \sum_k \beta_{2k}\, trend_{kit} + \varphi_i + \omega_t + \varepsilon_{it}$ where $trend_k$ is a trend variable for region k

Table D5 production by lag in impact and definition of intervention EPAs (with trends)

production	impact starts in (years delay):			
definition of intervention EPAs	**2004-05 (0)**	**2005-06(1)**	**2006-07(2)**	**2007-08 (3)**
md1	0.133 (.135); 0.942	0.255^{**} (.123); 0.942	0.114 (.117); 0.942	0.156^{*}(.094); 0.942
md$^{\#}$	0.112 (.138); 0.942	0.243^{*} (.125); 0.942	0.101 (.119); 0.942	0.141 (.097); 0.942
md$^{\$}$	0.043 (.147); 0.942	0.199 (.124); 0.942	0.078 (.119); 0.942	0.129 (.098); 0.942
md$^{\#+}$	0.027 (.134); 0.942	0.147 (.120); 0.942	0.033 (.119); 0.942	0.072 (.100); 0.942
md$^{\$+}$	-0.028 (.140); 0.942	0.111 (.122); 0.942	0.014 (.120); 0.942	0.063 (.102); 0.942
md$^{\$++}$	-0.175 (.153); 0.942	-0.034 (.132); 0.942	-0.084 (.123); 0.942	-0.018 (.110); 0.942

Note to table: see table 1a; estimated specification as in Table 2a.

Table D6 crop area by lag in impact and definition of intervention EPAs (with trends)

crop area	impact starts in (years delay):			
definition of intervention EPAs	**2004-05 (0)**	**2005-06(1)**	**2006-07(2)**	**2007-08 (3)**
md1	0.123 (.098); 0.960	0.270^{***}(.098); 0.960	0.262^{***}(.094); 0.960	0.281^{***}(.082); 0.960
md$^{\#}$	0.114 (.100); 0.960	0.263^{***}(.100); 0.960	0.262^{***}(.097); 0.960	0.286^{***}(.087); 0.960
md$^{\$}$	0.104 (.100); 0.960	0.254^{**} (.100); 0.960	0.259^{***} (.097); 0.960	0.289^{***} (.089); 0.960
md$^{\#+}$	0.066 (.092); 0.960	0.189^{*} (.096); 0.960	0.204^{**} (.090); 0.960	0.238^{***} (.081); 0.960
md$^{\$+}$	0.059 (.093); 0.960	0.183 (.097); 0.960	0.203^{**} (.091); 0.960	0.243^{***} (.083); 0.960
md$^{\$++}$	0.000 (.092); 0.960	0.097 (.103); 0.960	0.116 (.100); 0.960	0.168^{*} (.092); 0.960

Note to table: see table 1a; estimated specification as in Table 2a.

Appendix E Robustness for inclusion of covariates: without cleaning for outliers

The estimation results in the table below repeat the estimations of Table 5 without cleaning for outliers. The results are similar to the results in the main text. However, alike the estimation results in Table 4, statistically significant impacts shift, to a certain degree, away from production and towards crop area. Measurement error and the impulse response figures make us confident that the results '*after cleaning for outliers*', presented in the main text are most trustworthy (Table 4 and 5).

Table E1 Market access in tobacco: OLS, fixed effects, trends and covariates

Dependent variables, all by Extension Planning Area (EPA):	ln(production per hectare)		ln(production)		ln(area)	
	(1)	(2)	(3)	(4)	(5)	(6)
ln(distance to auction floor)	-0.189***	-0.209***	0.048	0.100	0.237**	0.309***
	(0.0539)	(0.0540)	(0.1241)	(0.1265)	(0.0967)	(0.1059)
ln(rainfall)		0.128**		0.294***		0.166**
		(0.0555)		(0.0982)		(0.0836)
ln(tobacco price$_{t-1}$)		0.046		0.208*		0.162**
		(0.0975)		(0.1251)		(0.0745)
ln(maize price$_{t-1}$)		0.143*		-0.115		-0.258***
		0.0733		(0.1229)		(0.0895)
ln(production/area)$_{tb,t-1}$		0.047		0.271**		0.224***
		(0.0657)		(0.1159)		(0.0700)
ln(agglomeration index)		4.657		-4.247		-8.904
		(4.559)		(9.938)		(8.232)
number of observations	1046	1046	1046	1046	1046	1046
R^2	0.5335	0.5418	0.9474	0.9496	0.9637	0.9655
RMSE	0.29659	0.29482	0.52031	0.51084	0.42113	0.41191

