TI 2012-088/V
Tinbergen Institute Discussion Paper

# How does Market Access affect Smallholder Behavior? The Case of Tobacco Marketing in Malawi

*Wouter Zant*

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

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

Duisenberg school of finance is a collaboration of the Dutch financial sector and universities, with the ambition to support innovative research and offer top quality academic education in core areas of finance.

DSF research papers can be downloaded at: http://www.dsf.nl/

Duisenberg school of finance
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 8579

# How does Market Access affect Smallholder Behaviour?

# The Case of Tobacco Marketing in Malawi

Wouter Zant*

Version August 28, 2014

*Abstract*

Transaction costs play a key role in the behaviour of smallholders in developing countries. We exploit the introduction of an additional tobacco auction floor in Malawi to investigate the impact of a reduction in transaction costs and improved market access on production per hectare and the underlying smallholder's decisions on production and cultivated area of tobacco, the major cash crop in Malawi. Given the non-experimental nature of the data we use matching and potential outcome models to identify impact. Estimations are based on annual data by Extension Planning Area, 198 in total, fully covering Malawi, for the period 2003-04 to 2009-10. The estimation results support a statistically significant positive impact of the introduction of a new auction floor on smallholders' behaviour: production per hectare, production and area of tobacco has increased in the long run with respectively 20-25%, 36-38% and 15-21%. This outcome, and the increase in cultivated area in particular, suggests that lower transaction costs trigger smallholder farmers to shift to commercial agriculture.

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 5A-36, 1081 HV Amsterdam, The Netherlands; e-mail address: wouter.zant@vu.nl; tel: +31 20 598 9592; fax: +31 20 598 6127.

## 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)[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 one can overcome this subsistence trap? A possible way out of this trap is to reduce transaction costs for smallholders. Transaction costs – all types of costs incurred in order to sell agricultural output on the market – include costs of information, collection, loading and transport of goods, bargaining on prices and conditions, monitoring, insurance and other costs associated with access to market. It is often claimed that transaction costs are large and a major cause of not selling on the market. Conversely, (improved) access to markets – both the mere existence of markets but also the logistical and marketing infrastructure that facilitates agricultural crop sales on these markets – decrease transactions costs and should, thereby, potentially trigger smallholders to produce for the market.

To analyse this latter claim empirically we exploit the introduction of an additional tobacco auction floor in Malawi, on top of the three already existing auction floors. We quantify the size of the impact of reducing transaction costs on production per hectare in

---

[1] Food crops are not necessarily or exclusively used for subsistence, but may also be sold on the market.

[2] Promotion of either food crops or commercial crops also lies at the heart of policy discussions on economic growth and development (see e.g. Harrigan, 2003, 2008).

tobacco cultivation and the underlying smallholders' tobacco production and tobacco area decisions. Tobacco in Malawi is – with a distance – the major cash crop, grown throughout the country and by regulation exclusively sold on auctions. Malawi does not have a domestic cigarette industry and, hence, all tobacco is exported. In the 2000/01 season FAO estimates average transaction costs in Malawi tobacco to be in the range of 14.5% to 22.5% of sales value (see FAO, 2003). This estimate of transaction costs is likely to be a lower bound since many transaction costs are difficult to observe.

The paper is organised as follows. In Section 1 we review the literature on the role of market access and transaction costs in developing countries. In Section 2 we describe developments in the Malawi tobacco industry: we discuss 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. We also analyse a full year of auction transaction data. In Section 3 we show how we plan to 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 give the summary and conclusion of this paper.

## 1. Market access and transaction costs in the literature

Part of the literature that studies how transaction costs affect market participation and behaviour of farmers, is structural in nature and attempts to model the decision to grow either subsistence crops or cash crops, and the decision to participate in the market. What drives 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? De Janvry et al. (1991) discuss the implications of market failure in food markets, cash crop markets and labour markets, by assessing the effects of shocks against the benchmark of complete markets. Under market failures households have a tendency to get trapped into self-sufficiency, and the limited participation in the market also explains a sluggish supply response. Goetz (1992) develops a model where the decision of a food producing household is split up in a decision to enter the market and trade, and a decision how much to trade conditional on participating on the market as a buyer or seller. With a switching regression estimation, empirically applied to household survey data for Senegal, it is shown that improved market information for households significantly raises the probability of participation. Jayne (1994) argues that the wedge between producer prices of home produced maize and consumer prices of 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. A decrease of consumption prices with 5% to 30% is needed to make cash crop production attractive, using Zimbabwe survey data for 1990. Omamo (1998) uses a household model with transport costs to explain why farmers in the Kenyan Siaya District allocate larger shares of land to low yielding food crops rather than high return cash crops. Simulations incorporating transport costs generate observed food dominated cropping patterns as the optimal responses to high transport costs between farms and markets. Transport costs alone are sufficient for these results, and uncertainty and risk aversion is not required for this purpose. Key et al. (2000) extend and generalize the model proposed by Goetz (1992) by incorporating proportional and fixed transaction costs into an agricultural household model of supply response. Supply is shown to depend on proportional transaction costs and market participation on fixed transaction costs, which allows identification of both types of transaction costs. Empirical estimations, based on data of Mexican corn producers, indicate that both types of transaction costs matter, both for sellers and buyers. They find on the basis of their empirical work that 60% of the supply response to a price increase is due to producers who enter the market, while 40% is due to those producers who are already sellers on the market. Reduction of transaction costs is concluded to be an important complement in improving supply response. Using a similar model as developed by Key et al. (2000), Renkow et al. (2004) investigate – on the basis of data of rural households in Kenya – to what extent autarkic households need to be compensated with higher market prices to offset the fixed transaction costs that are made to either sell or buy on the market. They find that the ad valorem tax equivalent of fixed transactions costs is 15%. Public investment in reducing transaction costs is claimed to hold a greater potential for poverty alleviation than additional agricultural research.

Another line of recent research[3] makes use of impact evaluation techniques and is less concerned with developing structural models on transaction costs and estimation of marginal effects. This literature has focused on a specific type of transaction costs, notably information and search costs and aims to show the importance of information for the proper functioning of markets. The empirical studies focus for a large part on the introduction of mobile phone services, how this new information technology has impacted on access to and costs of information (search costs), on market prices and on economic behaviour. Often the introduction of mobile phone coverage is used for the identification of impacts, but also experimental designs are documented. Jensen (2007) makes use of micro level survey data to show that price dispersion on fish markets in Kerala, India has dramatically reduced after the introduction of mobile phones. This reduction in price dispersion is claimed to have established a nearly perfect adherence to the Law of One Price. The evidence further supports increased fishermen's profits and consumer welfare due to mobile phones. The gains from mobile phones services are not exclusively reserved for the wealthy but shared by smaller and poorer fishermen as well.

[3] The review of structural models is not exhaustive. For example, Cadot et al. (2006) approach the "commercial versus subsistence farming" issue from a different angle. They use a simple asset-return model of occupational choice in order to provide estimates of the cost of moving out of subsistence farming. The model is applied to data of Madagascar farmers. They find that the entry cost associated with moving out of subsistence are in the range of 124% to 153% of subsistence farmers' annual production. Several other contributions in this area offer useful insights of marketing behaviour of farmers, transaction costs, food prices and domestic trade in developing countries (e.g. Fafchamps et al., 2005; Fafchamps and Vargas Hill, 2005, and Minten and Kyle, 1999).

Easy and timely access to information is also shown to prevent waste, inefficiency and spoilage of production of perishable crops (Jensen, 2007; see also Muto and Yamano, 2009, on bananas). Aker (2010) uses market and trader level data to estimate the impact of mobile phones on price dispersion across grain markets in Niger. The empirical evidence on the introduction of mobile phone services in Niger between 2001 and 2006 supports a 10 to 16% reduction in price dispersion. The reduction in price dispersion is stronger for market pairs with higher transport costs, and larger once a critical mass of market pairs has mobile phone coverage. Reduction in search costs and inter market price dispersion is associated with improvements in trader and consumer welfare. The lower reduction in price dispersion compared to Jensen (2007) is attributed to better storability of grain and less perishability than fish. Along similar lines Muto and Yamano (2009) investigate the reduction in marketing costs of agricultural commodities due to the introduction of a mobile phone network in Uganda using household data for 2003 and 2005. They investigate marketing and trade of maize and bananas and find that the improved information due to mobile phone coverage has induced market participation of farmers in remote areas who produce bananas. Their study does not find impacts of mobile phone expansion on maize marketing. Mobile phone services cannot avoid asymmetric information between traders and farmers, that block potential benefits for farmers. Farmers' organizations are suggested to tackle this problem and to strengthen the bargaining power of farmers. Fafchamps and Minten (2012) estimate the benefits for farmers of SMS based agricultural information in Maharashtra, India, using a randomized controlled trial. The information includes prices, weather forecasts, crop advice and new items. They find no effect of this service on the prices received by farmers, value added, crop losses, crop choices and cultivation practices. These disappointing and somewhat disturbing results are in line with the limited commercial take-up of the information service, but difficult to reconcile with previous investigations on the impact of information (see above). A comparative advantage in transport is suggested as an explanation why benefits accrue in the first place to traders and not to producers. Finally, Goyal (2012) – offering an example of a change in marketing services rather than mobile phone services – investigates the impact of a change in marketing of a major private company in the soy market in the central Indian state of Madhya Pradesh. The company aimed at an improvement in procurement efficiency of soybeans to be achieved by the creation of a direct marketing channel (internet kiosks and warehouses) and by a reduction in transaction costs. After the introduction of kiosks and warehouses, soybean prices increased, price dispersion decreased and area under soy cultivation increased. This study highlights the benefits from direct interaction between producers and processors in agricultural marketing. For a full welfare assessment, however, the loss to traders needs to be quantified.

In summary, we find persuasive and rigorous evidence in the literature, both theoretical and empirical, of the key role that transaction costs play in explaining subsistence farming, on the impact of transaction costs on prices, arbitrage and economic behaviour, and on the potential welfare improvements of reductions in transaction costs. The current paper aims to contribute to this literature by investigating the impact of market access – caused by the introduction of a new auction floor – on the household decision to grow a cash crop. For this purpose we look at tobacco production in Malawi. Tobacco in Malawi is – with a distance – the major cash crop, grown in nearly all Malawi districts and exclusively sold through auctions. 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); see Appendix B for a map of Malawi with the locations of the auction floors). We exploit the introduction of this new auction floor: a comparison of supply response of those producers who did and those who did not benefit from this new auction floor provides a measure of the impact of market access. For the empirical measurement of impact we make use of aggregate annual area and production data of smallholders at Extension Planning Area level (EPA). There is a total of 198 EPAs, covering the whole of Malawi, for a period of seven years, from 2003 to 2009. Prior to elaborating the research methodology and showing the impact results we present an overview of the Malawi tobacco industry.

## 2. The Malawi tobacco industry

Various articles and publications describe market developments in the Malawi tobacco sector, the marketing and regulatory infrastructure that evolved over time and the (nearly complete) transformation that took place from estate based production – as a colonial heritage – to smallholder based production, since the end of the 1980s (see e.g. Kydd and Christiansen, 1982; Orr, 2000; Jaffee, 2003; World Bank, 2004; Poulton et al., 2007; Tchale and Keyser, 2010). Rather than making repeated references we note that we draw extensively on these sources to highlight the key developments and institutional changes which are relevant to the subsequent analysis. Complementary to this description we analyse the 2009 transaction data of all Malawi tobacco auctions.

*The role of tobacco in the domestic economy of Malawi*

Tobacco is, with a distance, the major 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 product (alternately tea or sugar) is only a fraction of tobacco exports. 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% – is poor or ultra poor (Economic Council (2000)). In the period from 2003 to 2010 smallholder crop area allocated to tobacco varies from 120,000 to 185,000 hectare, and smallholder crop production from 90 to 210 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) varies 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, export tax 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 government tax revenue (Jaffee (2003) reports 23%; FAO (2003) reports: "…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 its major driver of economic growth (see e.g. Lea and Hammer, 2009).

*Tobacco cultivation in Malawi: from colonial heritage to smallholder domination*

The Special Crops Act of 1964 continued pre-independence existing restrictions, that made the cultivation of tobacco the exclusive domain of estates. A new government – elected in 1993 and following an era of one party rule since independence – shifted the policy stance to a more broad based economic growth aiming at poverty alleviation. Under the new government – but also because of pressure by donors to liberalise the tobacco industry – amendments to the Special Crop Act were realised which allowed smallholders to grow burley tobacco (Jaffee,

[4] The estimate of direct employment in tobacco production and marketing depends both on the applied tobacco data (MoAFS, FAO), the labour supply data and how employment is related to area and production.

