THE GRADUATE INSTITUTE | GENEVA

INSTITUT DE HAUTES ETUDES INTERNATIONALES ET DU DÉVELOPPEMENT

GRADUATE INSTITUTE OF INTERNATIONAL AND DEVELOPMENT STUDIES

Graduate Institute of
International and Development Studies Working Paper
No: 04/2012

The Impact of a Feeder Road Project on Cash Crop Production in Zambia's Eastern Province between 1997 and 2002

Christian K.M. Kingombe
Overseas Development Institute / Graduate Institute of International Studies

Salvatore di Falco
London School of Economics

Abstract

This paper investigates the dynamic impacts of rural road improvements on farm productivity and crop choices in Zambia's Eastern Province. There are several channels through which the feeder road improvements impact on farmers. Our aim is to estimate whether the differential outcomes in the five treatment districts and three control districts generated by the expansion of market agricultural activities among small to medium scale farmers could be explained by rural road improvements that took place after the new Chiluba MMD government in 1995 had completed an IMF rights accumulation programme bringing the principal marketing agent system to an end. Our district-level empirical analysis is an extension to the Brambilla and Porto(2005, 2007) cross-provincial level approach which proposes a dynamic approach accounting for entry and exit into the agricultural cotton sector to avoid biases in the estimates of aggregate productivity, when measuring productivity in agriculture applied to a repeated cross-sections of farm-level data from the Zambian post-harvest survey (PHS). Despite the limitations of the PHS data covering the period from 1996/1997 to 2001/2002 when the Eastern Province Feeder Road Project (EPFRP) was being implemented. The identification strategy relies on differences-in-differences of outcomes (i.e., cotton productivity) approach across two phases (pre-treatment and post-treatment). We use maize productivity to difference out unobserved household and aggregate agricultural year effects. Through our descriptive analysis we do find that changes in land allocation and in yields to Eastern Province's most important cash crop – cotton did occur at the district level. However, it is difficult to conclude that these changes are linked directly to the improved accessibility obtained from the implementation of the EPFRP based on our differences-in-differences estimator or our Tobit model.

The Impact of a Feeder Road Project on Cash Crop Production in Zambia's Eastern Province between 1997 and 2002

Christian K.M. Kingombe[1] and Salvatore di Falco[2]

Abstract:

This paper investigates the dynamic impacts of rural road improvements on farm productivity and crop choices in Zambia's Eastern Province. There are several channels through which the feeder road improvements impact on farmers. Our aim is to estimate whether the differential outcomes in the five treatment districts and three control districts generated by the expansion of market agricultural activities among small to medium scale farmers could be explained by rural road improvements that took place after the new Chiluba MMD government in 1995 had completed an IMF rights accumulation programme bringing the principal marketing agent system to an end. Our district-level empirical analysis is an extension to the Brambilla and Porto(2005, 2007) cross-provincial level approach which proposes a dynamic approach accounting for entry and exit into the agricultural cotton sector to avoid biases in the estimates of aggregate productivity, when measuring productivity in agriculture applied to a repeated cross-sections of farm-level data from the Zambian post-harvest survey (PHS). Despite the limitations of the PHS data covering the period from 1996/1997 to 2001/2002 when the Eastern Province Feeder Road Project (EPFRP) was being implemented. The identification strategy relies on differences-in-differences of outcomes (i.e., cotton productivity) approach across two phases (pre-treatment and post-treatment). We use maize productivity to difference out unobserved household and aggregate agricultural year effects. Through our descriptive analysis we do find that changes in land allocation and in yields to Eastern Province's most important cash crop – cotton did occur at the district level. However, it is difficult to conclude that these changes are linked directly to the improved accessibility obtained from the implementation of the EPFRP based on our differences-in-differences estimator or our Tobit model.

Key words: Impact evaluation, Cash crop choice; Cotton productivity; Africa; Zambia (Eastern Province).
JEL-codes: C2; C83; D2; O12; O13; Q12; R3.

---

[1] Senior Research Officer, Trade, Investment and Growth Programme, Overseas Development Institute, 111 Westminster Bridge Road, London SE1 7JD and Visiting Research Fellow, Development Studies, Graduate Institute of International and Development Studies (IHEID), (c.kingombe@odi.org.uk or Christian.kingombe@graduateinstitute.ch).
[2] Assistant Professor, London School of Economics, and Programme Director, MSc in Environment and Development, London School of Economics, Department of Geography and Environment, Houghton Street, London WC2A 2AE, United Kingdom LSE Room: K204 Kings Chambers (s.difalco@lse.ac.uk).
We have benefited from discussions with Colin Thirtle, Bhavani Shankar, Peter Hazell, Jonathan Kydd, Natascha Wagner, and the participants of seminars at the IHEID and Oxford (CSAE).

1. Introduction

In Zambia's Eastern Province poverty is largely a rural phenomenon. Agriculture is an important part of the livelihoods of many poor people, and it is frequently argued that agricultural growth is a fundamental pre-requisite for widespread poverty reduction. Paradoxically, in Zambia agriculture's share of GDP in terms of value added is higher today, than what it was at independence in 1964. This has led to questions about the benefits of attempts to promote directly agricultural growth and development, and about the best means to promote such growth (Dorward et al, 2004).

Fafchamps, Teal and Toye(2001:13) argue that *"while higher rates of growth achievable in export manufacturing may make it theoretically the best sector to support poverty reducing growth, in practice "only a handful" of African countries will be able to achieve this, so that "the 45 or so other African countries that do not become export platforms,"* e.g. through special economic zones (Brautigam, 2009) e.g. benefiting from the preferential access of clothing items to the American market within the African Growth and Opportunity Act(AGOA) scheduled to expire in 2015 (Kingombe and te Velde, 2012), *"must rely on other engines of growth: Agriculture, mining, tourism or a combination of them* (quoted in Dorward et al., 2004:75f)."

Despite the strong arguments for agriculture having provided the main engine of growth for rural poverty reduction in the past, reliance on pro-poor agricultural growth as the main weapon against rural poverty today may not be appropriate if the areas where today's rural poor are concentrated face severe difficulties in *raising agricultural productivity* or in accessing wider agricultural markets (Dorward et al., 2004) created by poor rural infrastructure (Thurlow and Wobst, 2005).

Transport infrastructure investments in rural areas are hypothesized to affect poverty through various channels. It increases *agricultural productivity*, which in turn directly increases *farm incomes* and helps reduce rural poverty (Fan and Chan-Kang, 2004). Thus, hypothetically speaking, improved provision of rural feeder roads should lead to lower transaction and farm production costs, while facilitating trading of cash crops and fostering long-run economic growth that contribute to the expansion of the economy. Hence, in this paper we want to test the following hypothesis:[3]

> The mean response in agricultural production and productivity growth to labour-based investment in rural roads within the treatment areas is the same as the mean response in the control areas.

In the case of Zambia's Eastern Province, the *main agricultural activities* are cotton, tobacco, maize, vegetables, and groundnuts. Since, income earners in the rural areas are not going to benefit much from price increases alone, so *quantity responses* are going to be critical for poverty reduction (Balat et al., 2004).

We use a *repeated annual (i.e. equal spaced) sequence of independent* cross-sections of farm-level data based on a random sample of the population. This dataset is extracted from the Zambian Post-Harvey Surveys (PHS), which covers all the districts of Zambia's Eastern

[3] In the words of Lofgren et al.(2004), spending on feeder roads in rural areas leads to the strongest reduction in national and rural poverty.

Province in the period from 1996/1997 to 2001/2002, allowing us to measure the short-term and medium-term gains from an UN sponsored Eastern Province Feeder Road Project (EPFRP) covering five districts in Eastern Province (Chadiza; Chipata; Lundazi; Katete; and Petauke districts), which was implemented during this period. These data are used to set up an empirical model of cash crop choice and cotton productivity.

Our *comparative inter-district analysis* builds on two studies by Brambilla and Porto(2005) and Brambilla and Porto(2007) both of which propose a dynamic approach accounting for entry and exit into the agricultural cotton sector. This is done to avoid biases in the estimates of aggregate productivity when measuring productivity in agriculture, which can be applied to our observational data for Zambia's Eastern Province.

The objective of our paper is to quantify the direct and indirect rural transport infrastructure investment impacts of the EPFRP. Although, the estimation of this supply response has proved difficult in the preceding literature, we will nevertheless attempt to explore the impacts on the production of the main cash crop – cotton – in Zambia's Eastern Province. The aim is to estimate whether *the differential cotton yield* generated by increased market agricultural activities arguable associated with the EPFRP treatment, with rural roads understood as a non-traditional production factor input in the production function.[4]

The remainder of the paper is organized as follows. Section 2 presents the public works programme and also offers a discussion of cotton production trends in Zambia's Eastern Province. Section 3 briefly surveys the literature and presents the equations to be estimated, and discusses the way in which the impact of the EPFRP on productivity is identified. Section 4 briefly presents the data covering the period from 1996/1997 to 2001/2002. Section 5 presents and discusses the empirical results for cotton productivity along with showing a number of robustness checks. Finally, section 6 concludes and discusses a few policy implications of the empirical findings.

2. Background

Zambia's Eastern Province covers an area of 69,106 square kilometres and today has 8 districts namely Chadiza, Chama, Chipata, Katete, Lundazi, Nyimba, Petauke and Mambwe. In 2000(2010) Eastern Province had a population of 1,300,973(1,707,731). Of this population, 49.4 (48.96) per cent were male and 50.6 (51.04) per cent were female. Eastern Province was growing at an average annual population growth rate of 2.72 (2.66) per cent between 1990 and 2000 (2000 and 2010) (CSO, 2001; CSO, 2011).

There are several *unique features to the Eastern Province*. Despite experiencing the least percentage growth among Zambia's nine provinces, *the agricultural households* in Eastern Province still constitutes Zambia's largest population.[5] The 2000 census of population also found that the whole province had the highest number (231,120) and the second highest percentage of *female headed households* (19.8%) (CSO, 2001).

[4] Only a total of 34,329 worker days were generated in Mambwe by Rehabilitation works which is less than 20% of the average workers days of the catchment districts. Moreover no workers days were created by Maintenance Road Works, therefore Mambwe is categorised as a control district.
[5] At the same time in 2000 Eastern Province had the lowest proportion of urban population at only 9 percent. Moreover, 4.1% of the agricultural households had an urban residence in 2000.

Of the 221,703 *agricultural households* in rural Eastern Province in 2000, the majority (38.6 percent of the total) were engaged in the *three major agricultural activities*: Crop growing, livestock- and poultry rearing while 28.8% were involved in crop growing and poultry rearing only. Only 20.6% were exclusively engaged in crop growing (CSO, 2001) .

Eastern Province's economy is *agro-based* and depends entirely on the soil with maize, cotton and tobacco being the *major cash crops* most of which are *intended for the export market*. The small scale farmers remain the key players of the local economy (Lungu, 2006).