Notes – The table reports estimates of average impact of distance to auction floor on tobacco production per hectare, production and area. Estimations are based on annual data from 2003-04 to 2009-10. All estimations include a constant term, EPA and year dummies ($dEPA_i$, $dYEAR_t$) and trends ($trend_k$) of which we do not report coefficients and standard errors. All equations are estimated with OLS. Robust standard errors clustered by EPAs are given in parentheses (.) below the coefficient. R^2 = coefficient of determination, and RMSE = Root Mean Squared Error. * means significant at the 10% level ($p < 0.10$), ** at the 5% level ($p < 0.05$), and *** at the 1% level ($p < 0.01$).

Appendix F

Table F1 Impact estimations with placebo crops

Dependent variable:	ln(production per hectare)		ln(production)		ln(area)	
Crop: maize	(1)	(2)	(3)	(4)	(5)	(6)
ln(distance to auction floor)	-0.076 (0.0635)	-0.083 (0.0617)	-0.037 (0.0653)	-0.061 (0.0664)	0.039 (0.0507)	0.022 (0.0521)
ln(rainfall)		0.281^{***} (0.0591)		0.286^{***} (0.0677)		0.0052 (0.0362)
ln(maize $price_{t-1}$)		0.106^{*} (0.0618)		0.147^{**} (0.0688)		0.041 (0.0397)
$\ln(\text{production/area})_{mz,t-1}$		0.030 (0.0456)		0.088^{*} (0.0518)		0.058^{**} (0.0255)
ln(tobacco $price_{t-1}$)		0.077 (0.0503)		0.049 (0.0658)		-0.029 (0.0350)
ln(agglomeration index)		-1.857 (2.429)		3.354 (3.192)		5.211^{**} (2.233)
Number of observations	1108	1108	1108	1108	1108	1108
R^2	0.7997	0.8105	0.8757	0.8810	0.9452	0.9463
RMSE	0.24426	0.23825	0.29186	0.28634	0.16229	0.16118
Crop: groundnuts	(1)	(2)	(3)	(4)	(5)	(6)
ln(distance to auction floor)	-0.107^{*} (0.0613)	-0.152^{***} (0.0579)	-0.008 (0.0646)	-0.052 (0.0693)	0.098 (0.0611)	0.101 (0.0620)
ln(rainfall)		0.452^{***} (0.0981)		0.222^{**} (0.0909)		-0.230^{***} (0.0609)
ln(groundnut $price_{t-1}$)		0.105 (0.0877)		0.038 (0.1076)		-0.067 (0.0762)
$\ln(\text{production/area})_{gn,t-1}$		0.229^{***} (0.0605)		0.203^{***} (0.0616)		-0.027 (0.0469)
ln(tobacco $price_{t-1}$)		0.089 (0.0543)		0.092 (0.0691)		0.004 (0.0484)
ln(agglomeration index)		6.717 (4.200)		11.06 (6.354)		4.343 (4.574)
Number of observations	1108	1108	1108	1108	1108	1108
R^2	0.6101	0.6428	0.9150	0.9175	0.9407	0.9420
RMSE	0.35118	0.33705	0.43817	0.4328	0.30064	0.29799
Crop: pulses	(1)	(2)	(3)	(4)	(5)	(6)
ln(distance to auction floor)	-0.086 (0.1277)	-0.121 (0.1145)	0.035 (0.2024)	0.001 (0.1875)	0.121 (0.1090)	0.122 (0.1091)
ln(rainfall)		0.189^{***} (0.0618)		0.143 (0.0919)		-0.046 (0.0691)
$\ln(\text{production/area})_{pl,t-1}$		0.102^{**} (0.0480)		0.159^{**} (0.0754)		0.057 (0.0441)
ln(tobacco $price_{t-1}$)		-0.031 (0.0467)		-0.004 (0.0620)		0.028 (0.0438)
ln(agglomeration index)		1.931 (2.876)		-0.893 (4.928)		-2.824 (3.958)
Number of observations	1108	1108	1108	1108	1108	1108
R^2	0.6183	0.6280	0.9348	0.9360	0.9618	0.9621
RMSE	0.27257	0.26970	0.38579	0.38315	0.28106	0.28086

Notes – The table reports estimates of average impact of distance to auction floor on maize, groundnut and pulses production per hectare, production and area. Estimations are based on annual data from 2003-04 to 2009-10. All estimations include a constant term, EPA and year dummies ($dEPA_i$, $dYEAR_t$) and trends ($trend_k$) of which we do not report coefficients and standard errors. All equations are estimated with OLS. Robust standard errors clustered by EPAs are given in parentheses (.) below the coefficient. * means significant at the 10% level ($p < 0.10$), ** at the 5% level ($p < 0.05$), and *** at the 1% level ($p < 0.01$).