[5] The large share of tobacco proceeds that flows to the Government of Malawi makes the government a major stakeholder in the tobacco industry. In view of its large interest, it is an open question if the Government of Malawi is well suited to defend the interests of the tobacco industry. With at least partially exogenous world market prices a conflict of interest arises between farmers, buyers and the government. Farmers complain nearly each year about too low prices offered by colluding buyers at the auctions. A wide variety of rents due to lack of competition, lack of transparency and lack of accountability are observed (see e.g. Koester et al., 2004).

2003)[6]. In the course of the 1990s, these developments have given rise to a complete transformation from estate based tobacco cultivation with a high share of western type tobacco's (Flue Cured, NDDF, SDDF, Sunair), to a smallholder based tobacco cultivation with a high and dominant share of burley tobacco (see Figure 1). The changes in regulation laid the foundation for this transformation. High profitability of tobacco as a cash crop – the only really remunerative cash crop available to smallholders – and the broad spread of technical knowledge on tobacco cultivation – since many farmers worked previously on estates as labourers – triggered high growth of smallholder tobacco production. Simultaneously the more labour, capital and input intensive quality tobacco's, mainly grown by large estates, were quickly losing commercial viability because of high production costs (labour, fuel wood and capital). The increase of smallholder production was further supported by the formation of burley clubs, the introduction of intermediate buyers who provided the logistical link from farmers to auction floors[7] and the availability of credit to smallholders provided by the Malawi Rural Finance Company (Jaffee, 2003).

Since all tobacco exported from Malawi is required to be sold at auction, unit values and sales volume at auctions can be considered to be reasonable indicators of Malawi market prices and Malawi (aggregate) production[8]. Over the years, sales volume – broken down by burley tobacco, flue cured and other tobacco's – shows a nearly continuous upward development (see Figure 1). High growth rates in tobacco, however, are entirely on account of burley production. Especially, since the end of the 1980s burley sales realize high growth rates,

---

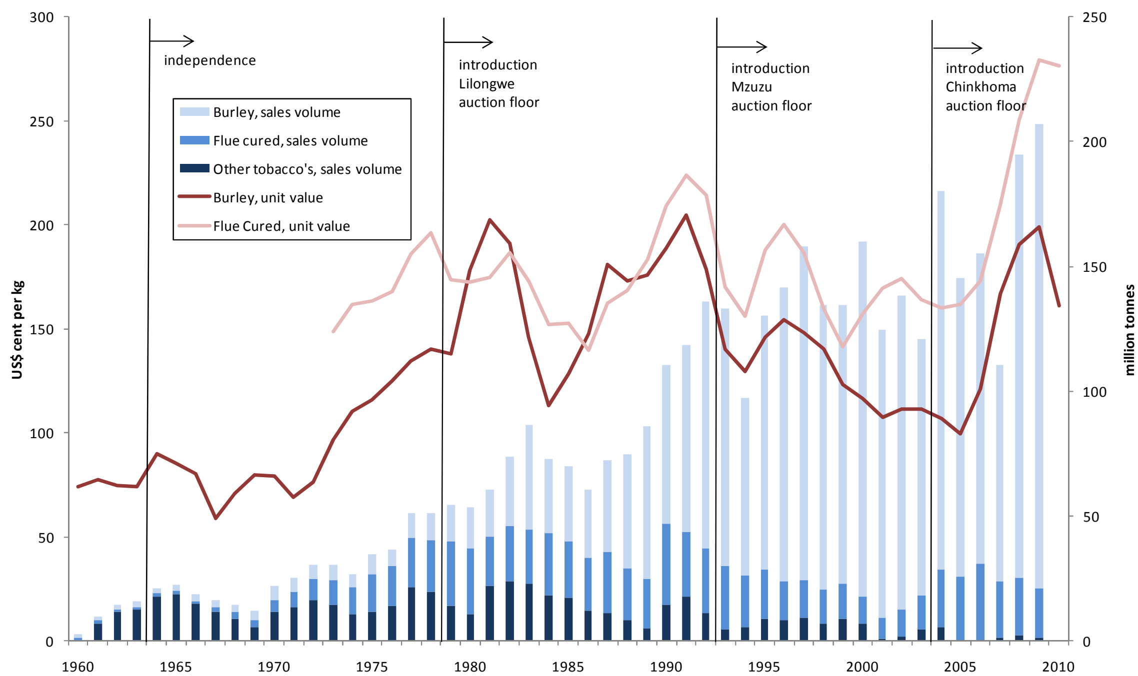
[6] Burley tobacco is a light air-cured tobacco used primarily for cigarette production. Flue-cured tobacco (also Virginia), NDDF and SDDF (respectively Northern and Southern Division Dark Fired) tobacco are other Malawi types of cigarette tobacco, that are smoke and fire dried and aged in curing barns.

[7] The minimum requirement for selling tobacco on the auction was overcome with the introduction of the intermediate buyer (FAO, 2003).

[8] Some authors report sales of tobacco from Zambia or Mozambique at Malawi auctions (see e.g. Koester et al., 2004). This is likely to occur occasionally but we assume that the size of these sales are modest and do not disturb the general message of the auction data for the Malawi tobacco industry. The empirical investigations in this study, however, are based on data from the Ministry of Agriculture and Food Security (Agro-Economic Survey).

which slightly level off by the end of the 1990s. Flue cured production is stagnant during the 1980s, declines during the 1990s to reach a production low in 2001, after which year a moderate recovery has taken place. Flue cured tobacco is nearly exclusively grown by estates. As a result of the higher capital requirements in the form of curing barns and other supporting equipment, smallholders are reluctant to take up the production of flue cured tobacco. Hence, it is safe to assume that the decline in sales volume of flue cured tobacco since 1990 should be attributed to the collapse of the estate sector[9].

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




[a] Notes – 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.

[9] It should be noted that the flue cured-burley divide does not run entirely parallel with the estate-smallholder divide as some estates are active in burley production.

Figure 1 further illustrates distinct periods of increases and decreases in burley auction prices. Flue cured auction prices are on average higher than burley auction prices but follow a similar development. 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. This observation is confirmed by other work in this area (see e.g. Jaffee, 2003) and also supported by our empirical estimations (see following sections).

*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 previous century. In the early days marketing institutions were adapted to production and marketing needs of estates. We focus on post-colonial times – Malawi gained independence in 1964 – and specifically the period since the start of 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, in the 1990s – in the course of liberalisation of the Malawi tobacco industry – a logistical infrastructure for tobacco transport and marketing from rural areas to auctions had to be put in place to service smallholder farmers. A variety of institutions and organisations came into being. 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 clubs were allocated a quota and were entitled to receive burley seed, fertilizer, advice on cultivation and extension support. From 1991/92 onwards clubs were authorized to sell directly on the auction floors and, since 1994, also to intermediate buyers (Orr, 2000). Access to auctions and thereby access to world market prices, credit facilities and economies of scale in transport are the major incentives for smallholders burley growers to join a burley club (Orr, 2000; Negri and Porto, 2008). Over the last decade (2000-2010) 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: Tobacco Control Commission). For accessing the auction floor, growers' clubs of at least 10 to 15 members can register at the Tobacco Control Commission for a fee and some administrative requirements (from the website of TCC, accessed in 2013). The relative size of the registration fee, depends on quotum, realized volume and realized auction prices: our calculations suggest that small farmers pay around 1% of sales value for registration at an auction floor. Additionally, the intermediate buyer system was introduced in 1994 to help smallholders to transport their burley tobacco to the auctions. The existing Tobacco Association of Malawi (TAMA) and the National Association of Smallholders Farmers of Malawi (NASFAM), which was established in the 1990s, also assisted in the organisation, collection, storage, transport and sale of smallholder tobacco from rural areas to the auction floors, in cooperation with burley clubs, estates and intermediate buyers (see Appendix C for more information on the major institutions of tobacco sector in Malawi). Shortcomings to this marketing infrastructure – which is continuously developing – were experienced in the area of widely diverging transport rates, storage losses and lack of accountability (see Jaffee, 2003).

Tobacco marketing is regulated by the Tobacco Control Commission (TCC), a government statutory body. TCC is responsible for market regulation and control, licensing of farmers, quality standards, and data and statistics of the tobacco sector. The TCC also advises 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 substantial 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 and well after independence 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 Appendix B for a map of Malawi with locations of auction floors and Figure 1 for the timing of the introduction of the auction floor in relation with aggregate tobacco sales volume and sales prices). The tobacco auctions normally open from mid-March and close towards the end of October. Tobacco is packaged in bales with a weight of 80-90kg. In the 2009 marketing season a total of 2.3 million bales were sold. A total of 25,000 to 30,000 bales are sold daily amongst the four auction floors. According to weekly reports from TCC[10] direct trade and contract trade is primarily important for specialty tobacco's (Flue Cured, NDDF and SDDF) and plays a minor role for burley tobacco. At least until 2010. Moreover, direct trade and contract trade transactions also need to be settled through auctions. Sellers that are not satisfied with the contract prices have the opportunity to switch to the auction market. The traditional auction system has four national sales daily: two in Lilongwe and one each in Limbe and Mzuzu. The Chinkhoma floor trades twice weekly on which days Lilongwe only has one auction sale. The sale of tobacco by auction is concluded to the highest bidder.

On the demand side there is a limited number of companies active, notably Limbe Leaf Tobacco Company, Dimon, Standard Commercial, Universal Leaf, Alliance One and Premium Tama. 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, the degree of 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

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

collusion between buyers at the auction floors (see e.g. Koester et al. 2004, Otañez at al., 2007)[11]. This is particularly manifest with occasional outbursts of protest from tobacco farmers who complain about the low prices at the auction.

*Auction transactions: comparison of district composition and unit values*

Auction transaction data make it possible to analyse the composition of sales volume and unit values, by auction floor and by district of origin (see Appendix E for an overview of sales volume by auction floor and by district of origin)[12]. In 2009, a share of 56.6% of all tobacco sales originates from the central districts Kasungu (21.1%) and Dowa (18.3%) and the northern district Mzimba (17.2%), comprising the three largest tobacco producing districts. A few districts in the south – Mwanza, Chikwawa and Nsanja – have negligible or no tobacco sales. In 2009 33.6% of total sales volume is traded on the Mzuzu auction floor, 13.4% on the Chinkhoma auction floor, 26.4% on the Lilongwe auction floor and 26.5% on the Limbe auction floor[13]. The Mzuzu auction floor in the north trades in particular tobacco from Mzimba (50%), Rumphi (20%) and Kasungu (13%); the Chinkhoma auction floor in the central region trades tobacco from Kasungu (60%), Dowa (20%) and Ntchisi (7%); the Lilongwe auction floor, also in the central region, trades tobacco from Dowa (54%) and Kasungu (30%); and finally the Limbe auction floor, in the south, trades tobacco from

[11] Another issue, and beyond the scope of this study, concerns the producer share of prices. How do farm gate prices compare with auction prices and export unit values? Our evidence suggests that farm gate prices are around 32% of auction prices, where the difference is explained by transaction costs and rents. Such (low) producer shares are not uncommon: e.g. for the former Zaire, Minten and Kyle (1999) observe similarly low producer shares and demonstrate on the basis of data from a traders' survey that transportation costs explain most of the differences in food prices between producer regions and that road quality is an major determinant of transportation costs. Further, auction prices are 40 to 70% of export unit values. The difference between auction prices and export unit values is explained by value addition from 'after auction' processing, export tax and also includes rents from buyers (and from the government).

[12] 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.

[13] This clearly diverges from the situation at the start of the 2000s when the Lilongwe auction floor at Kanengo was the largest (see Jaffee, 2003) and this relatively large throughput (close to 60%) is also confirmed by other sources (see Koester et al. 2004). The available 2009 TCC transaction data indicate that Lilongwe has lost substantial market share.

Mangochi (24%), Machinga (19%), Ntcheu (13%) and Phalombe (15%). To a large extent these figures reflect the relative importance of nearby production areas.

All tobacco growers in the north and the south – and this is no surprise – transport their tobacco almost exclusively to one single auction floor: the Mzuzu auction floor for the northern districts and the Limbe auction floor for the southern districts. Tobacco from more centrally located districts is sold on two or even three auction floors, with the exception of Salima and Ntcheu. Preferred floor of sale is not only determined by distance: for example, 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 Table 1 and Appendix E). This is likely to be explained by congestion at the Lilongwe auction floor, cheap transport alternatives from Mchinji and Salima (transport by train[14]) 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 are, however, negligible in 2009. This may be caused by the (still) moderate trading volume at the Chinkhoma auction floor and the – on average – higher prices at the Mzuzu auction floor.