Eastern Province has two distinct agro-ecological regions—the Eastern Plateau and the Luangwa Valley (see **Map A1** in appendix). Central, Southern and Eastern Plateau known as *agro-ecological region II* covers the Central, Southern and Eastern fertile plateau of Zambia (CSO, 1994). It is characterised by: Moderate rainfall ranging from 800 and 1,000 mm of annual rainfall.[6] The years 1996/1997 and 2003/2004 were periods of above average rainfall levels, whereas the agricultural seasons 1997/98 and 2004/2005 were respectively below and above the average rainfall levels in Eastern Province. In Eastern Province the rainfall is concentrated between October/November and April/May, during the other months there is no rainfall at all.

2.1. The Eastern Province Feeder Road Project

In the first half of the 1990s the road network in Eastern Province was in bad shape due to lack of maintenance and repair.[7] Generally the road accessibility within the province was very poor especially in the rainy season. As a result most private transporters were not willing to put their vehicles on the neglected routes (Eastern Province Chamber of Commerce and Industry, 2005).

In 1991, when the UN General Assembly classified Zambia as a Least Developed Country (LDC), this rendered Zambia eligible to receive UNCDF assistance. UNCDF in turn fielded an identification mission in 1993, which resulted in the formulation of two projects in the Eastern Province:

- The Rehabilitation and Maintenance of Feeder Roads (FRP) project; and
- The District Development Planning and Implementation (DDP) project.

The FRP started **12 June 1996** as a pilot for the introduction of Labour-based Technology (LBT) for road construction and the establishment of maintenance systems for the feeder roads. It was executed based on the Project document: *'Addendum to Project Agreement'* signed 21st of October 1996. The FRP ended **31 December 2001**. The FRP was a project within the Ministry of Local Government and Housing (MLGH) and implemented by the District Councils of Eastern Province,[8] which was designed to build and strengthen capacities in the local authorities and local private sector to rehabilitate and maintain feeder

[6] In the valley areas, the rainy season tends to begin and end earlier than elsewhere.

[7] Eastern Province lies between Latitude 10 and 15 degrees South and Longitude 30 and 33 degrees East. The Province lies between two international boundaries Malawi in the east and Mozambique in the South. The North Western boundary is marked by the Luangwa River, which separates Eastern Province from Lusaka, Central and Northern Provinces.

[8] *The District Councils* being the feeder roads authority in Zambia, act as client organizations whose responsibilities include: to select and prioritise roads, to prepare and sign rehabilitation and maintenance contracts and tender documents, to supervise and certify the works, to pay the contractors, etc. The District Works Departments are trained to act as contract managers in all these aspects.

and urban roads through contracting systems. The FRP was enhanced through the linkages to the DDP, which supported participatory planning and strengthening of the service delivery capacity of District Councils. In principle, FRP operated within the framework of the DDP, which aimed at developing the capacity within the districts to plan and manage public works, and involving communities in all development processes (Rwampororo, 2002; Clifton, 1998).[9]

The project had four immediate objectives:

i. Develop capacity of district councils' works departments to plan, design, implement and manage road rehabilitation and maintenance works by establishing Contract Management Units (CMU).
ii. Develop a private sector construction industry capable of rehabilitating and maintaining feeder roads using labour based methods by training and equipping small-scale road rehabilitation and maintenance contractors.
iii. Improve access to highly productive agricultural areas.
iv. Create direct employment in the rural communities by encouraging participation of local communities in road works (Clifton et al., 2001; Rwampororo et al., 2002).

Table 1: Eastern Province Road Sector Network, (km)

i.	ii.	iii.	iv.	v.	vi.
Trunk	**Main**	**District**	**Feeder (*)**	**(Primary Feeder)**	**Total (i+ii+iii+iv)**
415	179	1,516	3,862	-2,359	5,972

Source: Road Sector Investment Programme (ROADSIP), Bankable Document, August 2001.
Note: (*) Feeder roads have a further internal classification of primary, secondary and tertiary.

The 3,862 km of rural feeder roads constitute 65% of the total road network in Eastern Province (**table 1**). It is worth noting that only 21 km of the 450 km under maintenance were on roads rehabilitated under the FRP. This means that (404-21 = 383 km) of rehabilitated feeder roads have had no maintenance in 2002 since they had been improved. The remaining (450-21 = 429km) of maintained feeder roads were improved and maintained through enhanced maintenance contracts. This brings the total to 833 km of road addressed by the FRP (**table 2**) (Rwampororo, 2002). These maintained feeder roads are illustrated in **Map A2** (see appendix).

Table 2: Eastern Province Feeder Road Network

		Km	Total KM	Percentage of total feeder road network	Percentage of primary feeder road network
EPFRP	Rehabilitated Roads	21	404	404/3,862 = 10%	404/2,359 = 17%
	Rehabilitated and Maintaned Roads	383			
Enhanced Maintenance Roads		429		429/3,862 = 11%	429/2,359 = 18%
Total Maintained Road		833		833/3862 = 22%	833/2,359 = 35%

Source: Authors' calculations based on Rwampororo et al. 2002.

In relation to the feeder road network in the district the improvement to roads would account for 833 / 3,862 = 22% of the entire feeder road network in Eastern Province (**table 2**). However, given that *the roads were prioritised* and the most important links were

[9] The FRP operated alongside the DDP, also funded by UNDP and UNCDF, and had the same Project Manager. In principle, FRP operated within the framework of the DDP, which aimed at developing the capacity within the districts to plan and manage public works, and involving communities in all development processes (Rwampororo et al., 2002; Kalinda, 2001).

identified to be rehabilitated and maintained, *the impact* in the Province would be greater than the proportion of the network addressed (see last column in **table 2**).[10]

2.2. Cotton Production in Zambia's Eastern Province

Cotton is one of the key agricultural activities in rural Zambia and the cotton sector has been a success story in Zambia since a process of liberalization in cotton production and marketing began in 1994 (Balat and Porto, 2005b). Significant percentages of cotton farmers (due to soil characteristics) are observed only in the Southern Province, Mumbwa in Central Province, and the **Eastern province**, where it is the most relevant cash crop activity. In these three cotton-growing provinces, a large share of the cash income of rural farmers comes from the sale of cotton seeds. Eastern Province is the most important area for cotton production and it share of Zambia's total output increased from 15% in 1994 to 23% in 1997 (**table 3**).[11]

In the agricultural season 1995/1996 we notice a 100 per cent jump in the share of cotton area in Eastern Province (**table 3**). However, it was the year where the out-grower programmes were offered by two private companies Lonrho Cotton and Clark Cotton to provide participating small-to medium scale farmers with inputs and extension services.[12] However, it did not succeed in introducing much competition in the sector. Instead, the liberalization in 1996 gave rise to geographical monopsonies (i.e. the initial phase of regional private monopsonies) rather than national oligopsonies {Balat and Porto, 2005a; Brambilla and Porto, 2005, 2012; Chauvin and Porto, 2011).

Table 3: Share of Cotton Area in Total Cropped Area, 1993-1998

Harvest Year	Province (%)			Zambia	Province (km^2)			Zambia
	Central	Eastern	Southern		Central	Eastern	Southern	
1993	9%	6%	3%	3%	8495,55	4146,36	2558,49	22578,48
1994	7%	5%	5%	3%	6607,65	3455,3	4264,15	22578,48
1995	10%	7%	1%	4%	9439,5	4837,42	852,83	30104,64
1996	9%	14%	4%	6%	8495,55	9674,84	3411,32	45156,96
1997	16%	15%	6%	6%	15103,2	10365,9	5116,98	45156,96
1998	13%	17%	7%	7%	12271,35	11748,02	5969,81	52683,12

Source: Zambia Food Security Research Project, 2000 based upon Post-Harvest Surveys, Ministry of Agriculture, Food & Fisheries, Database Management Unit, Central Statistical Office.

In particular, instead of the localized monopsonies, entrants and incumbents started *competing in cotton trade* in many districts of Eastern Province, which was dominated by two businesses Clark Cotton, a South African firm which took over the Chipata Ginnery, and Sable Limited a completely new entrant into the market, which diversified into cotton trading from other trading activities in 1992/93 (Chiwele et al., 1998; Poulton et al., 2004). In addition, some entrants that were not using out-grower schemes started offering higher prices for cotton seeds to farmers who had already signed contracts with other firms. This caused

[10] The 'appraisal' stage used to select the road links to be rehabilitated can be thought of as 'ex-ante evaluation (Van de Walle, 2009).'
[11] As much as 45% in 1996 if we only include these three provinces in the total area sum.
[12] Before taking over the operations of Lintco in Mumbwa through privatization, Lonrho went into cotton trading in the 1992/93 marketing season following liberalization and traded alongside Lintco for some-time (Chiwele et al., 1998).

repayment problems and increased the rate of loan defaults (Balat and Porto, 2005a; Brambilla and Porto, 2005, 2012).

Table 4: Percentage of Farmers Growing Cotton in Eastern Province, 1997 – 2002

District	1996/97	1997/98	1998/99	1999/2000	2000/2001	2001/2002
Chadiza (301)	44,79%	27,27%	11,24%	10,00%	26,14%	27,00%
Chipata (303)	40,92%	33,90%	34,54%	25,74%	36,16%	36,97%
Katete (304)	50,51%	53,03%	35,18%	39,55%	52,17%	52,36%
Lundazi (305)	24,11%	25,78%	43,23%	11,92%	27,04%	24,52%
Petauke (308)	24,34%	16,79%	20,30%	8,44%	21,37%	21,31%
Total Catchment Districts	35,48%	30,99%	31,04%	19,55%	32,50%	32,20%
Chama (302)	18,92%	33,33%	30,26%	11,25%	10,00%	16,88%
Mambwe (306)	59,62%	61,82%	50,00%	49,15%	54,90%	54,24%
Nyimba (307)	6,25%	8,11%	11,32%	6,67%	22,22%	23,73%
Total Control Districts	29,93%	38,28%	28,22%	21,11%	26,86%	30,26%
Total	34,86%	31,75%	30,68%	19,76%	31,71%	31,93%

Source: Authors' calculations based on CSO's Post Harvest Surveys 1997-2002.

On top of all this, world prices in the 1998/99 agricultural season began to decline, and farm-gate prices declined as a result. After many years of high farm-gate prices, and with limited information on world market conditions, felt that out-growers' contracts were being breached, and the participation rates in the whole Eastern province declined from 32% percent in 1997/1998 to 20% in 1999/2000 (**table 4**).

From 2000/2001 to 2004/2005 the percentage of cotton growers returned to the previous level of more than 30% (**table 4**) and is correlated with entry into cotton. In 2004, there were three companies based in Eastern Province's capital Chipata: Clark Cotton Zambia Ltd; Dunavent Zambia Ltd. Cotton;[13] and Zambia-China Mulungushi Textiles Joint Venture (ZCMT).