Appendix G Finding evidence for substitution and expansion

Is the increase in tobacco area realized by substitution of area of other crops, or the result of area expansion? In order to investigate this we propose the following decomposition: Total crop area is the sum of tobacco area and non-tobacco area ($a_{all,t} = a_{tb,t} + a_{ntb,t}$) where non-tobacco area is an aggregate of all alternative crops including maize, groundnuts and pulses. We introduce λ, the share of tobacco area in total crop area, hence, $a_{tb,t} = a_{all,t} - (1- \lambda_t)a_{all,t}$. After lagging and differencing, we write the change in tobacco area as: $\Delta a_{tb,t} = a_{all,t} - a_{all,t-1} - (1-\lambda_t)a_{all,t} - (1-\lambda_{t-1})a_{all,t-1}$. If the share is constant ($\lambda_t=\lambda_{t-1}$) and all changes in tobacco area stem from a change in total crop area, than $\Delta a_{tb,t} = \lambda_t(a_{all,t} - a_{all,t-1})$. If total crop area is constant ($a_{all,t}=a_{all,t-1}$) and all changes in tobacco area stem from a change in the share of tobacco, than $\Delta a_{tb,t} = (\lambda_{t-1}-\lambda_t)\ a_{all,t}$. On the basis of this decomposition we regress the change in tobacco area on the share of tobacco area (λ_t) and total crop area ($a_{all,t}$) to find the average contribution of expansion and substitution as the coefficients of these regressors.

The estimation results for the intervention locations during the intervention years (column 4) indicates that the change in tobacco area correlates positively and statistically significant with the previous year share of tobacco area. Correlation with previous year total crop area is at most weakly significant (if we include EPA fixed effects). Hence, both estimations indicate that variation in tobacco area is dominated by variation in total crop area. The estimations do not support substitution of tobacco area with area of other crops.

Table G1 Substitution and expansion of tobacco crop area

dependent variable	change in tobacco area ($\Delta a_{tb,t}$)			
Explanatory variables: change in	(1)	(2)	(3)	(4)
total crop area ($a_{all\ crops,t}$)	-0.0003	0.0004	0.0016	-0.0014
	(0.0010)	(0.0080)	(0.0010)	(0.0035)
tobacco area share	-143.9	919.7	1167.1***	2995.6***
($a_{tb,t}/a_{all\ crops,t}$)	(423.3)	(597.4)	(365.7)	(738.9)
$dEPA_i$	no	no	no	no
crop years	2003/04 - 2005/06	2003/04 - 2005/06	2006/07 - 2009/10	2006/07 - 2009/10
(non) intervention EPAs	non-intervention	intervention	non-intervention	intervention
number of observations	294	77	592	157
R^2	0.0015	0.0014	0.0621	0.0698
RMSE	202.6	457.3	237.1	643.5
total crop area ($a_{all\ crops,t}$)	0.0256*	0.0638***	0.0030	0.0290*
	(0.0138)	(0.0196)	(0.0092)	(0.0158)
tobacco area share	12238***	17024**	12291***	24957***
($a_{tb,t}/a_{all\ crops,t}$)	(2955.0)	(6255.4)	(1209.7)	(4346.9)
$dEPA_i$	yes	yes	yes	yes
crop years	2003/04 - 2005/06	2003/04 - 2005/06	2006/07 - 2009/10	2006/07 - 2009/10
(non) intervention EPAs	non-intervention	intervention	non-intervention	intervention
number of observations	294	77	592	157
R^2	0.7816	0.7003	0.4792	0.5252
RMSE	134.2	361.5	204.6	544.0

Notes – The table reports estimates of (1) previous year total crop area and the share of tobacco area in total crop area, on tobacco crop area. Estimations are based on annual data from 2003-04 to 2009-10. All equations are estimated with OLS. Robust standard errors clustered by EPAs are given in parentheses (.) below the coefficient. R^2 = coefficient of determination, and RMSE = Root Mean Squared Error. * means significant at the 10% level ($p < 0.10$), ** at the 5% level ($p < 0.05$), *** at the 1% level ($p < 0.01$).