Our primary interest is the districts of origin of tobacco auctioned on the Chinkhoma floor. We are also interested if tobacco is either exclusively sold on the Chinkhoma floor or combined with sales on other floors. Hence, for these districts we have calculated average unit values of burley tobacco for the different auctions (see Table 1). The table indicates that the highest average prices for burley are realized on the Mzuzu auction floor, while the

[14] The Malawi rail network consists of a rail line running from Zambia in the west, via Lilongwe to the south where it splits into a line further south to Blantyre and Beira in Mozambique (the Beira corridor), and a line to the east, to Nacala in Mozambique (the Nacala corridor). Both Mchinji, Salima and Ntcheu have railway stations along the 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 – substantial quantities of tobacco are transported from rural areas to the Lilongwe and Blantyre auction floors, both located along the railway line.

lowest average prices are realized on the Lilongwe auction floor. From the perspective of 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]

| district \ auction floor | 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. Figures are omitted in the case of less than 100 transactions.

## 3. Methodology

*How do transaction costs influence farmers' behaviour?*

We investigate the impact of (a reduction in) transaction costs on production per hectare, crop area and crop production of tobacco smallholders in Malawi that occurs with the introduction of a new auction floor. Farmers' area and production decisions are both driven by expected profits: farmers are assumed to maximize total expected profits. Total farmers' profits at time t in location i can be written as the sum of profits over all activities j:

$$\Pi_{it} = \Sigma_j \pi_{jit} \tag{1}$$

Without loss of generality, profits of each activity can be written as revenue minus costs, or market price minus unit production costs, and minus unit transaction costs, times output:

$$\pi_{jit} = [p_{jit} - c_{jit,production} - c_{jit,transaction}] \cdot q_{jit} \tag{2}$$

where $p, c_{production}, c_{transaction}$ $and$ $q$ are respectively output price, unit production cost, unit transaction cost and output. Under the assumption that cost and prices of all activities other than tobacco do not change, farmers' maximization of expected profits will simultaneously determine tobacco area, tobacco production and the preferred auction floor for selling tobacco[15]. The introduction of a new auction floor improves markets access and leads to a reduction of transaction costs for some tobacco growers. We assume that this reduction is due, in the first place, to a reduction in transport costs and transport costs are postulated to be a quadratic function of distance. Hence, we have:

$$c_{jit,transaction} = \gamma_0 + \gamma_1 \tau_{ji} + \gamma_2 \tau_{ji}^2 \tag{3}$$

Where $\tau$ is the distance to the nearest auction floor. The size of the impact of changes in transaction costs on total profits hence depends on distance:

$$\partial \Pi_{it} / \partial c_{jit,transaction} = -[\gamma_1 + 2.\gamma_2 \tau_{ji}] \tag{4}$$

A new auction floor that reduces the distance to the nearest auction, will lead to a reduction in transaction costs and to an increase of profits, following equation (4). The resulting increase in expected profits will trigger increases in tobacco production and larger allocations of area to tobacco cultivation.

*Estimating Impact of Interventions with Panel Data*

The impact of an intervention is estimated using a fairly standard and straightforward framework for impact evaluation with panel data. A reasonably general specification to estimate the dynamic impact of an intervention with a panel of observations runs as follows:

---

[15] The specification of profits is sufficiently general to incorporate differential auction prices, congestion costs and alternative modes of transport. All these variables are also important in the context of Malawi tobacco: see Table 1 for the extent of differential auction prices; the Kanengo auction floor has a reputation for high congestion costs and tobacco transport by train is a real alternative for locations along the Malawi railway line. These costs and prices will simultaneously determine farmer's tobacco area, tobacco production and preferred auction floor. Unfortunately data are lacking to estimate a complete structural model. Therefore we simplify the exercise by restricting transaction costs to transport costs only, with a single mode of transport (truck transport) and assuming no differences in prices and costs on different floors (or at least negligible relative to transaction costs).

$$Y_{it} = \beta_0 + \beta_1 I_{it} + \sum_k \beta_{2k} X_{k,it} + \varphi_i + \omega_t \tag{5}$$

where $Y_{it}$ is the outcome indicator, $I_{it}$ is a binary intervention variable with a unit value in case of an intervention and zero elsewhere (hence, zero before the intervention period for all i, and during the intervention period for those i's without intervention), $X_{k,it}$ is a set of k exogenous covariates, and $\varphi_i$ and $\omega_t$ are not observed time invariant and location invariant effects. The coefficient of the intervention variable ($\beta_1$) is the parameter of interest. We may allow impacts to vary over time, thereby increasing the flexibility of the specification and the informational content of the data. Covariates ($X_{k,it}$) may also contain time-invariant or location invariant variables.

*Impact of market access to the Malawi tobacco sector: empirical specification*

We implement the framework set out in the previous section to estimate the impact of the introduction of a new auction floor in Malawi (the intervention) on production per hectare, production and area of tobacco. The exercise aims to measure if and to what extent this new auction floor has given rise to changes in production per hectare, and if these changes are due to changes in production (the intensive margin), changes in area (the extensive margin) or both. In other words, the exercise aims to measure the supply effect of market access. In order to transform equation (6) to an equation that can be estimated empirically, we simply insert tobacco production per hectare, tobacco area and tobacco production as outcome indicator and add error terms:

$$y_{it}/y_{i0} = \alpha_0 + \alpha_1 I_{it} + \sum_k \alpha_{2k} X_{k,it} + \theta_i + \vartheta_t + \varepsilon_{it} \tag{7a}$$

$$q_{it}/q_{i0} = \gamma_0 + \gamma_1 I_{it} + \sum_k \gamma_{2k} X_{k,it} + \varphi_i + \omega_t + \mu_{it} \tag{7b}$$

$$a_{it}/a_{i0} = \delta_0 + \delta_1 I_{it} + \sum_k \delta_{2k} X_{k,it} + \rho_i + \tau_t + \nu_{it} \tag{7c}$$

Where $y_{it}$, $q_{it}$ and $a_{it}$ are tobacco production per hectare, production and area of EPA i in year t, $y_{i0}$, $q_{i0}$ and $a_{i0}$ are base period tobacco production per hectare, base period tobacco production and base period area cultivated with tobacco of EPA i, and $\varepsilon_{it}$, $\mu_{it}$ and $\nu_{it}$ are error terms with zero mean and constant variance. Production per hectare is in kilogram per hectare and can directly be compared between different EPAs. However, tobacco production is in kilogram and crop area in hectare, and both cannot directly be compared between different EPAs and years. Therefore, to make variables comparable between EPAs and years, we have expressed all dependent variables relative to their pre-intervention or base period level[16,17]. The difference-in-difference approach is not affected be this procedure: strictly, the EPA fixed effects ($\theta_i$, $\varphi_i$ and $\rho_i$) should cancel out (but need to be maintained if covariates ($X_{k,it}$) are not expressed relative to their base levels). The specification also enables to incorporate EPA fixed effects in the estimation of matching and potential outcome models. Outcome variables and covariates are transformed into natural logarithms. We use the specifications of equation (7a) to (7c) as the basic specification of our estimations, with restrictions on coefficients which are apparent from the tables.

Covariates ($X_{k,it}$) for both production per hectare, production and area primarily derive from profit maximizing behaviour of farmers. Expected profits are determined by production per hectare, and expected input and output prices. In rain-fed agriculture production and production per hectare of tobacco cultivation are, in the first place, determined by climatic variables, in particular rainfall. Output in tobacco cultivation depends strongly on rainfall. Both current rainfall and (cumulative) rains in the past contribute to growth of tobacco crops: in the empirical work we use current rainfall relative to long run averages (20-30 year averages) and a

[16] We have run alternative estimations of tobacco production and area normalized with the EPA population involved in tobacco cultivation. The latter is approximated with EPA population times district level share of households active in tobacco cultivation, extracted from IHS-2 (2003/04) and IHS-3 (2008/09). The impact estimations support a significant impact in both production and area (available on request). However, the size of the coefficient is large and difficult to explain, most likely due to measurement error in the normalization (see also the comparison between IHS and MoAFS data. Hence, and instead, we use a specification with current production relative to base production (or current area relative to base area) as independent variable.

[17] We have used the crop years 2003-2004 and 2004-2005 and the combination of these years as base period. Estimation results are robust for different specifications of the base period.

(weighted) sum of past levels of rainfall. Fertilizer inputs used in the cultivation of tobacco are likely to be important: agronomic evidence confirms the importance of NPK applications in tobacco cultivation. Next, both tobacco production and area are likely to respond to expected prices (see e.g. Figure 1 in this study, but also Jaffee, 2003). We used previous year farm gate tobacco prices for this purpose, since these prices – in contrast with tobacco auction prices – exclude transaction costs and drive smallholder decisions. We have also included previous year maize farm gate prices, the major alternative crop, since it is likely that tobacco production and area decisions are also influenced by alternative cultivation options. Then we employ a number of spatial variables to approximate transaction costs and agglomeration effects: For within-EPA transaction costs and agglomeration effects we use population density, for outside-EPA transaction costs and agglomeration effects we use two measures of distance to city or town and a distance weighted city and town size index. Data sources and data construction is extensively explained in Appendix A.

Finally we have included EPA and year fixed effects that represent a variety of unobserved influences, which are constant within EPAs and years, but varying between EPAs and years. In the context of the tobacco industry some of these fixed effects are clearly identifiable: growers that use the Kanengo auction floor in Lilongwe face large costs due to congestion. Also, some tobacco growers have the opportunity to transport their output to auction floors by rail (see previous section), since they are located near to the railway line. Some tobacco growers may have different opportunities because of proximity to a large city or proximity to a neighbouring country. Also presence of buyers' clubs or extension workers and other infrastructure or institution will contain a large EPA fixed effect component. Of course, there is also a wide range of general, not observed fixed effects outside the domain of tobacco cultivation that may, however, have an impact (e.g. education, religion, ethnicity, language). Also various influences are captured by year fixed effects, for example changes in general economic policy (agricultural subsidy programs), general weather conditions, broad market trends, international trade and exchange rate issues, etc.

*Intervention locations*

Tobacco farmers that benefit from the introduction of the new auction floor in Chinkhoma in 2004 (the intervention group) are identified by determining the minimum of the distances from each location to the different auction floors. For locations, for which this new auction floor has become the closest auction floor, the binary intervention variable is attributed the value of 1, while it is zero elsewhere. Practically this implies that all locations 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 F). Adhering to the rule that “the Chinkhoma auction floor has become the closest auction floor” for the identification for intervention locations, is apparently too strict. We assume that this is caused, at least partially, by inaccuracies in the measurement of distance (see *data for estimations* paragraph). Therefore we further consider those locations where 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 auction floor to each EPA. Potential measurement error in minimum distance to auction is approximated on the basis of (the root of the) EPA land area. On these grounds we have identified another 17 locations, summing to a number of 48 intervention locations (24%), out of a total of 198 locations, that potentially benefit from the newly established auction floor.

*Using matching models and potential outcome models 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 several alternatives and investigated the optimal location for doing this investment, basing its eventual choice on an assessment of 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, estimations are likely to be biased if this problem is not appropriately addressed. We propose matching models and potential outcome models to overcome this problem[18].

We briefly highlight the intuition of the estimation methods. The impact estimation methods either model both the impact variable and the treatment variable (inverse probability weighted regression adjustment (ipwra) and augmented inverse probability weighing (aipw)), compare observations that are similar (nearest neighbor matching (nn)), or use the estimated probability of treatment or propensity scores (propensity score matching ps). Both ipwra and aipw are best explained by starting with regression adjustment (ra) and inverse probability weighing (ipw). Under de regression adjustment method (ra) the outcome variable is regressed on a number of covariates for the treatment observations and for the control observation. Regression outcomes are used to predict potential outcomes (counterfactuals) and these predictions are used to estimate the population average of the treatment effect. Inverse probability weighing (ipw) fits a model on the probability of treatment using whatever characteristic that is available for all observations. This model is used to construct weights. For the (non)treated the weight is equal to the reciprocal of the predicted probability

[18] Practically we exploit the possibilities offered by the STATA command for treatment-effects estimation for observational data, *teffects.*

of (not receiving) treatment, where the probability of not receiving treatment is simply one minus the probability of treatment. Predicted probabilities close to zero or one make this technique unstable (and this corresponds to the requirement that every subject in the sample needs to have a non-zero probability to be treated). The outcome modeling strategy of ra and the treatment modeling strategy of ipw are combined in inverse probability weighted regression adjustment (ipwra) and the augmented inverse probability weighting (aipw). In the ipwra method inverse probability weights are used in the ra estimation to correct for misspecification in the regression function. If the regression function is correctly specified the weights do not affect the estimations. In the aipw method the treatment model includes a term that corrects this model if this model is mis-specified. In using both the ipwra and the aipw the overlap assumption needs to hold, i.e. all observation must have a non-zero probability of treatment. This requirement may be critical for the estimations. If the overlap assumption holds, both the ipwra and aipw estimations have the double-robust property for some functional form combinations, meaning that if either the outcome model or the treatment model is correctly specified, impacts are consistently estimated.