Table 5: Fraction of Land Allocated to Cotton, 1997 - 2002

District	1996/97	1997/98	1998/99	1999/2000	2000/2001	2001/2002
Chadiza (301)	19,80%	10,76%	8,77%	5,09%	15,38%	17,12%
Chipata (303)	24,97%	21,03%	18,81%	14,24%	12,66%	13,07%
Katete (304)	21,19%	26,77%	20,02%	14,79%	18,24%	19,14%
Lundazi (305)	11,96%	15,38%	23,67%	6,77%	8,71%	8,89%
Petauke (308)	7,41%	9,66%	9,19%	3,23%	8,39%	7,35%
Total Catchment Districts	***16,37%***	***17,07%***	***17,02%***	***9,77%***	***11,50%***	***11,50%***
Chama (302)	8,38%	15,55%	14,66%	3,59%	3,48%	6,01%
Mambwe (306)	33,45%	21,22%	30,74%	23,89%	18,13%	22,21%
Nyimba (307)	1,56%	1,60%	7,87%	4,84%	9,72%	9,87%
Total Control Districts	***21,84%***	***13,97%***	***14,54%***	***9,99%***	***13,73%***	***16,37%***
Total	**17,06%**	**16,67%**	**16,69%**	**9,79%**	**12,76%**	**12,16%**

Source: Authors' calculations based on the Post Harvest Surveys 1996/1997-2001/2002.

Cotton farming was also affected by a decline of **the land area devoted to cotton** in 1998-1999 (when the outgrowing scheme was failing). This, however, was followed by a significant increase in area planted in 2000-2001, when the out-grower scheme was perfected with the entrance of the private company Dunavant, which in 1999 had initiated a distributor

[13] Its predecessor was Lonrho Cotton.

system (Tschirley et al., 2006; Poulton et al., 2004).[14] The fraction of land allocated to cotton by the average farmer in the Eastern province's catchment districts plummeted from 17% in 1998/99 to less than 10% in 1999/2000, and subsequently stayed below the percentages reached in the control districts (**table 5**).

On the other hand, the average **cotton yield** of the catchment districts not only increased continuously from 1996/1997 to 1999/2000, but from 1998/1999 when the EPFRP was implemented exceeded mean of the control districts until the trend changed in 2000/2001, where the overall Eastern Province cotton productivity dropped to less than 1 MT/HA from 1,64MT in the previous season equivalent to a percentage fall of more than 40 percent (**table 6**).

Evidently these results beg the question as to why rural roads improvements are considered an important explanatory variable of the Yields per Hectare in Cotton given the higher average yield in the control districts. The most important indicator is the fact that the average cotton yield in the catchment districts exceeded the mean yield in the control districts exactly in the period from 1998 to 2000 when most of the feeder roads in the treatment zones had just been rehabilitated. Hence, in the first instant these descriptive statistics do seem to have some bearing on the question.

Table 6: Yields per Hectare in Cotton (MT/HA), 1997 - 2002

District	1996/97	1997/98	1998/99	1999/2000	2000/2001	2001/2002
Chadiza (301)	0,739	2,154	1,869	1,477	0,777	0,770
Chipata (303)	1,935	1,316	1,848	1,470	0,870	0,869
Katete (304)	1,069	1,648	2,241	1,817	1,103	1,099
Lundazi (305)	0,925	1,219	1,204	1,078	0,783	0,773
Petauke (308)	1,137	1,430	1,507	2,686	1,065	1,090
Total Catchment Districts	***1,304***	***1,476***	***1,689***	***1,677***	***0,943***	***0,944***
Sub-total Observations	380	329	337	242	397	417
Chama (302)	0,988	1,452	1,320	1,863	1,016	1,017
Mambwe (306)	1,841	1,651	1,056	1,358	1,164	1,130
Nyimba (307)	0,271	0,651	0,774	0,434	1,081	1,081
Total Control Districts	***1,580***	***1,541***	***1,151***	***1,381***	***1,105***	***1,088***
Sub-total Observations	41	49	46	37	70	75
Total Eastern Province	**1,331**	**1,485**	**1,625**	**1,638**	**0,967**	***0,966***
Total Observations	**421**	**378**	**383**	**279**	**467**	**492**

Source: Authors' calculations based on the Post Harvest Surveys 1996/1997-2001/2002.

3. Literature Survey and Framework

The linkage between rural transport infrastructure and economic development can be identified and expressed using economic development measures such as cash crop productivity or production (Banister and Berechman, 2000). The impact of a transport infrastructure project on a regional economy varies depending on *the phase of the project*, because the interrelationships are not instantaneous and, in general, require considerable periods of time to transpire. Transportation spending for maintenance and rehabilitation of rural feeder roads affects current economic activity but also represents an investment in future growth. The main reasons for this are the long period necessary for investment

[14] Partly as a result of the failure of the out-grower scheme, *Lonrho* announced its sale in 1999 and Dunavant Zambia Limited entered the market.

implementation (1998-2001) as well as the time needed for the demand side adjustment.[15] The longer-term effect fosters economic growth that contributes to the expansion of a regional economy (New York State Department of Transportation, 2000).[16]

Underlying these *time lags* are market imperfections including incomplete information concerning infrastructure development, uncertainty regarding the behaviour of public authorities and private entities, high transaction costs emanating from imperfect land market and general market externalities (Dorward et al., 1998; Kydd and Dorward, 2004; Kydd et al., 2004). All of these make the transformation of transport improvements into economic benefits highly *time dependent*. The overall result is *a dynamic process* whose evolution depends on the initial conditions of local transport and activity systems and on the local transport and economic policies (Banister and Berechman, 2000).[17]

In the next three sub-sections we respectively provide a brief literature survey, a presentation of the equations to be estimated and finally a discussion of the identification strategy for agricultural productivity.

3.1. Literature Survey

The downside of the **structural approaches** *(e.g. CGE models or Macro-style simultaneous-equation econometric models*, (Fan and Chan Kang, 2004; or Chauvin and Porto, 2011)) is that their assumptions have to be plausible and they may not be empirically testable. *The upside* is that they gain on what we can learn as long as the assumptions are valid. In contrast **impact evaluation** is highly a-theoretical and basically reduced form. As mentioned by van de Walle(2009) often there is very little that we can understand about why it is that we have impacts in general and specifically what the channels are through which roads are having impacts. Nevertheless, the two types of approaches are considered to be complementary (van de Walle, 2009), because they are asking different questions and are looking at different things.

Impact evaluations of road infrastructure are complex because of the economy-wide effects that roads create. Roads influence a wide array of economic and social activities. Acting through lowered transport costs, roads might promote market activities, the availability and use of social services; affect the division of labour inside and outside the household; etc. A thorough evaluation of all these effects is necessary in order to assess the contribution of this type of investment on the welfare of the population (World Bank, 2010).[18]

The central empirical obstacle to estimating road impact is *reverse causation* (Fan and Chan Kang, 2004). Roads are not randomly placed and people do not randomly settle next to roads once they have been constructed. The causal link between better road access and the

[15] As the effects of a transport project reverberate through the economy, increasing income levels, consumer spending, etc., government coffers will increase, allowing for an expansion and / or improvement of public services.

[16] Cost related indirect economic benefits of transportation investment do not materialize *instantaneously* because they involve *long-term* business and household location decisions. In fact, a prevalent view is that economic effects are realized after lags between 4 and 7 years in the case of highway developments.

[17] There is an alleged complementarity between transport and telecommunication technologies. The ability to use telecommunications (e.g. Agricultural Extension Services through radio programmes) may affect travel needs of the agricultural extension service officers.

[18] Source: World Bank Development Impact Evaluation Initiative Website consulted February 2010.

benefits of such access may thus be obscured. Longitudinal (i.e. micro-level panel) data spanning a period of road construction can alleviate the endogenous road placement problem insofar as the unobservables determining such placement are fixed over time (Jacoby and Minten, 2009), while overcoming the limitations of the aggregate cross-country approach.

Most recent impact evaluation studies that take into account *the endogeneity issues* use **double-difference (DD)** combined with other methods to deal with the initial conditions that affect the trajectory of impacts. For instance DD combined with **propensity score matching (PSM)** (Mu and van de Walle, 2007). **DD** focuses on difference in outcomes over time between project and non-project communities. This approach purges additive time-invariant observables and unobservables (and deals with *time invariant selection bias*). But, initial conditions may also influence subsequent changes and trajectories (*time varying selection bias*). **PSM** is used to select ideal comparison communities (Van de Walle, 2009; Jalan and Ravallion, 2003; McCord and Wilkinson, 2009).

The results from a long-term evaluation of a World Bank-funded roads rehabilitation and improvement project in **Vietnam** by Mu and van de Walle(2007) show heterogeneous impacts across regions and socio-economic groups. For example, it finds that markets are more likely to develop as a result of road improvements where communities have access to extended networks of transport infrastructure.

Another method used by Gibson & Rozelle(2003) has been to combine DD with **instrumental variables (IV)** in Papua New Guinea. Khandker et al.,(2006) use DD and controls for initial conditions through OLS for Bangladesh.

A different approach applied by respectively Jalan & Ravallion(2002) and Dercon et al.,(2006) uses **dynamic panel data models** to look at either initial assignment or changes in road assignments. Here the approach is to look at the impact on consumption and poverty reduction. The key is to look at least *three waves of a panel* to be able to use this approach adequately. However, it is very rare to find this kind of data on access to roads.

A number of panel data sets allow an analysis of similar questions in **Africa**. A shorter panel from **Ethiopia** demonstrates the importance of price variables as well as exogenous shocks (rainfall) for analysing growth at the household level (Dercon, 2001).

Micro-level survey and a longer panel data evidence from **Uganda** spanning 1992-2000 by Deininger and Okidi(2003) confirms the benefits from Uganda‘s decisive liberalisation of output markets. It demonstrates the importance of improving access to basic education and health care emerges more clearly than in cross-country analysis, but benefits depend e.g. on complementary investments in electricity and other infrastructure.

An impact evaluation of rural roads investments in **Bangladesh** by Khandker et al.,(2006) finds that roads improvements led to lower input and transportation costs, higher production, higher wages, and higher output prices as well as to increases in both girls‘ and boys‘ schooling.

A study of rural roads in **Nepal** by Jacoby (2000) finds that access to roads improves the productive capacity of poor households. However, the study also concluded that the impact of roads on poverty reduction was limited and had no effect on inequality.

An impact evaluation by Escobal and Ponce (2003) find that the rehabilitation and maintenance of roads in **Peru** improved some measures (access and attendance to schools and child health centres) or had no significant impact on others (agricultural production, income, poverty).

In order to impose further structure on the **microeconometric approach** Ravallion and Jalan (1996); Jalan and Ravallion (1997) and Dercon and Hoddinott(2005) borrow from the conceptual framework used to understand growth at the national or cross-national country level.

One limitation to the studies, that specify an aggregate production function which includes transportation infrastructure among the set of explanatory variables, is the failure to take **road quality** into account. Road quality can vary greatly within a country and different quality roads can act in different ways (see **table 1** above). *Failure to discriminate amongst types of roads* can also lead to biased estimates according to Fan(2004).

Dercon and Hoddinott(2005) estimate a series of *probit regressions* through which they find that *an increase of 10 km in the distance* from the rural village to the closest market town has *a dramatic effect on the likelihood* that the household purchases inputs, controlling for the effect of other factors. However, they get mixed results in terms of the likelihood of engaging in various activities when *roads of poor quality* (accessible only to carts, animals, or people) were replaced by *good quality roads* (reasonable access to any vehicle).