Appendix H Tobacco auction transactions

Table H1 Sales volume by tobacco auction floor* and by district of origin, 2009

	Sales volume (in 1000 kg)					per district distribution over auctions in %					per auction distribution over districts in %				
District	**MZZ**	**CNK**	**LIL**	**LMB**	**TOTAL**	**MZZ**	**CNK**	**LIL**	**LMB**	**TOTAL**	**MZZ**	**CNK**	**LIL**	**LMB**	**TOTAL**
Chitipa	2991	0.1	118	0.1	3110	96.2	0.0	3.8	0.0	100	4.3	0.0	0.2	0.0	1.5
Karonga	12288	16	24	4.0	1272	96.5	1.3	1.9	0.3	100	1.8	0.1	0.0	0.0	0.6
Rumphi	16383	469		4.0	16856	97.2	2.8		0.0	100	23.6	1.7		0.0	8.2
Nkhatabay	1530	3		2.7	1535	99.6	0.2		0.2	100	2.2	0.0		0.0	0.7
Mzimba	34913	550	34	46	35543	98.2	1.5	0.1	0.1	100	50.2	2.0	0.1	0.1	17.2
Nkhotakota	472	568	29	403	1472	32.1	38.6	1.9	27.4	100	0.7	2.0	0.1	0.7	0.7
Kasungu	9228	16737	16424	1139	43528	21.2	38.5	37.7	2.6	100	13.3	60.2	30.1	2.1	21.1
Ntchisi	415	1977	73	633	3098	13.4	63.8	2.4	20.4	100	0.6	7.1	0.1	1.2	1.5
Dowa	1361	5494	29337	1593	37786	3.6	14.5	77.6	4.2	100	2.0	19.8	53.7	2.9	18.3
Mchinji	400	678	421	1580	3079	13.0	22.0	13.7	51.3	100	0.6	2.4	0.8	2.9	1.5
Salima	54	45	4.9	1345	1450	3.7	3.1	0.3	92.8	100	0.1	0.2	0.0	2.5	0.7
Lilongwe	481	1199	6364	1428	9471	5.1	12.7	67.2	15.1	100	0.7	4.3	11.7	2.6	4.6
Dedza	4.1	22	1733	629	2388	0.2	0.9	72.6	26.3	100	0.0	0.1	3.2	1.1	1.2
Ntcheu	7.4	4.7	13	5583	5608	0.1	0.1	0.2	99.5	100	0.0	0.0	0.0	10.2	2.7
Mangochi	4.8	1.8		10225	10232	0.0	0.0		99.9	100	0.0	0.0		18.6	4.9
Machinga	2.4	1.5	2.2	8050	8056	0.0	0.0	0.0	99.9	100	0.0	0.0	0.0	14.7	3.9
Balaka		1.6	0.3	1080	1082		0.1	0.0	99.8	100		0.0	0.0	2.0	0.5
Zomba	10	2.1		9015	9027	0.1	0.0		99.9	100	0.0	0.0		16.4	4.4
Mwanza				48	48				100.0	100				0.1	0.0
Blantyre	5.9	3.0	3.3	390	402	1.5	0.8	0.8	96.9	100	0.0	0.0	0.0	0.7	0.2
Chiradzulu	13	11	6.7	2071	2102	0.6	0.5	0.3	98.6	100	0.0	0.0	0.0	3.8	1.0
Phalombe	0.4	3.2		6429	6433	0.0	0.0		99.9	100	0.0	0.0		11.7	3.1
Mulanje	0.7			1441	1442	0.1			99.9	100	0.0			2.6	0.7
Thyolo			0.0	1714	1714			0.0	100.0	100			0.0	3.1	0.8
Chikwawa	0.0	0.0	0.0	19	19	0.0	0.0	0.0	100.0	100				0.0	0.0
Total	69504	27788	54588	54871	206750	33.6	13.4	26.4	26.5	100	100	100	100	100	100

Source: transaction data for 2009 from the Tobacco Control Commission, Malawi (around 60,000 observations); * CNK = Chinkhoma; LIL = Lilongwe (Kanengo); LMB = Limbe; MZZ = Mzuzu; Notes – Districts of origin and auction floors, ordered from north to south. Regions (north, central and south) are distinguished by shading.