The basic intuition behind matching models (nearest neighbor matching (nn) and propensity score matching (ps)), is that outcomes are compared of observations that are as similar as possible, with the only exception of their treatment status. In nearest neighbor matching (nn) the similarity of observations with multiple covariates is calculated by constructing the distance between pairs of observations in terms of these covariates. Different scales of covariates and correlation between covariates is dealt with by calculating the so-called Mehalanobis distance. For removing large sample bias that arises because no formal outcome or treatment model is specified, a bias correction term needs to be included in the estimations in case of more than one continuous covariate. In propensity score matching (ps) a model of the probability of treatment (propensity score) is estimated. The sample is stratified in such a way that each stratum covers a subset of observations with similar characteristics. Impacts are calculated by comparing treatment and control observations within each stratum and subsequently use stratification weight to construct the aggregate impact. The treatment effects are calculated on the basis of matching the estimated probability of treatment.

We use these methods to estimate the supply response of a reduction in transaction costs (the treatment) caused by the introduction of a new auction floor. A few choices need to be specified: in the ipwra and aipw estimations the outcome model is linear since the outcome variable is continuous and the treatment model is probit since the treatment variable is binary. In the estimations we use variables as explanatory variables in the outcome model, as explanatory variables in the treatment model, and as matching variables in both nearest neighbour estimation and estimation of the propensity scores. We do, however, leave out variables if this generates a better fit. An overview of these variables is supplied in the tables.

*Data for estimations*

Impact estimations 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 [19]), 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[20]. There are a few EPAs that have no or negligible tobacco cultivation, notably EPAs in the districts Chikwawa and Nsanje.

[19] A few EPAs are larger than two times the mean size, but the bulk of the EPAs (90%) is smaller than 835km2.
[20] The EPA data cover the area of Malawi that is relevant for agriculture. Some parts of the country (e.g. various national parks and lakes) are excluded from the EPA 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 A). From the investigations we conclude that TCC auction sales volume data and AES-MoAFS production data show some, but justifiable discrepancies. The distinction between smallholders and estates, and burley tobacco and other tobacco's go a long way in accounting for the major differences. Across the border trade and lags in sales may also contribute to differences. However, tobacco production and area estimates based on IHS-2 household data are quite off the mark. In all respects the EPA data are most complete, comprehensive and detailed about smallholder area and production dynamics. The major drawback is the lack of other information (covariates) on EPA level. We have overcome this drawback in various ways (see below).

Distance from EPAs to the different auction floors is crow's eye distance[21]. This calculation involves a variety of errors: it is assumed that the earth is a perfect sphere (which is not the case): given the small distances on a global scale this entails only a marginal error. More important, 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. Also road distance may differ from crow's eye distance in a non-systematic way[22]. With a few exceptions,

[21] We use standard Great Circle Distances on the basis of latitude and longitude coordinates: Great Circle Distance is the shortest path between two points on the surface of a sphere (for calculation see e.g. www.cpearson.com).

[22] This applies most explicitly for a number of EPAs in the district of Mangochi: crows eye distance assumes that transport from some EPAs in this district to the Lilongwe, Chinkhoma and Mzuzu auction floors runs across

however, we find that the difference between road distance and crows-eye distance is more or less proportional[23] and using road distance data will lead to similar conclusions.

There is a nearly complete lack of directly publicly accessible country wide data at EPA level that can be used as covariates of production, area and production per hectare in tobacco[24]. We have resolved this by matching data from other sources to our EPA data. Consequently, the level of aggregation differs between these data. Measurement error is likely to increase in attributing data to EPAs, in constructing missing observations and in constructing complete EPA level series. This applies to prices, rainfall and fertilizer use. Farm gate price data for tobacco and maize and market prices for groundnuts are available for respectively 50, 60 and 70 markets / locations, and incomplete in varying degrees. Rainfall data are available for around 30 weather stations (but fortunately no missing observations). Fertilizer use is constructed on the basis of time series of aggregate Malawi fertilizer use and attributed to districts and EPAs on the basis of (district wise) fertilizer subsidies and (EPA) population. Hence, we use constructed and actual EPA level data on fertilizer use. Also 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. Note that the multiple levels of aggregation of covariates and their attribution to EPAs makes clustering of errors in the estimation virtually impossible.

In summary, we use a total number of nine control variables / covariates in estimations (rainfall (2x), fertilizer use, tobacco and maize farm gate prices, population density, a measure

---

lake Malawi. In principle transport by ship is feasible, but transport costs differ drastically from transport costs by truck. Hence, the crow's eye distance is not a good approximation for transport costs by truck.

[23] Note that road distance varies over time (and we do not always know how), while crows-eye distance 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 crows-eye distance.

[24] In its 2006/07 and 2007/08 Annual Agricultural Statistical Bulletin the Ministry of Agriculture and Food Security publishes only a few EPA level variables for specific years (amongst others the number of agricultural extension development officers and the number of farmer clubs).

of agglomeration and distance to nearest town or city (2x)). Statistics of these variables are shown in Table 2: tobacco and maize prices, fertilizer use and long run rainfall are similar, the other variables differ between treatment and control group. Current rainfall is slightly more favourable in the treatment EPAs, but production per hectare is lower. Spatial and integration variables indicate the treatment EPAs are somewhat more remote, less integrated and with lower population density. Further details on data, data sources and variable construction is in Appendix A.

**Table 2 Summary statistics of variables**

| | | treatment | | control | | |
|---|---|---|---|---|---|---|
| Variable | obser-vations | mean | standard deviation | Mean | standard deviation | \| t \| |
| market level data | | | | | | |
| real tobacco price, lagged (MK) | 357 | 69.4 | 34.9 | 66.7 | 28.5 | 1.4 |
| real maize price, lagged (MK) | 406 | 20.6 | 7.2 | 20.5 | 7.6 | 0.3 |
| district level data | | | | | | |
| chemical fertilizer per ha (kg/ha) | 189 | 0.109 | 0.051 | 0.114 | 0.078 | 1.0 |
| weather station level data | | | | | | |
| current rainfall (mm) | 224 | 981.6 | 298.4 | 929.2 | 242.3 | 3.0 |
| 3yr lagged average rainfall (mm) | 224 | 989.2 | 217.9 | 978.2 | 220.4 | 0.7 |
| EPA level data | | | | | | |
| tobacco production (ton) | 1266 | 1734 | 1416 | 515 | 767 | 18 |
| tobacco area (ha) | 1266 | 1954 | 1396 | 560 | 798 | 21 |
| tobacco production/area (ton/ha) | 1175 | 0.869 | 0.248 | 0.934 | 0.339 | 2.8 |
| population density (people/km2) | 1360 | 141.3 | 55.6 | 202.6 | 178.4 | 5.7 |
| distance to town (km) | 1360 | 39.2 | 22.4 | 36.2 | 19.7 | 2.2 |
| distance to city (km) | 1360 | 88.6 | 28.6 | 66.1 | 46.5 | 7.8 |
| agglomeration index (.) | 1360 | 0.202 | 0.051 | 0.259 | 0.152 | 6.2 |

Note to table: observations pertain to the number of independent observations in the original data set and cannot accurately be attributed to (EPA level) treatment and control; mean, standard deviation and t test for all variables are calculated on the basis of EPA level data or the EPA level variants of market, district or weather station level data.

## 4. Impact of market access on tobacco production per hectare, area and production

Selected estimation results for equation (7a) to (7c), using basic OLS and not controlling for the endogeneity of the intervention variable (and therefore labelled *naïve OLS*) , are reported in Table 3 (production per hectare), 4 (production) and 5 (area). Due to the endogeneity problem, estimates may be biased in the sense that they do not reflect the isolated impact of a change in transaction costs. However, the estimates do reflect the overall change in productivity, area and production that is associated with the introduction of an additional auction floor. If we succeed in estimating unbiased estimates we may also be able to assess the size and direction of the bias. We consider a specification with a fixed impact over the years (see Appendix D for OLS estimation results with variable impacts over the years) and we include EPA fixed effects, year fixed effects and covariates.

**Table 3 Market access and tobacco production per hectare: naïve OLS regression**

| Dependent variable: | natural logarithm of tobacco production per hectare, by Extension Planning Area ($\ln(q_{tobacco,i,t} / a_{tobacco,i})$) | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| I(2004-2010) | 0.161 | | | |
| | (3.6) | | | |
| I(2005-2010) | | 0.110 | | |
| | | (3.1) | | |
| I(2006-2010) | | | 0.137 | |
| | | | (4.2) | |
| ln(rainfall, current dif) | 0.167 | 0.137 | 0.151 | |
| | (3.8) | (3.2) | (3.5) | |
| ln(rainfall, lagged level) | 0.040 | -0.020 | 0.012 | |
| | (0.4) | (0.2) | (0.1) | |
| ln(chemical fertilizer use) | -0.034 | -0.031 | -0.040 | |
| | (1.8) | (1.7) | (2.1) | |
| ln(lagged real tobacco price) | 0.126 | 0.107 | 0.103 | |
| | (3.1) | (2.6) | (2.5) | |
| ln(lagged real maize price) | 0.034 | 0.051 | 0.044 | |
| | (0.9) | (1.3) | (1.1) | |
| ln(population density) | -1.193 | -1.062 | -0.994 | |
| | (2.2) | (2.0) | (1.8) | |
| ln(agglomeration index) | -3.137 | -3.705 | -3.121 | |
| | (3.0) | (3.6) | (2.9) | |
| dEPA(i) | yes | yes | yes | |
| dYEAR(j) | yes | yes | yes | |
| Number of observations | 1099 | 1099 | 1099 | |
| F (.) | (180, 918) 11.96 | (180, 918) 12.07 | (180, 918) 12.19 | |
| Prob > chi2 | 0.0000 | 0.0000 | 0.0000 | |
| Adjusted R2 | 0.6425 | 0.6447 | 0.6472 | |
| RMSE | 0.21633 | 0.21457 | 0.21441 | |

Notes – The table reports estimates of tobacco production per hectare. Estimations are based on annual data from 2003-04 to 2009-10 (seven years). All equations are estimated with OLS. Absolute *t*-statistics are given in parentheses (.) below the coefficient. Adjusted R2 = coefficient of determination, adjusted for degrees of freedom, and RMSE = Root Mean Squared Error. We do not report coefficients and t-values of the constant term and a complete set of EPA and year dummies (dEPA(i) and dYEAR(j)). Estimations follow the specification of equation (7a to c), with restrictions on the coefficients.

**Table 4 Market access and tobacco production: naïve OLS regression**

| Dependent variable: | natural logarithm of tobacco production relative to base period tobacco production, by Extension Planning Area ($\ln(q_{tobacco,i,t} / q_{tobacco,i,base})$) | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| I(2004-2010) | 0.277 | | | |
| | (2.6) | | | |
| I(2005-2010) | | 0.187 | | |
| | | (2.2) | | |
| I(2006-2010) | | | 0.158 | |
| | | | (2.1) | |
| ln(rainfall, current dif) | 0.144 | 0.103 | 0.117 | |
| | (1.5) | (1.0) | (1.2) | |
| ln(chemical fertilizer use) | -0.196 | -0.194 | -0.207 | |
| | (4.5) | (4.4) | (4.9) | |
| ln(lagged real tobacco price) | 0.244 | 0.232 | 0.171 | |
| | (2.6) | (2.5) | (1.9) | |
| ln(lagged real maize price) | 0.114 | 0.153 | 0.099 | |
| | (1.3) | (1.7) | (1.1) | |
| ln(population density) | -2.764 | -2.783 | -3.114 | |
| | (2.2) | (2.2) | (2.5) | |
| ln(agglomeration index) | -12.988 | -13.282 | -13.156 | |
| | (5.6) | (5.6) | (5.7) | |
| dEPA(i) | Yes | yes | yes | |
| dYEAR(j) | Yes | yes | yes | |
| Number of observations | 1100 | 1100 | 1100 | |
| F (.) | (179, 920) 11.16 | (179, 920) 11.10 | (179, 920) 12.00 | |
| Prob > chi2 | 0.0000 | 0.0000 | 0.000 | |
| Adjusted R2 | 0.6233 | 0.6218 | 0.6417 | |
| RMSE | 0.50478 | 0.50651 | 0.48732 | |

Notes – See previous table. The table reports estimates of tobacco production.