Dercon and Hoddinott(2005) also find from their *Fixed effect IV regression* that *increases in road quality* have strong positive growth effects: Improvement in roads leading to local towns, from a road *poorly accessible* to buses and trucks to one *reasonably accessible* for buses and trucks in the rainy season results in 3.5 percent higher growth. Furthermore, there is a persistent and divergent effect linked to road quality: The better level of past road quality increases growth.

This finding is corroborated by a number of other studies e.g. (Fan and Chan-Kang, 2004; Deichmann et al., 2002). Fan and Chan-Kang(2004) for example find that *low quality (mostly rural) roads* have benefit/cost ratios for *national GDP* in China that are about four times larger than the benefit/cost ratios for high quality roads. As far as *agricultural GDP* in China is concerned, high quality roads do not have a statistically significant impact while low quality roads generate 1.57 yuan of agricultural GDP for every yuan invested. Investment in low quality roads also generates high returns in *rural nonfarm GDP*. In terms of poverty reduction, Fan and Chan-Kang(2004) also find that *low quality roads* raise far more rural and urban poor above the poverty line per yuan invested than do high quality roads.

The presence or absence of road infrastructure is perceived to be one of the main determinants of food price variation. An analysis by Minten and Kyle(1999) show that in the case of the Democratic Republic of Congo (DRC), *food price dispersion* is significant both across products and across regions. They demonstrate that *transportation costs* explain most of the differences in food prices between producer regions and that *road quality* is an important factor in the transportation costs (see Ulimwengu et al., 2010).

Building on these *evaluation studies of rural roads* our paper seeks to contribute to the micro-level supply-side literature by seeking to answer the key research question: *Do rural assets created through the EPFRP build sustainable livelihoods in the treatment areas?* The paper fills a gap by looking at the medium-to long-term effects on livelihoods and local economies in terms of crop production and yield.

3.2. Agricultural Productivity: Estimation Strategy

The counterfactual framework is where each individual has an outcome with and without treatment. This approach allows us to define various treatment effects and it was pioneered by Rubin(1974) and since adopted by Rosenbaum (1983); Heckman(1992, 1997); Imbens and Angrist(1994); Angrist et al.,(1996); Heckman(1997); Angrist(1998) and Wooldridge(2002).

3.2.1. A Counterfactual Setting

A cause is viewed as a manipulation or treatment that brings about a change in the variable of interest, compared to some baseline, called the control. The basic problem in identifying a causal effect is that the variable of interest is observed either under the treatment or control regimes, but never both (Dehejia and Wahba, 2002).

Framework

We want to evaluate the causal effect of the binary rural transport infrastructure (RTI = EPFRP) treatment variable on a continuous 'logarithm of cotton productivity (or production)' outcome Y experienced by units in the population of interest.

Thus, in our observational study of the EPFRP's impact on cotton productivity (*logyield*), by definition there are no experimental controls. Therefore, there is no direct counterpart of the *Average Treatment Effect* **(ATE)**. In other words, the *counterfactual* is not identified. As a substitute we may obtain data from a set of *potential comparison units* that are not necessarily drawn from the same population as the treated units, but for whom the observable characteristics, **x**, match those of the treated units up to some selected degree of closeness. The average outcome for the *untreated matched group* identifies the *mean counterfactual outcome* for the treated group in the absence of the treatment. This approach solves ***the evaluation problem*** by assuming that selection is unrelated to the untreated outcome, conditional on **x** (Wooldridge, 2002).

3.2.2. Estimating Model of Cotton Productivity

In our **regression analysis** approach we seek to estimate whether the agricultural productivity (or production) outcomes are better in the **treatment districts**, where the existing feeder road network was partially improved, with similar outcomes from **"control" districts** where the EPFRP wasn't implemented while recognizing the difficulty of "controlling" for the myriad of other factors impacting on the comparative economic performance of these eight districts in Zambia's Eastern Province.[19]

We estimate **a simple model of cash crop productivity**. This is done by modifying the empirical model in Brambilla and Porto(2005, 2007), in which the dependent variable *agricultural productivity* (i.e. yield of cash crop) is defined in physical units (i.e. quantity per hectare of cultivated land). Let y_{ht}^{c} denote the volume of cotton production (in Metric Tonnes) per hectare produced by household **h** in period **t**. The log of output per hectare is given by:

[19] Thus, if the case study districts in which the EPFRP has been implemented can be shown to have experienced faster agricultural growth than would have been predicted on the basis of trends in the wider regional economy or with the *three control group districts* (Chadiza; Chama; and Mambwe), then this "additional" growth may be deemed to be due to the EPFRP (DTZ, 2004).

(3) $$\ln y^c_{ht} = x^c_{ht}\beta_c + I_t + \eta_{ht} + b_0\phi_{ht} + \varepsilon^c_{ht} + \alpha_1 RTI_t$$

Here, x^c_{ht} is a vector of *household determinants of cotton yields* including the age (i.e. experience) and sex of the household head, the size of the household, household demographics (i.e. share of male household members), input use (i.e. fertilisers), assets (i.e. livestocks), the size of the land allocated to cotton, farm size (i.e. stratum), and district dummies.[20] The rural roads dummy variable RTI_t captures whether the district has been *'treated'* by rural transport infrastructure improvements (i.e. this indicates the presence of the EPFRP) in the period from 1998/1999 to 2001/2002. Thus, we model *the productivity effects of the EPFRP* with one dummy variables (or alternatively the percentage share of the district feeder roads, which have been treated), RTI_t. The *impacts of these EPFRP treatments* are measured relatively to the excluded category, which is the introductory phase 1996-1998 (with $RTI_t = 0$) and the three control districts in the subsequent implementation phase of 1998-2002.

Following Brambilla and Porto(2005, 2007) the model includes a number of **fixed effects**, such as districts effects (included in **x**), year effects, $\mathbf{I_t}$, and idiosyncratic household level fixed effects $\boldsymbol{\eta_{ht}}$ and $\boldsymbol{\phi_{ht}}$ and ε_{ht}.

The district effects include market access, local infrastructure (road density), local knowledge and access to credit; they are controlled for with district dummies.

The year effects, $\mathbf{I_t}$, capture aggregate agricultural effects and other shocks that are common to all farmers in a given period **t**. In equation **(3)**, these effects *cannot be separately identified* from the infrastructure provision dummy RTI. To deal with this, following Brambilla and Porto(2005) we propose to **model productivity in other crops** (mainly maize) to difference out *time varying factors* that affect productivity in agriculture.

The household level fixed effect has two components: A farm effect, η, and a cotton-specific effect, ϕ.

The farm effect η captures all idiosyncratic factors affecting general agricultural productivity in farm **h** that are *not observed* by the econometrician and are thus not included in **x**. It includes soil quality, de jure (i.e., titles) and de factor land rights (Bellemare, 2009; Goldstein and Udry, 2008) , know-how, and other factors that affect productivity in all crops.

The cotton-(idiosynctratic) specific effect ϕ is a combination of *unobserved factors* that affect productivity in cotton, including ability and expertise in cotton husbandry and suitability of the land for cotton (Brambilla and Porto(2005). Although, cotton in Zambia's Eastern Province is intercropped with groundnut, we assume that the farmer's cotton characteristics are somewhat different than the farmer's characteristics in **x** e.g. given the different cotton cultivation practices (i.e. sowing, ploughing, harrowing, etc.) and marketing practices.

According to Brambilla and Porto(2005) there are two problems with *the household fixed* effects. First, both η and ϕ are unobserved by the econometrician but observed by the farmer when making *decisions on input and on land allocation to different crops*. Hence, some of the variables included in **x** may be *correlated* with these *unobservables*. In addition, *entry and exit into cotton farming* depend on these unobservables as well since farmers'

[20] It is a *production function*, not a supply function, since prices are not included in x^c_{ht}

decisions on *land allocation to different crops* may be based on η and ϕ. More importantly for our purposes, this *entry/exit component* affects the estimates of the **RTI** dummy by altering the composition of farmers that produce cotton in each time period.

3.3. Agricultural Productivity: Identification Strategy

We use the random sample statistics from these target areas, which the CSO collected in *the six year* period from **1996/1997 to 2001/2002**. This *pseudo-panel dataset* ideally should present us with an opportunity to use *panel data analysis* to test which factors that determine the variation of the productivity of cash crops. A panel data set would allow us to account for both *idiosyncratic effects*.[21] However, the PHS dataset is *a repeated independent cross section of farmers*, which makes it impossible to track the same household over time as required in a genuine panel, because the sample design does not attempt to retain the same units in the sample (Baltagi, 2001). We thus need additional modelling to deal with *the fixed effects*. Following Brambilla and Porto(2005) we propose to model agricultural productivity in maize to control for η (and the year effects, $\mathbf{I_t}$) and to model the share of land devoted to cotton to control for ϕ.

It is highly likely that the factors that attract better roads in certain areas also affect the agricultural productivity outcomes. Unless the comparison areas – the counterfactual – have the same factors, it will leave biased estimates. *Selection bias* occurs if for some reason roads are poor in participating area and being compared with places that don't have these factors.

The identification strategy relies on **a modified difference-in-differences (DD) approach** to get rid of endogeneity. First, again following Brambilla and Porto(2005) we take differences of outcomes (i.e., productivity or production) across the different phases. Second, we use maize productivity to *difference out* unobserved household and aggregate agricultural year effects. Finally, since more productive cash crop farmers are also more likely to allocate a larger fraction of their land to cash crop production, we use *cash crop shares*, purged of observed covariates, as a proxy for unobserved cash crop-idiosyncratic productivity. Failure to adequately control for time-varying initial conditions that lead to the road placement can lead to very large biases in estimates of impacts (Van de Walle, 2009).

Yields per hectare in maize, y^m_{ht}, are given by:

$$\text{(4)} \qquad \ln y^m_{ht} = x^{m}_{ht}{}'\beta_m + I_t + \eta_{ht} + \varepsilon^m_{ht},$$

Maize productivity depends on covariates, x^m_{ht}, including district effects, the agricultural year effects, $\mathbf{I_t}$, and the farm effects η_{ht}.

By taking **difference**, we get

$$\text{(5)} \qquad \ln y_{ht} = \ln(y^c_{ht} / y^m_{ht}) = \ln(y^c_{ht}) - \ln(y^m_{ht}) = x'_{ht}\beta + \alpha_1 RTI_t + b_0\phi_{ht} + \varepsilon_{ht}.$$

Here, the observed household covariates $\mathbf{x_{ht}}$ included in the estimation are based on the same determinants of productivity as mentioned above. The district EPFRP dummy capture local market access effects, we therefore allow marketing conditions to affect *cotton* (a cash

[21] The *unobservables* ϕ and η are **indexed by ht** because, given the *cross-sectional* nature of the data, the unit of observation is a household-time period combination. However, if the data were a *panel*, ϕ and η would be **indexed by h** only (Brambilla and Porto, 2006).

crop activity) and maize (a mostly subsistence crop) differently. The coefficient $\boldsymbol{\alpha_1}$ measure the impact of the implementation phase of the EPFRP on cotton productivity.