**Table 5 Market access and tobacco area: naïve OLS regression**

| Dependent variable: | natural logarithm of tobacco area relative to base period tobacco area, by Extension Planning Area ($\ln(a_{tobacco,i,t} / a_{tobacco,i,base})$) | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| I(2004-2010) | 0.083 | | | |
| | (0.9) | | | |
| I(2005-2010) | | 0.049 | | |
| | | (0.7) | | |
| I(2006-2010) | | | 0.015 | |
| | | | (0.2) | |
| ln(rainfall, lagged level) | -0.358 | -0.352 | -0.313 | |
| | (1.9) | (2.0) | (1.8) | |
| ln(chemical fertilizer use) | -0.160 | -0.177 | -0.173 | |
| | (4.3) | (5.0) | (4.9) | |
| ln(lagged real tobacco price) | 0.141 | 0.132 | 0.138 | |
| | (1.8) | (1.7) | (1.8) | |
| ln(lagged real maize price) | 0.121 | 0.034 | 0.051 | |
| | (1.6) | (0.5) | (0.7) | |
| ln(population density) | -1.247 | -1.344 | -1.347 | |
| | (1.2) | (1.3) | (1.3) | |
| ln(agglomeration index) | -12.619 | -12.483 | -12.224 | |

| | | | |
|---|---|---|---|
| | (6.3) | (6.5) | (6.3) |
| dEPA(i) | Yes | yes | yes |
| dYEAR(j) | yes | yes | yes |
| Number of observations | 1102 | 1102 | 1102 |
| F (.) | (179, 922) 8.67 | (179, 922) 9.56 | (179, 922) 9.55 |
| Prob > chi2 | 0.0000 | 0.0000 | 0.0000 |
| Adjusted R2 | 0.5549 | 0.5820 | 0.5816 |
| RMSE | 0.43354 | 0.40912 | 0.40933 |

Notes – See previous table. The table reports estimates of tobacco area.

We first consider the estimations of the impact of market access on production per hectare (Table 2). The coefficients of the intervention dummies are positive and statistically significant in all cases. The size of the impact varies from 11 to 16%. The included covariates – rainfall, chemical fertilizer use, tobacco prices, maize prices, population density and agglomeration – are well behaved: the signs of the coefficients are as expected, all of them are statistically significant at acceptable levels of accuracy. The contribution of covariates to the goodness of fit is very modest, increasing R2 with around 1% point. Note that the negative sign of chemical fertilizer use is due to interaction with year and EPA dummies (the sign of the coefficient becomes positive and significant if either of these fixed effects is omitted). We may conclude that the naïve OLS estimation results of Table 2 support a statistically significant impact of the new auction floor on production per hectare.

We are keen to find out to what extent this increase in production per hectare is associated with increases in production and increases in area. In other words does the effect of a reduction of transaction costs run through the intensive margin (an increase in production per hectare), through the extensive margin (an increase in area) or both. An increase in area has important implications: it indicates that smallholders are triggered to allocate (more) area to commercial agriculture. These questions are investigated on the basis of impact estimations using production and area as outcome variable, documented in Table 3 and 4.

The intervention variables in the production equation, reported in Table 3, are also statistically significant in all cases. The size of the impact varies from 16 to 28%. The performance of the covariates is good (statistically significant and expected sign), although rainfall is only weakly significant. Again, the sign of fertilizer use is negative, but this may be explained along the same lines as in the case of production per hectare. Impact on production is, as expected, higher than the impact on production per hectare: assuming an average impact on production per hectare of 16% and on production of 28%, we anticipate an impact on area of around 12%. But we may also obtain this result by direct estimation.

The intervention variables in the area equation, reported in Table 4, are less impressive: in all cases impact is statistically insignificant. With the exception of population density and real maize prices (which are both insignificant), covariates perform reasonable but less relative to the previous equations. With the insignificant estimates of impact we cannot check the consistency of the production per hectare and production estimates. We may conclude on the basis of the estimations in Table 2, 3 and 4 that market access has a statistically significant positive impact on production per hectare and production, if we use a naïve OLS estimation that does not take account of the endogeneity of the intervention.

*Estimation using matching models and potential outcome models*

In order to avoid the potential bias that arises with naïve OLS estimation, we employed matching techniques (nearest neighbour and propensity score matching) and potential outcome models (augmented inverse probability weighing and inverse probability weighted regression adjustment). These estimation techniques generate unbiased estimates of the population Average Treatment Effect (ATE). The estimations, reported in Table 6, 7 and 8, document the results of the matching techniques and potential outcome models. For both outcome and treatment model we have used all variables that are included as covariates in the naïve OLS estimation, or a subset of these variables.

From Table 6, 7 and 8 we see that ATE for the period 2005/10 and 2006/10 is statistically significant, for all estimation techniques and for both production per hectare, production and area. The impact on production per hectare (Table 6) varies from 14% to 20% for the period 2005/10, and from 20% to 25% for the period 2006/10; the impact on production (Table 7) varies from 21% to 24% for the period 2005/10, and from 36% to 38% for the period 2006/10 and the impact on area varies from 7% to 18% for the period 2005/10, and from 15% to 21% for the period 2006/10. The ATE estimates become increasingly precise if the intervention variable moves to a later period. Also, the ATE estimates for the sequence of intervention periods suggest that impact reaches full maturity for the period 2006/10: ATEs are very similar if we use the period 2007/10 (not shown, available on request from the author). It is also clear that impact is statistically insignificant for the period 2004/10, for all estimation techniques and for both production and area: apparently impact only starts in the crop season 2005/06 and only gradually.

**Table 6 Market access and tobacco production per hectare: using matching techniques and potential outcome models**

| Dependent variable: | natural logarithm of tobacco production per hectare, relative to base period tobacco production per hectare, by Extension Planning Area $(\ln[(q_{tobacco,i,t} / a_{tobacco,i,t}) / (q_{tobacco,i,base} / a_{tobacco,i,base}))$ | | | |
|---|---|---|---|---|
| | (nn) | (ps) | (aipw) | (ipwra) |
| I(2004-2010) | 0.100 | 0.016 | 0.107 | 0.110 |
| | (3.9) | (0.6) | (3.0) | (3.0) |
| I(2005-2010) | 0.159 | 0.132 | 0.186 | 0.197 |
| | (5.0) | (2.7) | (4.8) | (5.0) |
| I(2006-2010) | 0.204 | 0.245 | 0.241 | 0.253 |
| | (6.0) | (4.9) | (6.7) | (7.1) |
| Number of observations | 1107 | 1107 | 1107 | 1107 |

Notes – The table reports impact estimates of the auction floor introduction on tobacco production per hectare, using matching models (nearest neighbour (nn) and propensity score matching (ps)) and potential outcome models (augmented inverse probability weighted (aipw) and inverse probability weighted regression adjustment (ipwra)). Treatment models are probit models and outcome models are linear. All available continuous covariates, or a subset of these variables, are used in nn, ps, aipw and ipwra estimations, in case of the latter two, both for the outcome and the treatment model. These covariates are: rainfall, in deviation from long run values and average 3 year lagged levels, real lagged tobacco and maize price, population density, degree of agglomeration, distance to city and distance to town. Estimations are based on annual data from 2003-04 to 2009-10 (seven years). Absolute *z*-statistics are given in parentheses (.) below the coefficient.

**Table 7 Market access and tobacco production: using matching techniques and potential outcome models**

| Dependent variable: | natural logarithm of tobacco production relative to base period tobacco production, by Extension Planning Area ($\ln(q_{tobacco,i,t} / q_{tobacco,i,base})$) | | | |
|---|---|---|---|---|
| | (nn) | (ps) | (aipw) | (ipwra) |
| I(2004-2010) | 0.065 | 0.141 | 0.053 | 0.051 |
| | (1.2) | (2.1) | (1.0) | (1.0) |
| I(2005-2010) | 0.219 | 0.242 | 0.207 | 0.204 |
| | (3.4) | (4.4) | (3.7) | (3.7) |
| I(2006-2010) | 0.357 | 0.378 | 0.365 | 0.364 |
| | (5.5) | (5.2) | (6.1) | (6.3) |
| Number of observations | 1113 | 1113 | 1107 | 1107 |

Notes – See previous table. The table reports impact estimates of the auction floor introduction on tobacco production, using matching models and potential outcome models. Estimations are based on annual data from 2003-04 to 2009-10 (seven years). Absolute *z*-statistics are given in parentheses (.) below the coefficient.

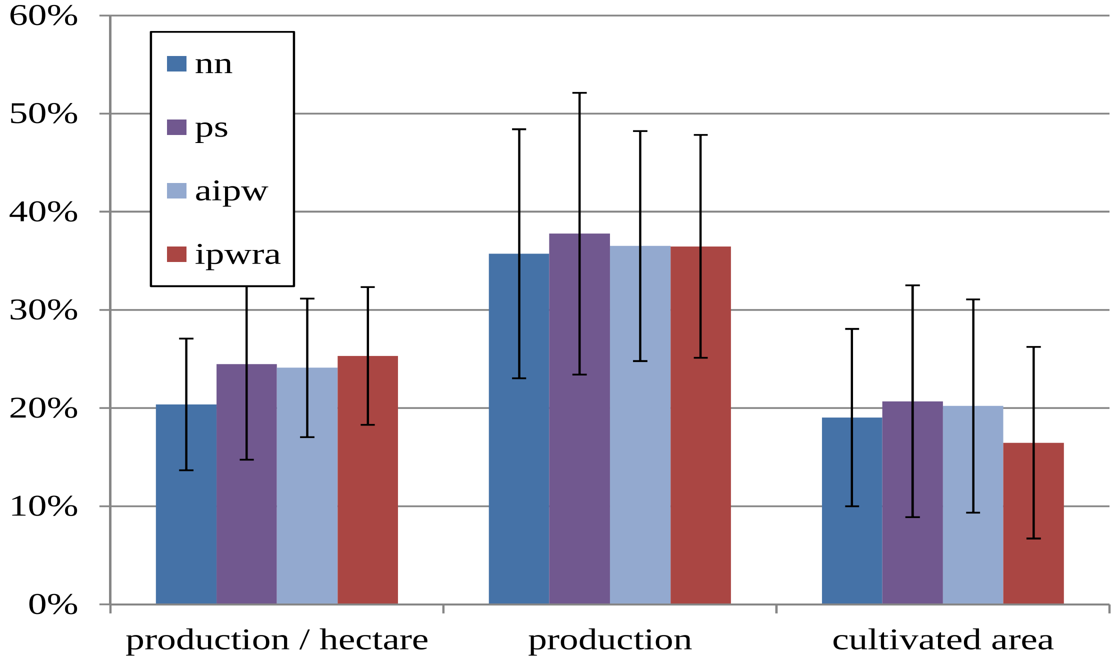
**Table 8 Market access and tobacco area: using matching techniques and potential outcome models**

| Dependent variable: | natural logarithm of tobacco area relative to base period tobacco area, by Extension Planning Area ($\ln(a_{tobacco,i,t} / a_{tobacco,i,base})$) | | | |
|---|---|---|---|---|
| | (nn) | (ps) | (aipw) | (ipwra) |
| I(2004-2010) | -0.005 | -0.097 | 0.001 | -0.005 |
| | (0.1) | (1.8) | (00.0) | (0.1) |
| I(2005-2010) | 0.074 | 0.182 | 0.099 | 0.084 |
| | (1.8) | (2.9) | (2.1) | (1.9) |
| I(2006-2010) | 0.201 | 0.207 | 0.202 | 0.148 |
| | (4.1) | (3.4) | (3.7) | (3.1) |
| Number of observations | 1107 | 1107 | 1107 | 1106 |

Notes – See previous table. The table reports impact estimates of the auction floor introduction on tobacco area, using matching models and potential outcome models. Estimations are based on annual data from 2003-04 to 2009-10 (seven years). Absolute *z*-statistics are given in parentheses (.) below the coefficient.

From the impact estimations reported in Table 6, 7 and 8, we conclude that the evidence supports a statistically significant positive impact of the introduction of a new tobacco auction floor on tobacco production per hectare, tobacco production and tobacco area and the size of the impact is in the range of around 13-15% on production per hectare, 36-38% on production and 15-21% on area. We also conclude that a statistically significant impact only starts in the crop year 2005/06, and reaches full maturity from 2006/07 onwards. The estimation results reported in Table 6, 7 and 8 represent the key message of this study.

**Figure 2 Impact of market access on tobacco production per hectare, on tobacco production and on tobacco area**




Notes to figure – Figures are drawn on the basis of estimation results; The figure shows the estimated fixed impact for 2006/10; the lines in the bars indicate 95% confidence intervals.

The estimation results from the tables are summarized in Figure 2. Confidence intervals (95%) of the estimated impact are included in the figures. The figure is self-explanatory and summarizes the major message of this research: The EPA data used for this work support a statistically significant impact of market access on tobacco production per hectare, tobacco production and tobacco cultivated area. It appears that if the change has reached its full impact production per hectare, production and cultivated area has increased, respectively, with around 22%, 37% and 19%. The estimates are more or less mutually consistent. A substantial part of the increase in production is due to increase in area.

*Alternative explanations and potential threats*

*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. Unfortunately we have few observations in the pre-intervention period: strictly, we have only one year of pre-intervention data (2003-04) since the auction floor started operation in 2004. However, from the estimations it is evident that the measured impact is delayed a few years, and therefore we may move up in the intervention period a few years for this purpose, and include 2004-05 and 2005-06: this allows us to investigate if the intervention and non-intervention EPAs were on a similar time path during the pre-intervention period.

**Table 9 Comparison of variables of intervention and non–intervention locations in the pre-intervention period**

| | | unconditional mean | | | |
|---|---|---|---|---|---|
| | **period** | **intervention** | **non-intervention** | **difference** | **F test** |
| **area (trend)** | 2003/04-2005/06 | 12.3 | 15.8 | -3.5 | 0.00 |
| | | (0.1) | (0.3) | | (0.972) |
| **area (growth)** | 2004/05-2006/07 | 0.029 | 0.024 | 0.005 | 0.01 |
| | | (0.6) | (0.8) | | (0.935) |
| **production (trend)** | 2003/04-2005/06 | 83.3 | 63.8 | 19.5 | 0.06 |
| | | (1.2) | (1.6) | | (0.811) |
| **production (growth)** | 2004/05-2005/06 | 0.275 | 0.285 | -0.010 | 0.01 |
| | | (3.7) | (5.7) | | (0.910) |

Notes – The table reports means and differences of the group of intervention EPAs and the group of non-intervention EPAs, both during the pre-intervention period. The table additionally reports the F test and its p-value on the significance of the difference. Trend is estimated, combined with a constant term.

Table 9 shows the unconditional means of intervention and non intervention variables of both outcome variables, its difference and mean difference tests, all during the pre-intervention period. We are concerned with developments over time and hence we look specifically at trends and growth rates. Tests on trend developments and annual rates of change of the outcome variables (area and production), 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 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 (see also Zant, 2012, for further details on the dominant role of maize). Expansion of tobacco crop area is realized either through expansion of total crop area or through substitution with other crops. We calculate the potential for expansion of total crop area, by computing the gap between the maximum total crop area over the years and actual total crop area by EPA ($pa_{it,expansion} = MAX_t(AREA_{it}) - AREA_{it}$ 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). The intuition is simple: we do not observe the potential crop area available for expansion, but assume that the maximum of total crop area realized over a period of several years, minus actual crop area, is a good approximation. Next, we calculate substitution for other crops by computing the difference between "other crop" area by EPA and the minimum area allocated to other crops in this EPA ($pa_{it,substitution} = area_{other\ crops,it} - MIN_t(area_{other\ crops,it})$ 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). Implicitly, we assume that all potential substitution takes place in the "other crops" sector and only to the extent that crop area allocated to other crops exceeds a minimum level. 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 the numerical exercise we find very few of the control EPAs to have expansion opportunities for tobacco cultivation less than 100% of existing tobacco area in any year. Only a few EPAs in the northern districts Rumphi and Mzimba[25] have below 100% potential area expansion opportunities. 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[26].

*Impact on other crops*

The statistically significant impact on tobacco production per hectare and tobacco production 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 production per hectare, production and area of an alternative crop, notably groundnuts. 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, although less important than tobacco. The OLS results of these estimations are reported in Table 10. The Table shows that the coefficient of the intervention is small, sometimes negative and always insignificant. Hence, the estimation results do not support a systematic and significant impact on production per hectare, production or area of groundnuts and the hypothesis that the estimated impact on tobacco production per hectare,

[25] Notably Bolero, Katowo, Mhuju and Mphompha in Rumphi, and Bwenga and Malidadi in Mzimba.

[26] We cannot analyze if availability of labour is a restriction to growth of tobacco production in the control EPAs. However, it is hard to believe that labour is used for low return activities rather than profitable and commercially attractive crops. Hence, we assume there is no restriction on labour, either because of a large number of unpaid workers in rural areas, but also because of easy conversion from low return activities to high return activities.

production and area applies to other crops as well, is not confirmed by impact estimations for groundnuts. Estimation on the basis of matching and potential outcome models (not shown) leads to similar conclusions. This provides further support for the impact of market access on production per hectare, production and area of smallholder tobacco farmers.

**Table 10 Impact estimations with placebo crop (groundnuts)**

Dependent variable:
natural logarithm of groundnut production per hectare ($\ln(q_{groundnut,i,t} / a_{groundnut,i,t})$, column (1) and (2)),
natural logarithm of groundnut production relative to base ($\ln(q_{groundnut,i,t} / q_{groundnut,i,base})$, column (3) and (4))
and natural logarithm of groundnut area relative to base ($\ln(a_{groundnut,i,t} / a_{groundnut,i,base})$, column (5) and (6))
all by Extension Planning Area (EPA)

| | production per hectare | | production | | area | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| I(2005/10) | -0.025 | | -0.073 | | -0.032 | |
| | (0.7) | | (1.2) | | (0.7) | |
| I(2006/10) | | 0.030 | | 0.002 | | -0.027 |
| | | (0.9) | | (0.0) | | (0.6) |
| ln(rainfall, lagged level) | 0.084 | 0.079 | 0.019 | 0.008 | -0.059 | -0.062 |
| | (0.9) | (0.9) | (0.1) | (0.1) | (0.5) | (0.5) |
| ln(rainfall, current dif) | 0.187 | 0.185 | 0.027 | 0.027 | -0.193 | -0.190 |
| | (4.4) | (4.4) | (0.4) | (0.4) | (3.7) | (3.6) |
| ln(chemical fertilizer use) | -0.033 | -0.035 | -0.023 | -0.026 | 0.040 | 0.040 |
| | (2.0) | (2.1) | (0.9) | (1.0) | (2.0) | (2.0) |
| ln(lagged real gnut price) | 0.051 | 0.059 | -0.044 | -0.030 | -0.106 | -0.104 |
| | (1.3) | (1.5) | (0.7) | (0.5) | (2.3) | (2.2) |
| ln(lagged real maize price) | 0.034 | 0.043 | 0.145 | 0.159 | 0.133 | 0.135 |
| | (0.9) | (1.1) | (2.4) | (2.6) | (2.8) | (2.8) |
| ln(population density) | -1.072 | -1.101 | 0.831 | 0.790 | 1.842 | 1.838 |
| | (1.9) | (2.0) | (1.0) | (0.9) | (2.7) | (2.7) |
| ln(agglomeration index) | -5.606 | -6.017 | -3.068 | -3.607 | 4.687 | 4.672 |
| | (5.6) | (6.0) | (2.0) | (2.3) | (3.8) | (3.8) |
| dEPA(i) | yes | yes | yes | yes | yes | yes |
| dYEAR (t) | yes | yes | yes | yes | yes | yes |
| Number of observations | 1237 | 1237 | 1214 | 1214 | 1225 | 1225 |
| F (.) | (209,1027) 15.88 | (209,1027) 15.89 | (198,1015) 11.00 | (198,1015) 10.09 | (199,1025) 10.89 | (199,1025) 10.89 |
| Prob > chi2 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| Adjusted R2 | 0.7156 | 0.7157 | 0.6201 | 0.6195 | 0.6165 | 0.6165 |
| RMSE | 0.22866 | 0.22863 | 0.35584 | 0.3561 | 0.28148 | 0.28149 |

Notes – The table reports estimates of groundnut 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. Absolute *t*-statistics are given in parentheses (.) below the coefficient. Adjusted R2 = coefficient of determination, adjusted for degrees of freedom, and RMSE = Root Mean Squared Error. We do not report coefficients and t-values of the constant term, a complete set of EPA and year dummies (dEPA(i), dYEAR(t)).

*Quality of the data*

The comparison of AES-MoAFS data with IHS-2 household data (see Appendix A) shows a number of serious discrepancies between these data. 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: the underestimation of area and production in the AES-MoAFS data in the years before the introduction may explain the observed impact in later years, in EPAs that benefit from the new auction floor. The comparison suggests that both production and area in the AES-MoAFS data are lower, making the possible error in production per hectare less or even negligible (see Figure A5). To investigate if the estimated impacts are statistical artefacts, we have checked the robustness of the results by omitting data from (parts of) these districts. The checks confirm that the estimation results can be maintained. Alternatively, we have varied the construction of the base period (2003-2004 and 2004-2005 and the combination of these years) in the production and area equation. This also did not affect the estimation results fundamentally. Finally, if estimated impact in tobacco cultivation would be due to a statistical problem with the data (measurement error), it is likely that a similar error occurs in other crops. This, however, is not confirmed in the placebo crop estimations for groundnuts.

## 5. Summary and conclusion

We have investigated the impact of improved market access on production per hectare and the underlying smallholder’s decisions on production and area. For this purpose we have exploited 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 the entire country and exclusively sold on official auctionfloors. There are four tobacco auction floors (Limbe (Blantyre), Kanengo (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 impact of the introduction of the Chinkhoma auction floor is estimated using matching models and potential outcome models. The estimation results support a statistically significant positive impact of the introduction of the new auction floor on tobacco production per hectare, increasing from 6% to 25%. The evidence further suggests statistically significant positive impacts on tobacco area (the extensive margin), increasing from 8% to 21%. The impact on production is estimated to increase from 21% to 38%, which reasonably well fits outcome of the production per hectare estimates. All impacts are statistically significant. Alternative explanations for the estimated impact (intervention and non-intervention on a different time path, restrictions to expansion in non-intervention locations, estimated impact applies to all crops, 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 production per hectare, production and area. This outcome, and the increase in area in particular, suggests 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 is classified as subsistence farmers (Integrated

Household Survey-2, 2004-05) it is unlikely that the increase in production per hectare, production and area has not also affected subsistence farmers[27].

**Acknowledgements**

I would like to thank Eric Bartelsman, Hans Binswanger, Chris Elbers, Christopher Gilbert, Menno Pradhan, Asger Moll Wingender, Bram Thuysbaert, Thea Nielsen, Frank Bruinsma, participants of the Nordic Conference in Development Economics in Götenborg, Sweden, 2012, the PEGNet Conference 2012 in Dakar, Senegal, 2012, the DIAL Conference, Université Dauphine, Paris, France, 2013 and colleagues of VU-University Amsterdam 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.

[27] 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.

## 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.

Cadot, O., L. Dutoit and M. Olarreaga, 2006, 'How Costly Is It for Poor Farmers to Lift Themselves Out of Subsistence?', The World Bank.

Fafchamps, M., 1999, Rural poverty, Risk and Development, Rome, Food and Agriculture Organization.

Fafchamps, M., E. Gabre-Madhin and B. Minten, 2005, 'Increasing Returns and Market Efficiency in Agricultural Trade', *Journal of Development Economics,* 78, 406-442.

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.

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

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.

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.

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

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, 2008, 'Burley Tobacco Clubs in Malawi: Nonmarket Institutions for Exports', The World Bank, Policy Research Working Paper 4561.

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.

Muto, M. and T. Yamano, 2009, 'The Impact of Mobile Phone Coverage Expansion on Market Participation: Panel Data Evidence from Uganda', *World Development*, 37, 12, 1887-1896.

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.

Zant, W., 2012, 'The Economics of Food Aid under Subsistence Farming with an Application to Malawi', *Food Policy*, 37, 124-141.

## Appendix A Data and variables

*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 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 crows-eye distance. 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 crows-eye distance 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 observations: the remaining observations are constructed by assuming a location specific fixed share of farm gate prices in national auction prices (Tobacco Control Commission (TCC)) and imputed these values to fill up the white spots. Auction prices are national and do not differ between the different auctions. 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 observations. The remaining share is 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).