There are two important *identification assumptions*. First, we assume that the agricultural effects, $\mathbf{I_t}$, have the same effect on growth of cotton and maize output per hectare. This according to Brambilla and Porto(2005) means that we can use *the trend in maize productivity to predict the counterfactual productivity in cotton in the absence of the EPFRP*. The assumption implies that we could *use productivity in other crops to difference out the agricultural effects*. Under the maintained hypothesis, the trend in maize productivity and the trend in the productivity of other crops should be similar. In the regression analysis, we use **maize as control** because, as opposed to the other crops, virtually all households produce it (**table 7**).

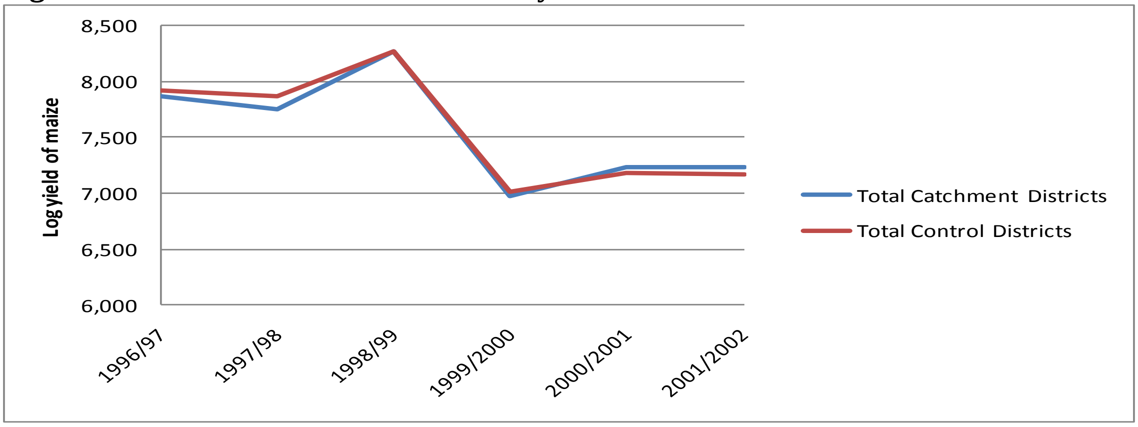
Table 7: Percentage of Households that Grow Maize, 1997 – 2002

District	1996/97	1997/98	1998/99	1999/2000	2000/2001	2001/2002
Chadiza (301)	100,00%	100,00%	98,88%	99,00%	97,73%	97,00%
Chipata (303)	99,34%	96,95%	95,39%	99,11%	99,35%	97,88%
Katete (304)	99,49%	98,48%	98,49%	98,64%	98,91%	98,58%
Lundazi (305)	99,55%	95,56%	95,63%	94,62%	94,85%	96,55%
Petauke (308)	99,25%	98,09%	97,42%	98,44%	99,24%	99,34%
Total Catchment Districts	*99,45%*	*97,47%*	*96,79%*	*97,90%*	*98,14%*	*98,01%*
Chama (302)	100,00%	94,44%	97,37%	97,50%	100,00%	100,00%
Mambwe (306)	100,00%	90,91%	97,06%	94,92%	98,04%	98,31%
Nyimba (307)	100,00%	100,00%	100,00%	98,33%	100,00%	100,00%
Total Control Districts	*100,00%*	*94,53%*	*98,16%*	*96,98%*	*99,43%*	*99,49%*
Total Eastern Province	99,51%	97,16%	96,97%	97,77%	98,32%	98,22%

Source: Author's calculations based upon PHS 1996/1997 – 2001/2002.

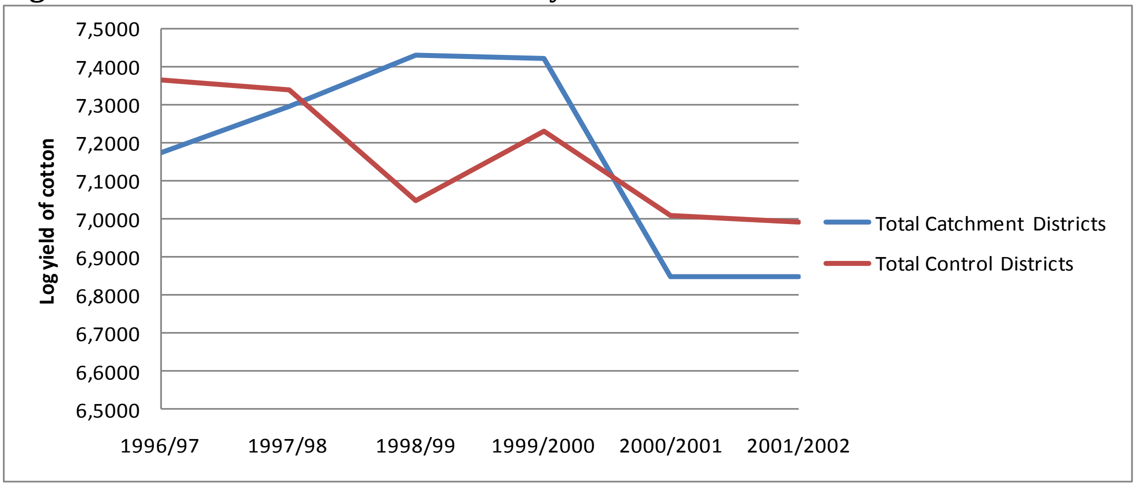
The second critical assumption of our **difference-in-difference model** is that the EPFRP implementation didn't affect maize productivity. By including measures of labour, fertilizers and land allocation in the observed covariates **x** of the regression, these direct and indirect effects will be accounted for.

Figure 1a: Trends in Maize Productivity EPFRP Catchment vs. Control Districts



Source: Authors' estimations based on the Post Harvest Survey.

Figure 1b: Trends in Cotton Productivity EPFRP Catchment vs. Control Districts



Source: Authors' estimations based on the Post Harvest Survey.

There is the possibility that the EPFRP affected maize productivity and that $\boldsymbol{\alpha_1}$ is a measure of the impacts of the EPFRP on cotton productivity relative to maize productivity. This possibility is ruled out from the plot (**figure 1a)** from where it can be seen that the parallel trends in maize productivity holds in our case between those districts that were affected versus those districts that were not affected by the EPFRP. Overall, this indicated that **the differencing** will *identify the impact of the EPFRP on cotton productivity only*.

The heterogeneity of the suitability of the land for cotton production and know-how of cotton husbandry leads to different entry-exit decisions regarding cotton production, which alters the composition of the group of farmers that produce cotton in each of the EPFRP phases. The estimates of the changes in productivity at the aggregate level comprise both the changes in productivity at the farm level and the changes in the composition of the farmers that produce cotton in each time period consistent estimation of the changes in productivity at the farm level requires that we *control for entry and exit* (Brambilla and Porto, 2005).

If there are fixed costs in cotton production, then cotton will only be profitable if productivity is high enough. This means that there is *a cut-off* (which depends on prices, market conditions, and infrastructure) such that farmers with productivity above this cut-off will *enter* the market and farmers below the cut-off will *not enter (or exit*, if they were in the market already). In consequence, measures of productivity that do not control for these *dynamic effects* may be artificially high (thus leading to downward biases in the estimates of productivity declines) (Brambilla and Porto, 2005).

Brambilla and Porto(2005) extend the *Industrial Productivity Analysis* literature by developing a method to deal with *entry and exit* in the estimation of agricultural production functions and crop choices. Furthermore, whereas this literature relies on *longitudinal surveys*, our method in line with that used by Brambilla and Porto(2005) can be used in *repeated cross-sections*.

Brambilla and Porto's solution to this problem is to construct *proxies* for the unobserved productivity parameter. The method exploits the idea that since households with high *unobserved cotton-specific effects* – ϕ_{ht} – *(see above)* are more productive in cotton, they are also more likely to devote a larger share of their land to cotton production. This means that we could use *land cotton shares* as a proxy for the unobservable $\boldsymbol{\phi_{ht}}$ in **(5)**. In practice, *consistent estimation* requires that we purge these shares of the part explained by observed determinants of cotton choice.[22]

Let a^c_{ht} be the fraction of land allocated to cotton. A *general model* of these shares is

(6) $$a^c_{ht} = m_t(z_{ht}, \phi_{ht}),$$

where **z** is a vector of regressors which includes *EPFRP's district effects* that affect selection into cotton production. We use the district EPFRP dummy to capture access to market and local feeder road network that facilitates farmer participation in market cash agriculture. The function ***m*** allows regressors **z** and unobservables **ϕ** to affect the shares **a** non-linearly.

We begin by considering the simplest model with *a linear functional form*

(7) $$a^c_{ht} = z'_{ht}\gamma_t + \phi_{ht},$$

Estimation of **(7)** is straightforward, except for the fact that *the share of land devoted to cotton* is censored at zero. This means that OLS may be inconsistent. A simple solution is to implement a **Tobit** procedure. More generally, we explore a more *semi-parametric estimation* of **(7)** by using a **censored least absolute deviation (CLAD)** model. We note that provided the right specification for the model is used, consistency follows because the regressors **z** are exogenous to **ϕ**. This requires that family composition or farm size does not depend on unobservables such as cotton-specific ability or land quality. Importantly, since we use data on *all households* to estimate **(7)**, this equation does *not suffer from a selection problem* like the one we are attempting to control for in the productivity model (Brambilla and Porto, 2005).

The allocation of land to cotton depends on several factors that we need to account for. In particular, the selection into cotton depends on the EPFRP. This means that we should include **RTI** in **(7)**. Cotton choices depend on output and input prices, too. Unfortunately, we do not have information on *prices at the farm level*. To the extent that prices vary by time, or by district, however, we can account for them with *year or district dummies*. In practice, we estimate a different model like **(7)** in each of the six years from 1996/1997 to 2001/2002 (notice that $\boldsymbol{\gamma_t}$ is indexed by **t** in **(7)**). This means that we will not be able to separate the effects of the implementation of the EPFRP from the effects of changes in international prices on land allocation, but we will be able to control effectively for **ϕ** in the productivity model.

Finally, we note that *identification of ϕ* requires that the selection into cotton is affected by the same unobservables that affect cotton productivity. In principle, it would be possible to argue that there are additional unobservable factors that affect the selection into cotton. Plugging in *the estimates* of *ϕ* in **(5)**, the **productivity model** is:

(8) $$\ln y_{ht} = x'_{ht}\beta + \alpha_1 RTI_{ht} + b_0\hat{\phi}_{ht} + \tilde{\varepsilon}_{ht}$$

[22] Notice that omitting **ϕ** not only leads to *inconsistencies* because of the entry-exit effects, but also may induce correlation between some variables in the vector **x** and the error term in the difference-in-differences model. For example, the choice of inputs, such as labour or fertiliser use, will depend on **ϕ** (so that higher levels of unobserved productivity may be positively correlated with input use). The model in (7) takes care of these biases (ibid.).

This **modified difference-in-differences approach** is consistent with entry and exit into cotton farming according to Brambilla and Porto(2005).