Aggregate annual data on fertilizer use in 1000kg (urea and 23:21:0:4s) are from the 2007-2008 Annual Agricultural Statistical Bulletin of the MoAFS and from IFDC. Annual aggregate fertilizer consumption by type of fertilizer is allocated to districts using the allocation of subsidized fertilizer packages (data also from the Ministry of Agriculture and Food Security). Fertilizer is subsequently attributed to EPAs using the share of the EPA population in the district population.

Data on the number of households by EPA for the year 2007/08 are from the Ministry of Agriculture and Food Security. Combined with district data on average household size and population growth, 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). Subsequently, the 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 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 two types of variables are constructed: a simple distance measure of proximity and an agglomeration index. For proximity we have calculated the distance to nearest city town, using a threshold population size of towns and cities of 20,000 and 50,000. Note that the proximity variables are EPA fixed effects. Alternatively, we have calculated an agglomeration index as 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 negative relationship with tobacco production per hectare, production and area. The proximity to town variable is different from the agglomeration variable the agglomeration variable reflects the entire network while proximity to town focuses on the nearest city or town. The two proximity variables also allow heterogeneity in the impact of different sized towns. The agglomeration variable varies over the years (unlike the proximity variables) and between EPAs.

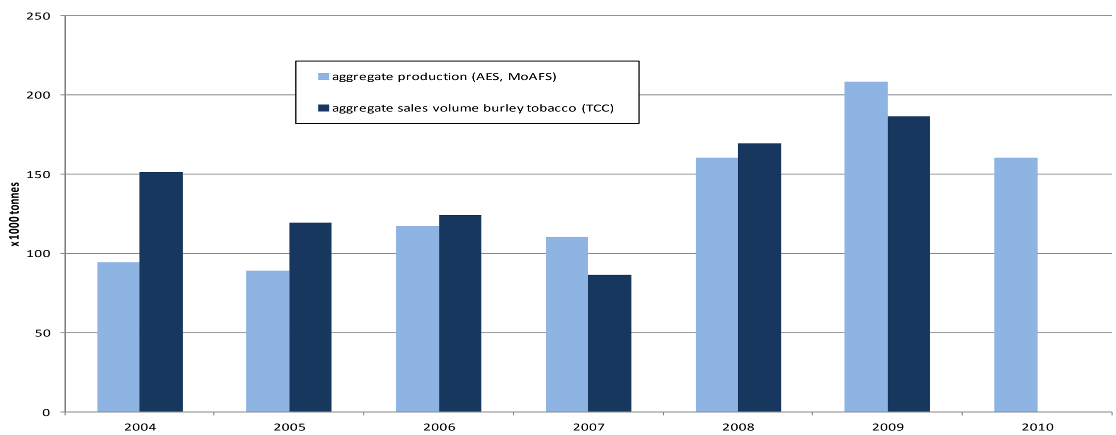
For descriptive purposes we use one year of auction transaction data (2009, a total of around 62,000 transactions), which were kindly made available by the Tobacco Control Commission (TCC). Each transaction contains information on type of tobacco, number of bales, volume in kg, value in US$, district 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).

*Production data vis-à-vis auction data*

How do the AES-MoAFS production data compare with the TCC auction sales data? On the basis of the available data we can compare the AES-MoAFS national production aggregates for the crop seasons 2003/04 to 2009/10 with TCC sales totals for the years 2004 to 2010 (see Figure A1) and the AES-MoAFS district composition of production of the crop season 2008/09 (2007/08) with the composition of sales by district of origin for the year 2009 from TCC transaction data (see Figure A2). These comparisons reveal a number of discrepancies: the AES-MoAFS aggregates are below the TCC sales totals in 2004 and 2005, while the difference gets smaller and even negative in later years. TCC sales data also contain sales by estates, while the production figures pertain to smallholders only: the burley tobacco sales by estates are 25,381 tons in 2004, 22,429 in 2005, 19,500 in 2006 and 10,750 in 2007, and hence go some way in explaining the differences. Additionally, different types of (non burley) tobaccos (included in TCC data), differences in time period covered, and illegal across-the-border trade in tobacco (both inflows and outflows) are likely to explain discrepancies between TCC and AES-MoADS series, but we cannot verify this.

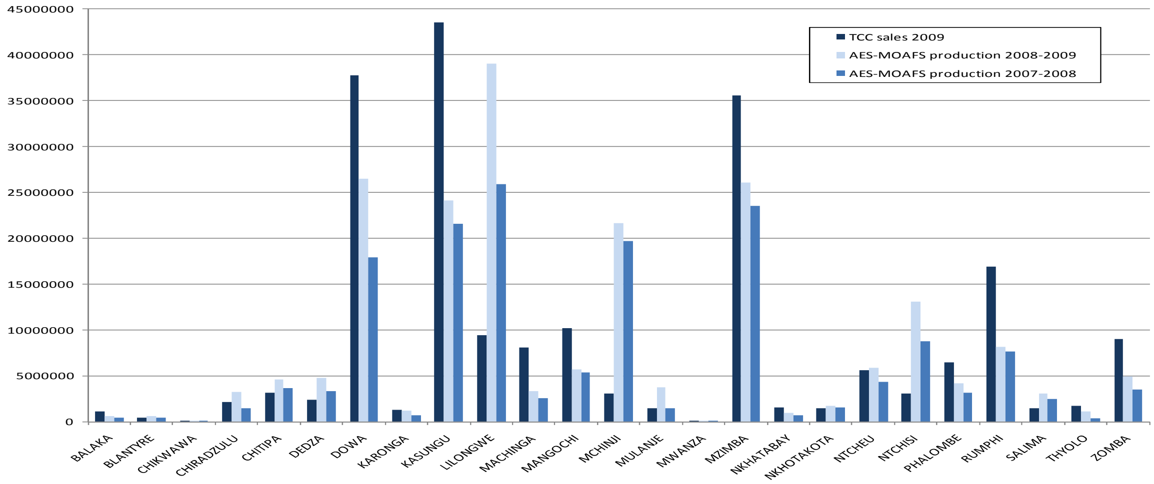
The district distribution of sales for 2009 shows large production overestimates / sales underestimates for Lilongwe, Mchinji and Ntchisi and large sales overestimates / production underestimates for Dowa, Kasungu, Mzimba and Rumphi. Possibly these differences are caused by the congestion at the Lilongwe auction floor, but we lack the information to verify this hypothesis. It is clear, however, that a systematic upward bias in the AES-MoAFS data is not supported by tobacco auction transactions data, collected by the TCC (see Figure A1).

**Figure A1 Tobacco aggregate production and auction sales volume over the years: TCC and MoAFS data**




Source: Tobacco Control Commission and Agro Economic Survey, Ministry of Agriculture and Food Security

**Figure A2 Tobacco production and auction sales volume by district in 2009: TCC and MoAFS data**




Source: Tobacco Control Commission and Agro Economic Survey, Ministry of Agriculture and Food Security

*Production and area data vis-à-vis household data*

How do the AES-MoAFS tobacco area and production data compare with tobacco area and production estimates based on household data of the Integrated Household Survey[28] (IHS-2, 2004-2005)? IHS-2 – which is based on a two stage stratified sample with a size of 11,280 households – contains a tobacco module (Section q) that reports if households have grown tobacco in the past 5 years, tobacco club membership, types of tobacco grown last season. For burley tobacco grown in last season, information is recorded on the area planted, net earnings, and on sales channels. To construct an estimate of area and production we make use of question q15a and b, and q18a and b[29]. There are 1466 households in the sample that report to have area cultivated with burley tobacco during the last season, and 1460 that report to have harvested burley tobacco during the last season (hence, on a national scale this corresponds with around 355,000 households, or 13% of all households). After conversion to

[28] The data of the Integrated Household Survey 2 (2004-2005) were kindly made available by the National Statistical Office in Zomba and the Poverty and Inequality Team of the World Bank.

[29] Question q15a: How much land did you plant to burley tobacco? And question q18a: How much burley tobacco did you harvest? Question q15b and q18b ask the corresponding unit of measurement. Both questions concern burley tobacco grown in the last completed season (2003-2004).

a common unit (hectare for area and kg for production), we use the household expansion factor to inflate tobacco area and production to their national representative level and collapse the data set to district aggregates[30].

The resulting district aggregates are summarized in Figures A3 to A5 and compared with area and production data from the AES-MoAFS. The comparison makes clear that both total area and total production for the whole of Malawi according to AES-MoAFS data is far below the IHS-2 estimates: the tobacco area aggregate based on IHS-2 (more than 1 million hectare) is highly questionable (AES-MoAFS, only smallholders, only burley tobacco: 141,406 hectare), but the production aggregate based on IHS-2 (180,839 tons) comes closer to a credible estimate (the AES-MoAFS production aggregate (only smallholders, only burley tobacco) is 94,444 tons; the TCC aggregate sales volume for 2004 (only burley tobacco, smallholders and estates: 151,453 tons; all tobacco, smallholders and estates:180,181 tons)[31]. At district level we further observe that the AES-MoAFS data for area and production are dramatically below the IHS-2 estimates for the districts Dowa, Lilongwe, Kasungu and Mchinji, although the discrepancy is smaller in the case of production (see Figure A3 and A4). Comparison of production per hectare by district suggest that the IHS-2 data for Dowa and Lilongwe are substantially off the mark, with average district production per hectare as low as 39 and 85 kg per hectare (see Figure A5). At the household level and at the other end of the scale, a substantial number of households have extremely high production per hectare: of the 1460 households with tobacco production there are 54 households with a per hectare production of more than 3000kg. This points at statistical errors in both area and production. The fluctuation in (average) production per hectare by district – measured with the coefficient

[30] The IHS-2 data are claimed to be representative at the district level. We cannot calculate EPA level aggregates on the basis of the IHS-2 data.
[31] It is unclear if tobacco production by estates is accounted for in the IHS-2 household data. This is not trivial: if estates are not accounted for (and since households are not estates this is likely), the IHS-2 production estimate is extremely high.

of variation – is nearly 4 times higher for the IHS-2 data. All these observations make us uncomfortable about the quality of the IHS-2 tobacco data.

**Figure A3 Tobacco data by district for 2003-2004 from IHS and MoAFS: tobacco area (upper panel), tobacco production (middle panel) and tobacco production per hectare (lower panel)**


Source: National Statistical Office and Agro Economic Survey, Ministry of Agriculture and Food Security

Researchers often point at alleged (but rarely substantiated) poor quality of administrative data, particularly those of Malawi maize production. From the investigations we conclude that TCC auction sales volume data and AES-MoAFS production data show some, but justifiable discrepancies. The distinction between smallholders and estates, and burley tobacco and other types of tobacco go a long way in accounting for the major differences. Across the border trade may also contribute to the differences between the series. Detailed sales data from TCC are not publicly available and, hence, we only have TCC data on aggregate sales volumes which are not helpful to investigate farmer responses. Data from the Integrated Household Surveys (we use IHS-2) does offer detailed information at the level of the household. However, in sharp contrast with TCC versus AES-MoAFS data, tobacco production and area estimates based on IHS-2 household data are quite off the mark, compared to AES-MoAFS data. Additionally, the IHS-2 data may be rich in terms of other household and agricultural variables, these data have little to offer in terms of dynamic behaviour of farmers.

## Appendix B Geographical location of auction floors

**Figure B1 Tobacco auction floors in Malawi**

## Appendix C Institutions in the tobacco commodity chain in Malawi

The Tobacco Association of Malawi (TAMA) was founded in 1929 and became a registered trust in 1983. TAMA is financed by the tobacco growers through a levy on the auction floors. Its objective is to take an active role in representing tobacco growers' interests in Malawi, to ensure profitable production and marketing of their tobacco through provision of such services as research, training, education and marketing promotion (www.tamalawi.com, accessed in June 2011). TAMA has a total number of association members (tobacco growers) of close to 50,000. Among a variety of activities like marketing, finance and inputs, TAMA runs – through their operations department – close to 100 satellite depots across the country to enable growers to deliver their tobacco at places near their farms. The operations department further coordinates tobacco haulage and tobacco re-handling through its fully fledged TAMA Re-handling Company.

The Tobacco Control Commission (TCC) was established in 1938 by the Minister of Agriculture under the control of the Tobacco Auction Floors Act, financed by tobacco growers through a levy on the auction floors. The major responsibilities of TCC are: control and regulation of tobacco marketing in Malawi; licensing and registration of tobacco growers and sellers; setting product quality standards (notably defining tobacco grades and classes); collecting, processing and distributing data and statistics on tobacco; carrying out market research and dissemination of markets studies; advising the government on tobacco issues and promotion and expansion of tobacco sales and the growth of tobacco industry and enhancing its contribution to the agricultural sector and GDP, and its contribution to Malawi's foreign exchange earnings (See TCC website (www.tccmw.com), accessed in June 2011).

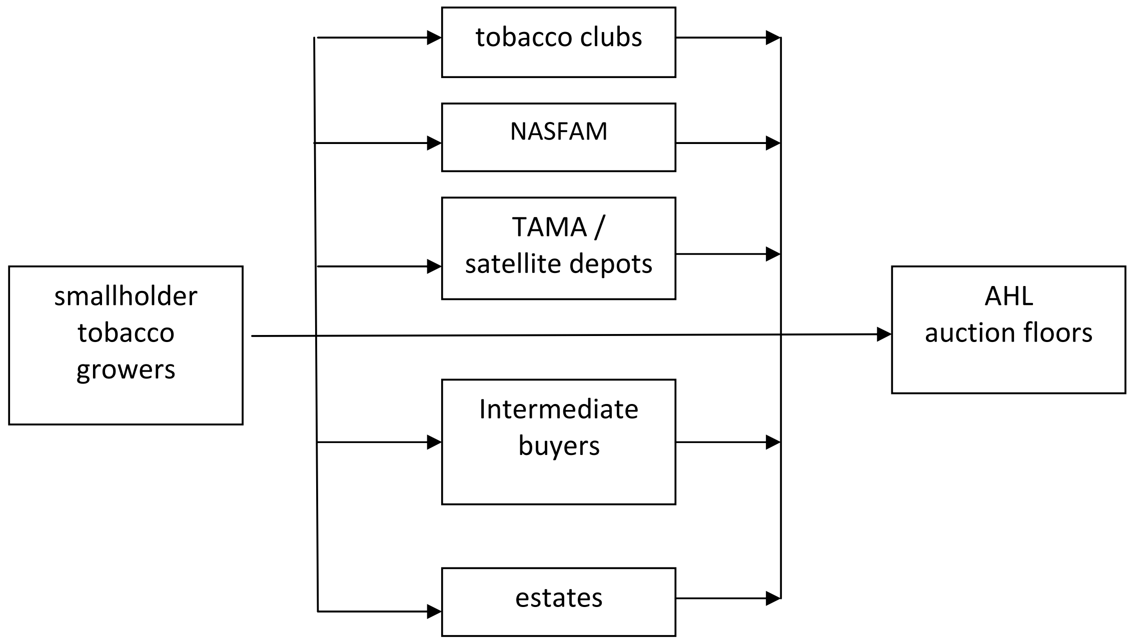
Auction Holdings Limited (AHL) is the private sector company that runs the auctions. Tobacco Auctions Limited and Producers Warehouse Limited, the predecessors of AHL whose first operations date back as early as 1936, operated in competition with each other until 1962, when they were amalgamated as Auction Holdings Nyasaland Limited. Following Malawi's Independence, the company was renamed Auction Holdings Limited in 1965 (See AHL website (www.ahlmw.com), accessed in June 2011). AHL currently runs operations at the four main tobacco auction floors (Limbe (Blantyre), Kanengo (Lilongwe), Mzuzu (Mzimba) and Chinkhoma (Kasungu)), together with a number of satellite auction floors, mini auction floors and rural markets[32]. In 2004 the Chinkhoma auction floor – near Kasungu in Central Malawi – has started operations as an independent scale auction floor. At all floors AHL provides tobacco marketing facilities and related support structures. The auction floor infrastructure includes large storage facilities and warehouses and after auction processing factories from buying companies. AHL claims to ensure that growers are paid the proceeds on their bank account within 24 hours after the sale.

The National Association of Smallholder Farmers of Malawi (NASFAM), grown out of the USAID funded Smallholder Agribusiness Development Project and founded in 1997, originally aimed at supporting and organizing smallholder tobacco production, with farmers clubs and associations as major organisational unit. NASFAM currently has an extension network that reaches over 100,000 smallholder farmers, an estimated share of 15 to 25% of all smallholder tobacco farmers (Jaffee, 2003; www.nasfam.org (accessed in June 2011) and supplies - amongst other things – insurance and transport services for transport of tobacco to the auction floors under the NASFAM transport program.

[32] TCC's tobacco market report covering the period from 1st March 2006 to 5th July 2006 states, for example: "Highlights of this year's tobacco marketing season include the opening of mini auction floors at Mgode in the Southern region, Linyangwa and Mpasadzi in the Central region and Kabwafu in the Northern region. The mini floors were opened to reduce congestion on the main floors of Lilongwe, Mzuzu and Limbe, and Chinkhoma satellite floors. The mini floors are also there to reduce the cost of transporting tobacco to the marketing that is entirely borne by the growers".

## Appendix D Trade Channels

**Figure D1 Trade channels for smallholder tobacco growers**



## Appendix E OLS estimations with variable impact by year

**Table E1 Market access&tobacco production per hectare: naïve OLS,variable impact**

| Dependent variable: natural logarithm of tobacco production per hectare, by EPA ($\ln(q_{tobacco,i,t} / a_{tobacco,i})$) | | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| I(2004-2005) | -0.254 | 0.073 | 0.148 | 0.152 |
| | (4.5) | (1.1) | (2.5) | (2.6) |
| I(2005-2006) | -0.008 | 0.324 | 0.044 | 0.085 |
| | (0.1) | (4.6) | (0.7) | (1.4) |
| I(2006-2007) | 0.214 | 0.536 | 0.121 | 0.150 |
| | (3.8) | (8.0) | (2.1) | (2.6) |
| I(2007-2008) | 0.208 | 0.536 | 0.122 | 0.173 |
| | (3.6) | (7.9) | (2.1) | (3.0) |
| I(2008-2009) | 0.371 | 0.700 | 0.174 | 0.213 |
| | (5.9) | (9.8) | (2.9) | (3.5) |
| I(2009-2010) | 0.213 | 0.543 | 0.277 | 0.324 |
| | (3.4) | (7.6) | (3.6) | (5.3) |
| ln(rainfall, current dif) | | | | 0.149 |
| | | | | (3.4) |
| ln(rainfall, lagged level) | | | | -0.015 |
| | | | | (0.2) |
| ln(chemical fertilizer use) | | | | -0.032 |
| | | | | (1.7) |
| ln(lagged real tobacco price) | | | | 0.111 |
| | | | | (2.7) |
| ln(lagged real maize price) | | | | 0.031 |
| | | | | (0.8) |
| ln(population density) | | | | -1.193 |
| | | | | (2.2) |
| ln(agglomeration index) | | | | -3.421 |
| | | | | (3.3) |
| dEPA(i) | no | yes | yes | yes |
| dYEAR(j) | no | no | yes | yes |
| Number of observations | 1098 | 1098 | 1098 | 1098 |
| F (.) | (6,1091) 15.34 | (172, 925) 4.31 | (178, 919) 11.43 | (185, 912) 11.67 |
| Prob > chi2 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| Adjusted R2 | 0.0727 | 0.3418 | 0.6286 | 0.6429 |
| RMSE | 0.34906 | 0.29409 | 0.22092 | 0.21662 |

Notes – The table reports estimates of tobacco production per hectare with impacts that vary over the years. Estimations are based on annual data from 2003-04 to 2009-10 (seven years). All equations are estimated with OLS. Absolute *t*-statistics are given in parentheses (.) below the coefficient. Adjusted R2 = coefficient of determination, adjusted for degrees of freedom, and RMSE = Root Mean Squared Error. We do not report coefficients and t-values of the constant term and a complete set of EPA and year dummies (dEPA(i) and dYEAR(j)). Estimations follow the specification of equation (7a to 7c), with restrictions on the coefficients. Outliers in estimations – up to a maximum of 1% of all observations and identified if size and (in)significance of coefficients are to a large degree due to one or a few observations – are removed.

**Table E2 Market access and tobacco production: naïve OLS, variable impact**

| Dependent variable: | natural logarithm of tobacco production relative to base period tobacco production, by Extension Planning Area ($\ln(q_{tobacco,i,t} / q_{tobacco,i,base})$) | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| I(2004-2005) | -0.427 | 0.148 | 0.252 | 0.259 |
| | (3.2) | (1.0) | (1.8) | (1.9) |
| I(2005-2006) | -0.213 | 0.424 | 0.093 | 0.179 |
| | (1.5) | (2.8) | (0.7) | (1.3) |
| I(2006-2007) | -0.128 | 0.436 | 0.190 | 0.316 |
| | (1.0) | (3.0) | (1.4) | (2.3) |
| I(2007-2008) | 0.369 | 0.941 | 0.262 | 0.436 |
| | (2.7) | (6.4) | (1.9) | (3.3) |
| I(2008-2009) | 0.467 | 1.103 | 0.037 | 0.180 |
| | (3.2) | (7.1) | (0.3) | (1.3) |
| I(2009-2010) | 0.135 | 0.772 | 0.021 | 0.207 |
| | (0.9) | (5.0) | (0.2) | (1.5) |
| ln(rainfall, current dif) | | | | 0.175 |
| | | | | (1.7) |
| ln(rainfall, lagged level) | | | | |
| ln(chemical fertilizer use) | | | | -0.193 |
| | | | | (4.4) |
| ln(lagged real tobacco price) | | | | 0.226 |
| | | | | (2.4) |
| ln(lagged real maize price) | | | | 0.078 |
| | | | | (0.8) |
| ln(population density) | | | | -2.938 |
| | | | | (2.3) |
| ln(agglomeration index) | | | | -12.864 |
| | | | | (5.4) |
| dEPA(i) | no | yes | yes | yes |
| dYEAR(j) | no | no | yes | yes |
| Number of observations | 1099 | 1099 | 1099 | 1099 |
| F (.) | (6, 1092)<br>5.55 | (172, 926)<br>5.26 | (178, 920)<br>10.42 | (184, 914)<br>11.11 |
| Prob > chi2 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| Adjusted R2 | 0.0242 | 0.4000 | 0.6044 | 0.6289 |
| RMSE | 0.81368 | 0.63807 | 0.51812 | 0.50179 |

Notes – See previous table. The table reports estimates of tobacco production with impacts that vary over the years.

**Table E3 Market access and tobacco area: naïve OLS estimation, variable impact**

| Dependent variable: | natural logarithm of tobacco area relative to base period tobacco area, by Extension Planning Area ($\ln(a_{tobacco,i,t} / a_{tobacco,i,base})$) | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| I(2004-2005) | -0.183 | 0.032 | 0.086 | 0.075 |
| | (1.8) | (0.3) | (0.8) | (0.7) |
| I(2005-2006) | -0.194 | 0.074 | 0.017 | 0.077 |
| | (1.8) | (0.6) | (0.2) | (0.7) |
| I(2006-2007) | -0.293 | -0.108 | 0.066 | 0.171 |
| | (2.8) | (1.0) | (0.6) | (1.6) |
| I(2007-2008) | 0.087 | 0.307 | 0.048 | 0.184 |
| | (0.8) | (2.8) | (0.4) | (1.7) |
| I(2008-2009) | 0.084 | 0.349 | -0.215 | -0.116 |
| | (0.8) | (3.0) | (1.9) | (1.0) |
| I(2009-2010) | -0.090 | 0.175 | -0.313 | -0.164 |
| | (0.8) | (1.5) | (2.8) | (1.5) |
| ln(rainfall, current dif) | | | | |
| | | | | |
| ln(rainfall, lagged level) | | | | -0.375 |
| | | | | (2.2) |
| ln(chemical fertilizer use) | | | | -0.161 |
| | | | | (4.6) |
| ln(lagged real tobacco price) | | | | 0.155 |
| | | | | (2.1) |
| ln(lagged real maize price) | | | | 0.075 |
| | | | | (1.0) |
| ln(population density) | | | | -1.296 |
| | | | | (1.3) |
| ln(agglomeration index) | | | | -11.633 |
| | | | | (6.1) |
| dEPA(i) | no | yes | yes | yes |
| dYEAR(j) | no | no | yes | yes |
| Number of observations | 1098 | 1098 | 1098 | 1098 |
| F (.) | (6, 1091) 2.68 | (172, 925) 5.48 | (178, 919) 9.10 | (184, 913) 9.68 |
| Prob > chi2 | 0.0139 | 0.0000 | 0.0000 | 0.0000 |
| Adjusted R2 | 0.0091 | 0.4127 | 0.5680 | 0.5929 |
| RMSE | 0.62234 | 0.4791 | 0.41091 | 0.39892 |

Notes – See previous table. The table reports estimates of tobacco area with impacts that vary over the years.

## Appendix F Tobacco auction transactions

### Table F1 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 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |

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