4. Data

Agriculture censuses and farm structure surveys are useful sources for rural development analysis. While the former is carried out every 10 years in Zambia, the latter is done annually. Agriculture censuses have the same advantage as the population censuses in that they provide exhaustive results with detailed territorial breakdown (Karlsson and Berkeley, 2005). The Agricultural statistical system in Zambia has been producing both structural[23] and performance data.[24] After the census of 1971/72, CSO extended the surveys to cover the subsistence or smallholder sub-sector of agriculture. In 1985/86 the two types of surveys were renamed the Crop Forecasting Survey (CFS) and Post- Harvest Survey (PHS), respectively.[25] These surveys are conducted in an integrated manner and as the core of the National Household Survey Capability Programme (NHSCP), which has been implemented since 1983. However, The Agriculture and Environment Department of Zambia's CSO only have agricultural production data at the district level going back until 1995.

4.1. The 1990 & 2000 Censuses of Population, Housing and Agriculture

The 1990 Census of Population and Housing provides such information as was required to create a sampling frame for inter-censal agricultural surveys.[26] At the time of the 1990 census there were 57 districts in the country. The Census Supervisory Areas (CSAs) were demarcated within each district while the Standard Enumeration Areas (SEAs) were demarcated within a CSA. A geo-coding system was, hence, developed with each SEA having a unique 6-digit code (Kasali, 2002).[27] The sampling frame comprised 4,193 CSAs out of which 3,231 are rural and 962 are urban. Each CSA is made up of about 3 SEAs. Out of a total of 12,999 SEAs, a sample of 610 SEAs had been selected. The rural stratum had been allocated 349 SEAs.[28]

Following the 1990 Census of Population, Housing and Agriculture a Master Sample of agricultural SEAs was set up and this sample was used to collect *Census of Agriculture data* during the period 1990/91 to 1991/92 and during the PHSs of 1992/93 to 1999/2000. The CSO conducted the 2000 Census Mapping exercise from 1998 to 2000. At the time of this

[23] Structural data or basic agricultural statistics relate to characteristics of agricultural holdings that vary slowly over time (are normally collected in a Census of Agriculture).
[24] Performance data or current agricultural statistics relate to: prices, quantities of inputs and outputs; enterprise costs and returns; and net farm incomes are collected mainly from current (annual) agricultural surveys. CSO and MAFF have been collecting current agricultural statistics since 1964.
[25] Up to 1978/79 agricultural season, the survey was called the Agricultural and Pastoral Production Survey, later renamed in 1982/83 as the Early Warning and Agricultural Survey to encompass the Crop Forecasting and Post-Harvest stages of the agricultural season during which period the two different types of surveys were conducted.
[26] Although the statistical unit in the Census was the agricultural holding, the agricultural household was used to identify the holding. All the data was collected by interviewing head of households or responsible adults.
[27] Provinces were identified by a 1-digit code, districts by a 2-digit code, and both CSAs and SEAs by a 5-digit code.
[28] *The "modified equal allocation method"* has been used to allocate the SEAs to provinces. The method allocates Units equally across all the provinces by dividing the sample size by the number of provinces. Then, considering the population size, heterogeneity and homogeneity of the province, the probability proportional to size method yielded additions and subtractions to some provinces. The final results are somewhere between equal and proportional to size allocation (IB Thomsen, 1996). This has been done at all levels and it increases the probability of including even the remote areas in the sample (CSO, 1996e).

mapping exercise a number of new districts had been created. The total number of districts had as a result risen to 72. Parliamentary Constituency and Ward boundaries were also taken into account in demarcating the 2000 Census CSAs and SEAs (CSO, 2001; Kasali, 2002).

4.2. Review of Sample Design for 1996/1997 - 1999/2000 Post-Harvest Survey

A stratified multi-stage sample design was used for the Zambia PHS. The sampling frame was based on the data and cartography from the 1990 Census described above.

The primary sampling units (PSUs) were defined as the CSAs delineated for the census. The CSAs were stratified by district within province and ordered geographically within district. A total sample of 405 CSAs was allocated to each province and district proportionally to its size (in terms of households). A master sample of CSAs was selected systematically with probability proportional to size (PPS) within each district at the first sampling stage; the measure of size for each PSU was based on the number of households listed in the 1990 Census.

The secondary sampling unit (SSU) is the SEA, that is, the sampling areas defined as the segment covered by one enumerator during the census. One SEA was selected within each sample CSA with PPS for the survey. A new listing of households was conducted within each sample SEA, and the farm size was obtained for each farm household. The listed households within each sample SEA were then divided into two groups based on farm size: **Category A** for households with less than 5 hectares (ha.) and **Category B** for households with 5 or more has (**table 8**).[29]

Table 8: Frequency of Holdings in Eastern Province, 1996-2002

	PHS 1996/97		PHS 1997/98		PHS 1998/99		PHS 1999/2000		PHS 2000/01		PHS 2001/2002	
	Frequency	Percent	Frequency	Percent	Frequency	Percent	Frequency	Percent	Frequency	Percent	Frequency	Percent
A-Small scale holding	956	78	1052	88	1111	88.5	1233	85.8	1060	84.9	1128	87.3
B-Medium scale holding	256	22	144	12	144	11.5	204	14.2	189	15.1	164	12.7
Total	1225	100	1196	100	1255	100	1427	100	1249	100	1292	100

Source: Author's calculation.

The original sampling plan was to select 10 households from each category within the sample SEA, for a total sample of 20 households per SEA. However, it was found that most sample SEAs had less than 10 households in Category B. In order to ensure a sample of 20 households within each sample SEA, the remaining households were selected from Category A (Megill, 2000; Megill, 2004).[30]

Given that a large majority of the rural households in Zambia are involved in agriculture, the sample of farm households is effective for most types of agricultural characteristics being measured by the PHS. In fact, the purpose of the PHS has been to

[29] Farm size for the small and medium farmers is difficult to obtain because the claims individual households have to land are not exclusive given that land is customarily owned. The problem is more acute in locations where **the Chitemene farming system** is practiced. *Chitemene System* involves the cutting of tree branches in a field to be used for crop production. CSO uses *the land under crops* to determine the extent of farm size.

[30] Following the data collection for the 1997/98 PHS, it was found *that more than 60 percent of the households selected in category B* actually had less than 5 HAs, according to the survey data. This is due to changes in the plans of individual households in the amount of land planted in crops, as well as non-sampling error in the listing data. Even with the current level of misclassification, this farm size stratification increases the sampling efficiency for producing estimates of total crop area and production.

capture relevant data from, and keep abreast with the changes occurring in the agricultural sector (**figures 1a-b**).

Specifically, the objectives of the PHS include provision of actual figures pertaining to: Area planted to individual crops (land usage - allocation); Realised Production quantities (output in physical units); Sales of produce and income realized; Numbers of livestock and poultry; Purchase and use of agricultural inputs; Capital formation and other operational expenses; Demographic characteristics of heads of rural households; Farming practices and soil conservation methods used; Access to agricultural loans; and, access to market prices information and agricultural extension services in general. The reference period for this information is the *agricultural season* starting 1st October ending 30th September.

However, the PHS estimates for some crops which are rare or limited to particular geographic areas have relatively high sampling errors.[31] In order to *evaluate the effectiveness of the PHS sample design* in meeting these survey objectives Megill(2000) use a CENVAR (Census Variance Calculation System) software to tabulate the standard errors for key survey estimates from the 1997/98 PHS data. The CENVAR results found by Megill(2000) illustrate that the main limitation of the sample design was that it didn't provide reliable results for rare crops. Moreover, over the period during which the PHSs have been conducted, *the survey questionnaire* has undergone several major revisions and differences in questions asked.

4.4.3. Review of Sample Design for 2000/2001 - 2001/2002 Post-Harvest Survey

The **PHS 2001/2002** also covered the whole country and was conducted in a sample of areas numbering 407 SEAs drawn using PPS sampling scheme, representing a sample proportion of about 5%. The survey was conducted in the same CSA and SEAs selected over the previous 4-5 years. The survey relied on the previous listing of household populations in **PHS 1999/2000** but with a new sample drawn from this listing.

Drawing on the experiences from the Census of Agriculture and the three PHSs that followed, it was realized that estimates for minor crops such as rice, sorghum, *cotton*, and tobacco were far from being satisfactory. Because of this, it became necessary to revisit the area frame in order to address the situation. In order to try and improve on the estimates for minor crops it was decided to create *Crop Zones* for these crops. In doing so a number of strata (Zones) were created in order to improve precision and accuracy in the estimates for minor crops.

In each district, the allocated sample size was shared proportionately among the crop strata, i.e., the more SEAs a crop stratum had the larger its share of the sample. This was done whilst ensuring that a minimum of two SEAs was selected from each stratum to facilitate computation of sampling error of the estimates.

Since the selection of participants in the PHS 2001/02 survey was not done with a simple random sample, a weight variable is used for our analysis.[32] We use the overall household weight.[33] The District level weight is developed using: The proportion of households who produce the crop in the district, the number of sampled SEA within each CSA, the number of households in the CSA, the number of households in SEA (all from

[31] The definition of *in-scope farm households* for the survey should also be examined. Megill recommends certain modifications to the sample design for *improving the sampling efficiency for future surveys*.

[32] The WGT variable in the *ID.dta* file is the appropriate weight to use. Another file has been created that contains the weighting value for specific crops. That file is called *cropwgt.dta*.

[33] The Weights (Boosting Factors) are the inverse of the probability that a given household has of being included in the sample. These factors are developed at the SEA level for each category of farmer.

Agricultural Census Survey Data). The district level weight is simply the probability that the number of households in a SEA will be selected as a primary unit from within a CSA within a particular District. After obtaining a complete list of the households in the SEA categorized as small or medium scale and the number of households to be sampled in each SEA, *the SEA level weight* is estimated. So with the District Level and SEA level weights, these two are multiplied and the product is the boosting factor.

Table 9: Post Harvest Survey, Sample sizes by District, 1997-2002

District	1996/1997	1997/1998	1998/1999	1999/2000	2000/2001	2001/2002
Chadiza (301)	96	88	89	100	88	100
Chipata (303)	303	295	304	338	307	330
Katete (304)	198	198	199	220	184	212
Lundazi (305)	224	225	229	260	233	261
Petauke (308)	267	262	271	320	262	305
Total Catchment Districts	***1088***	***1068***	***1092***	***1238***	***1074***	***1208***
Chama (302)	37	36	76	80	70	77
Mambwe (306)	52	55	34	59	51	59
Nyimba (307)	48	37	53	60	54	59
Total Control Districts	***137***	***128***	***163***	***199***	***175***	***195***
Total	**1225**	**1196**	**1255**	**1437**	**1249**	**1403**

Source: Authors' calculations based on CSO's Post Harvest Surveys 1997-2002.

The number of sample household in Eastern Province selected during the period 1996/97 – 2001/2002 was on average 1,274 households. They were interviewed during the period December and January using personal interviews with qualified respondents in sample households in sample areas (**table 9**). All PHSs were independent farm surveys and thus interviewed different households in each year. Consequently it is **not possible to construct a panel** of households using PHSs surveys in order to examine the correlates and causes of changes in the agricultural productivity of individual households over time (McCulloch, 2001).

5. Estimation Results and Discussion

This section focuses on the extent to which the productivity of cotton production in Zambia's Eastern Province from 1996/1997 to 2001/2002 is a result of the combined effects of the 1992 radical agricultural market liberalization and the subsequent rehabilitation of the feeder road network in Eastern Province in the period from 1996 to 2001.[34]

5.1. Descriptive Statistics

We are interested in measuring the impact of rural road interventions (i.e. access to local infrastructure and public goods and capital) on cotton yields per hectare (i.e. farm productivity).

Table 10 Descriptive Statistics, 1996/1997 – 2001/2002

		1996/1997		1997/1998		1998/1999		1999/2000		2000/2001		2001/2002	
		Full Sample		Full Sample		Full Sample		Full Sample		Full Sample		Full Sample	
Variable	Variable	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
Dependent variable	Volume of cotton production per hectare produced (MT)	1,33	2,31	1,48	2,09	1,62	3,06	1,64	3,02	0,97	0,68	0,97	0,68
	Log of cotton output (in kg) per hectare	6,54	1,10	6,83	0,96	6,75	1,17	6,55	1,40	6,65	0,71	6,64	0,71
Household determinants	Age of the household head	46,7	15,0	44,4	15,2	45,5	15,3	43,0	14,3	45,7	14,7	45,3	14,7
	Age Square of the household head	2404,0	1506,1	2205,4	1537,9	2307,8	1535,0	2056,0	1371,5	2309,7	1465,6	2270,4	1459,2
Household demographics	Size of the household	5,8	3,2	5,7	3,0	5,94	3,20	6,17	3,43	5,97	2,95	6,34	2,93
	Log of Size of the household	1,61	0,59	1,59	0,56	1,63	0,59	1,67	0,56	1,66	0,54	1,73	0,50
	Household category (stratum)	1,22	0,41	1,12	0,33	1,11	0,32	1,14	0,35	1,14	0,35	1,13	0,33
	Number of males in household	2,79	1,82	2,74	1,85	2,94	1,99	3,08	2,24	2,98	1,93	3,18	1,86
	Number of females in household	3,03	1,97	2,91	1,79	2,99	1,85	3,09	1,92	2,98	1,73	3,16	1,81
	Sex of head of household	1,23	0,42	1,23	0,42	1,24	0,43	1,24	0,43	1,25	0,43	1,25	0,44
Input use	Basal Quantity used (kg)	29,93	123,90	30,88	121,42	39,63	145,91	47,77	129,59	32,81	149,51	34,79	149,91
	Topdressing Quantity used (kg)	27,18	104,57	30,50	122,82	38,71	127,18	45,80	118,37	31,98	145,77	33,69	147,14
	Basal Fertilizers Used per cultiv. Area (kg per ha)	11,53	36,76	13,05	38,21	16,56	42,32	22,00	53,14	17,17	50,91	16,09	41,98
	Top Dressing Fertilizers Used per cultiv. Area (kg per ha)	10,43	28,63	13,32	41,86	16,74	38,74	21,01	47,71	16,10	40,45	15,56	37,37
	Value of Basal quantity used - (ZMK)	31920,3	92680,6	22202,3	238505,6	24409,9	87229,3	34564,7	95575,4	n.a.	n.a.	n.a.	n.a.
	Value of Topdressing quantity used - (ZMK)	27770,4	80701,8	23052,7	241685,4	25208,1	89689,2	33167,1	86535,9	n.a.	n.a.	n.a.	n.a.
	Expenditure on Basal fertilizers per cultivated area (ZMK/Ha)	12152,0	26389,7	7133,8	26384,2	10491,0	28902,8	15823,0	38979,2	n.a.	n.a.	n.a.	n.a.
	Expenditure on Topdressing fertilizers per cultivated area (ZMK/Ha)	10284,9	19167,9	8317,9	42566,3	10934,1	26732,4	15274,7	35724,6	n.a.	n.a.	n.a.	n.a.
Assets	Number of ploughs	0,374	0,865	0,29	0,77	0,30	0,77	0,27	0,65	n.a.	n.a.	n.a.	n.a.
	Number of draught animals	0,649	1,741	0,54	1,45	0,57	1,55	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
	Number of ploughs per household member	0,062	0,159	0,05	0,13	0,05	0,13	0,04	0,11	n.a.	n.a.	n.a.	n.a.
	Number of draught animals per household members	0,099	0,260	0,09	0,25	0,09	0,27	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
	Size of the land allocated to cotton	0,13	0,21	0,12	0,21	0,11	0,20	0,07	0,16	0,10	0,19	0,10	0,18
	Total area under crops (ha)	1,97	1,77	1,86	1,74	1,87	1,96	2,10	2,06	1,73	1,65	1,83	1,74
	Cultivated land per household member (ha)	0,38	0,33	0,37	0,33	0,35	0,32	0,39	0,45	0,34	0,37	0,36	0,31
	Livestock raising	0,58	0,49	0,48	0,50	0,48	0,50	0,50	0,50	0,55	0,50	0,47	0,50
	Usage of animal draught power for land preparation	0,27	0,45	0,25	0,43	0,24	0,43	0,28	0,45	0,35	0,48	0,35	0,48
	Received agricultural loan	0,323	0,468	0,265	0,441	0,32	0,47	0,16	0,37	n.a.	n.a.	n.a.	n.a.
EPFRP	*Rural transport infrastructure dummy (EPFRP)*	**n.a.**	**n.a.**	**n.a.**	**n.a.**	**0,84**	**0,37**	**0,83**	**0,37**	**0,83**	**0,37**	**0,83**	**0,37**
Aggregate agricultural - Year effects -	Length of Roads Network per total area of District (km / km2)	7,47	4,32	7,47	4,32	7,47	4,32	7,47	4,32	7,47	4,32	7,47	4,32
	Cotton-specific effect (OLS fitted values)	0,148	0,049	0,146	0,048	0,118	0,055	0,121	0,057	0,122	0,042	0,113	0,046
Agricultural extension services	Information on marketing for agricultural products	0,46	0,50	0,39	0,49	0,33	0,47	0,30	0,46	n.a.	n.a.	n.a.	n.a.
	Use any of the advice received on Crop husbandry	0,28	0,45	0,20	0,40	0,20	0,40	0,01	0,10	n.a.	n.a.	n.a.	n.a.
	Use any of the advice received on Crop diversification	0,23	0,42	0,12	0,32	0,16	0,37	0,14	0,35	n.a.	n.a.	n.a.	n.a.
	Information on agricultural input supply	0,41	0,49	0,35	0,48	0,32	0,47	0,23	0,42	n.a.	n.a.	n.a.	n.a.
Geographic Variables	Proportion of sample in Catchment Areas	0,85	0,36	0,85	0,36	0,84	0,37	0,84	0,37	0,83	0,37	0,83	0,37
	Proportion of sample in Control Areas	0,15	0,36	0,15	0,36	0,16	0,37	0,16	0,37	0,17	0,37	0,17	0,37
	Distance to the nearest all-weather road	1,374	0,603	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
	Distance to the nearest input market	1,855	0,784	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
	Rainfall	831,5	122,9	716,0	81,4	788,2	148,1	667,1	93,8	980,1	203,6	723,7	89,4
Cotton Observations		421		378		388		279		467		492	
Total number of Observations		1219		1197		1255		1427		1249		1403	

Source: Authors' estimations based on PHS.

[34] Grain marketing was liberalized in 1992 by the MMD government barely one month after coming to power, the same year that financial liberalization occurred.

This outcome of interest is a continuous variable with a mean ranging from 6.54 in 1996/1997 to 6.83 in 1997/1998 and a standard deviation from 0.71 in 2001/2002 to 1.40 in 1999/2000. The treatment EPFRP variable is discrete and of on/off variety (**table 10**).

Distribution of the Treatment and Comparison Samples

The sample characteristics of the comparison group and the treatment group highlight the role of randomization in the sense that the distribution of the covariates for the treatment and control groups are not significantly different. *The age of the head of household* in 1996/97 was only 2 years higher in the catchment districts, whereas in 2001/2002 it was almost similar. *The size of the household* was likewise equivalent in both 1996/97 and 2001/2002, although a bit higher in the catchment areas in entire period, exclusive in 1998/1999. The same could be said about *the number of males in the household* with the number in the catchment areas again being slightly higher (**table 10**). A more synoptic way to view these differences is to use the estimated **propensity score** as a summary statistic.

5.2. Evaluation of the EPFRP's impact on Cotton Productivity

The standard problem in **treatment evaluation** involves the inference of a causal connection between the treatment and the outcome. In our single-treatment case in each cross-section we observe (y_i, x_i, D_i; $i = 1, \ldots, N$) the vector of observations on the scalar-valued outcome variable **y**, a vector of observable variables **x**, a binary indicator of a treatment variable **D**, and let **N** denote the number of randomly selected individuals who are eligible for treatment. Let $\mathbf{N_T}$ denote the number of randomly selected individuals who are treated and let $\mathbf{N_{NT}} = N - N_T$ denote the number of non-treated individuals who serve as a potential control group.

We would like to obtain a measure of the impact of the EPFRP intervention in **D** on **y**, holding **x** constant. The situation is akin to one of missing data, and it can be tackled by methods of causal inference carried out in terms of (policy-relevant) **counterfactuals**. We ask how the outcome of an average untreated individual agricultural household would change if such a person were to receive the treatment. That is, the magnitude $\mathbf{\Delta y/\Delta D}$ is of interest. Fundamentally our interest lies in the outcomes that result from or are caused by the EPFRP interventions. Here the causation is in the sense of ceteris paribus (Cameron and Trivedi, 2005).

The average selection bias is the *difference* between programme participants and nonparticipants in the base state. *Selection bias* arises when the treatment variable is correlated with the error in the outcome equation (Baltagi, 2001). In our observational data the problem of selection of observables is solved in the subsequent sections using regression and matching methods.

Differences-in-Differences Estimators

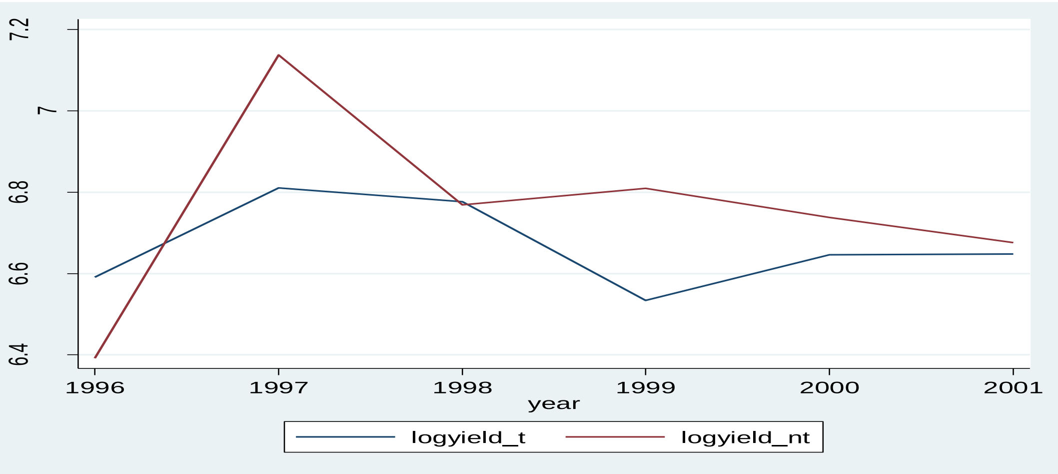
In our case we have data on the treated and the comparison (control) groups both before and after the experiment (i.e. implementation of the EPFRP). One way to improve on the one-group before and after design, which makes the strong assumption that the group remains comparable over time, is to include an additional untreated comparison group, that is, one not impacted by policy (i.e. EPFRP), and for which the data are available in both periods (Cameron and Trivedi, 2005).

Since the work by Ashenfelter and Card(1985), the uses in differences-in-differences methods has become very widespread. The first-differences estimator for the fixed effects model reduces to a simple estimator called the differences-in-differences estimator. The latter estimator has the advantage that it can also be used when *repeated cross-section data* rather than panel data are available. However, it does rely on model assumptions that are often not made explicit (Blundell and MaCurdy, 1999; Cameron and Trivedi, 2005).

In the context of the analysis of our experimental data *the simple comparison of the mean of the outcome* in treatment and control groups (**the 'differences' estimator**) is justified on the grounds that *the randomization* guarantees they should not have any systematic differences in any other pre-treatment variable.[35] In *the absence of treatment*, the unobserved differences between treatment and control groups are the same over time. In this case one could use data on treatment and control group before the treatment to *estimate the 'normal' difference* between treatment and control group and then compare this with the difference after the receipt of treatment. This removes biases in second period comparisons between the treatment and control group that could be the result from *permanent differences* between those groups, as well as biases from comparisons over time in the treatment group that could be the result of *trends*.

The validity of the differences-in-differences estimator is based on the assumption that the *underlying 'trends' in the outcome variable is the same* for both treatment and control group. This assumption is never testable and with only two observations one can never get any idea of whether it is plausible. But, from **figures 2a-b** depicting more than two observations we can get some idea of its plausibility.

Fig. 2a: 'Trends' in the Log of Cotton Yield for Treatment & Control group, 1996-2001

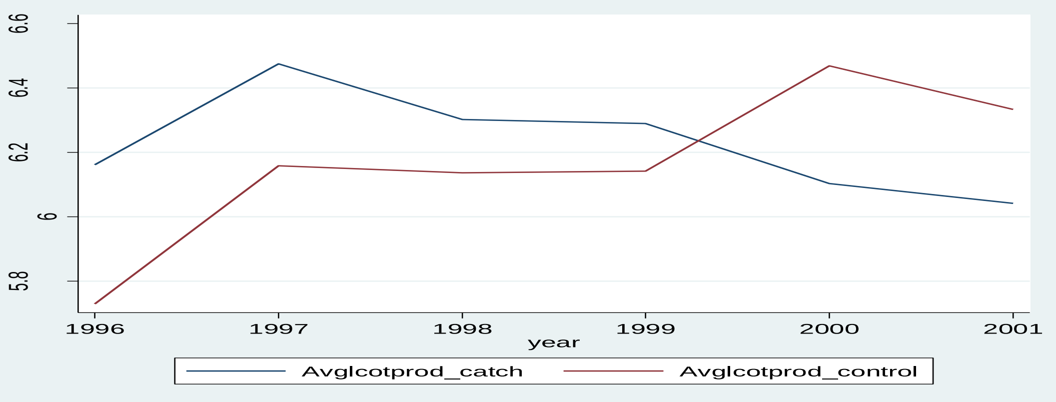


Source: Author's estimations.

We have two time periods, 1996/1997-1997/1998 (i.e. the pre-EPFRP treatment period) and 1998/1999 – 2001/2002 (i.e. the post-EPFRP treatment period), which straddle the policy change. We include a time period dummy variable for the second (post-policy change) time period 1998/1999 – 2001/2002, which we denote D2 to account for aggregate changes (factors) that affect the dependent variable over time in the same way for the two groups, the neighboring three **control districts** (no.: 302, 306-307) and the five similar **treatment districts** (no.: 301, 303-305, and 308). Define i = 0 for the control group and i=1 for the treatment (catchment) group. Define t = 0 to be a pre-treatment period and t=1 to be the post-treatment period.

[35] In economic applications treatment and intervention usually mean the same thing (op.cit., p.860). The term outcome refers to changes in economic status or environment on economic outcomes of individuals (ibid.).

Fig. 2b: 'Trends' in the log of Cotton Production for treatment & control group, 1996-2001



Source: Author's estimations.

The dummy variable (EPFRP) **Treatment** equals unity for those in the treatment group and is zero otherwise. The equation for analyzing the impact of the EPFRP treatment policy change known as the *pooled cross-section time-series model or constant-coefficient model* is:[36]

(9) $$y_{it} = \beta_0 + \delta_0 D2 + \beta_1 \text{ EPFRP} + \delta_1 D2*\text{EPFRP} + \beta_i X_{it} + u_{it},$$

where **y** is the outcome variable (log yield) of interest. The presence of **EPFRP** by itself captures possible differences between the treatment and control groups before the policy change occurred. The coefficient of interest, $\boldsymbol{\delta_1}$, multiplies the interaction term, **D2*EPFRP**, which is simply a dummy variable equal to unity for those observations in the treatment group in the second period.

X comprise the additional covariates in equation **(9)**, which account for the possibility that the random samples within a group have systematically different characteristics in the two time periods.

Define $\boldsymbol{\mu_{it}}$ to be the *mean* of the outcome in group **i** at time **t**. The **difference estimator** simply uses *the difference in means* between treatment and control group post-treatment as *the estimate of the treatment effect* i.e. it uses a estimate of ($\boldsymbol{\mu_{11} - \mu_{01}}$) (**table 11**). However, this assumes that the treatment and control groups have no other differences apart from the treatment, a very strong assumption with non-experimental data. A weaker assumption is that any *difference in the change in means* between treatment and control groups is the result of the treatment i.e. to use an estimate of:

(10) $$\hat{\delta}_1 = (\mu_{11} - \mu_{01}) - (\mu_{10} - \mu_{00}) = (\overline{y}_2^{tr} - \overline{y}_1^{tr}) - (\overline{y}_2^{nt} - \overline{y}_1^{nt}) = \Delta\overline{y}^{tr} - \Delta\overline{y}^{nt}$$

as an estimate of the treatment effect – this is the ***differences-in-differences estimator (DID)***.[37] The *DID estimator* is the OLS estimate of $\boldsymbol{\delta_1}$, the coefficient on the interaction

[36] In the statistics literature the model is called a *population-averaged model*, as there is no explicit model of y_{it} conditional on individual effects. Instead, any individual effects have implicitly been averaged out. The random effects model is a special case where the error u_{it} is equicorrelated over t for given i (Cameron & Triverdi, 2005:720).

[37] Since one estimates the time difference for the treated and untreated groups and then takes the difference in the time differences (Cameron & Trivedi, 2005:769).

between EPFRP and D2. This is a dummy variable that takes the value one only for the treatment group in *the post-treatment period*. In practice we write the *DID estimator* $\hat{\delta}_1$ as $(\mu_{11} - \mu_{01}) - (\mu_{10} - \mu_{00})$ note that *the first term* is the change in outcome for the treatment $\Delta\bar{y}^{tr}$ and the second term the change in outcome for the control group $\Delta\bar{y}^{nt}$ and then estimate the model **(9)**.

By comparing the time changes in the means for treatment and control groups, both group-specific and time-specific effects are allowed for. Nevertheless, *unbiasedness of the DID estimator* still requires that the policy change not be systematically related to other factors that affect **y** (and are hidden in **u**). The estimated equation (**column 4** in **table 11)** is

(9b)

$$\log(\hat{yield}) = \underset{(0.2364)}{7.201^{***}} + \underset{(0.0777)}{-0.0394\,\text{D2}^{*}} + \underset{(0.0035)}{-0.011\,\text{Treatment}^{**}} + \underset{(0.1090)}{0.0598\text{D2 x Treatment}} + \ldots$$

Therefore, $\hat{\delta}_1$ = 0.0598 (t= 0.55), which implies that the average cotton yield in Zambia’s Eastern Province increased by about 6 percent due to improved rural transport infrastructure development. The coefficient on **D2** is both fairly small and statistically insignificant. The coefficient on the **EPFRP Treatment** is also negative at the same level, but statistically significant at the 1 percent level. Thus, in the absence of change in treatment districts, the treatment districts actually saw *a fall* in cotton yield over time, which corresponds to **figures 2a-b** above.

(11) $\Delta y_i = y_{i1} - y_{i0}$,

Equation **(11)** is simply ***the difference estimator*** applied to differenced data.[38] The treatment effect **δ** can be consistently estimated by pooled OLS regression of Δy_{it} on ΔX_{it} and a full set of time dummies (**column 1** in **table 11**) (Cameron and Trivedi, 2005;Wooldridge, 2002).

[38] The individual-specific fixed effects α_i is eliminated by *first differencing* of the fixed effects model for $y_{it} = \phi D_{it} + \delta_t + \alpha_i + \varepsilon_{it}$.

Table 11 Basic Log-Productivity & Log-Production Regressions

		Productivity (Kg/HA)				Production (Kg)			
		Simple Difference		Difference-in-Differences Estimator		Simple Difference		Difference-in-Differences Estimator	
		Model 1	Model 2	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
		(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Type	**Independent Variables**	**dlogyield**	**dlogyield**	**logyield**	**logyield**	**dlcotprod**	**dlcotprod**	**lcotprod**	**lcotprod**
CV	Age of Head of HH	0.0106	0.0096	0.0070	0.0054	0.0175*	0.0183*	0.0152	0.0153
		(0.0085)	(0.0086)	(0.0085)	(0.0086)	(0.0100)	(0.0101)	(0.0101)	(0.0101)
CV	Age Squared	-0.0002*	-0.0001*	-0.0001	-0.0001	-0.0002*	-0.0002**	-0.0002*	-0.0002*
		(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
DV	Sex	0.1179**	0.1148**	0.1017*	0.1026*	0.2415***	0.2406***	0.2200***	0.2198***
	(M=1; F=0)	(0.0563)	(0.0561)	(0.0568)	(0.0565)	(0.0632)	(0.0632)	(0.0642)	(0.0643)
CV	Household Size	0.0012	0.0021	0.0027	0.0020	0.0342***	0.0344***	0.0367***	0.0367***
		(0.0076)	(0.0076)	(0.0077)	(0.0077)	(0.0080)	(0.0080)	(0.0081)	(0.0081)
CV	Share of Males in HH	-0.0669	-0.0746	-0.0846	-0.0989	-0.3063**	-0.3045**	-0.2950**	-0.2950**
		(0.1176)	(0.1177)	(0.1193)	(0.1190)	(0.1328)	(0.1327)	(0.1355)	(0.1355)
DV	Farm Type (Stratum)	-0.0811	-0.0845	-0.1108*	-0.1126*	0.4989***	0.5003***	0.4649***	0.4650***
	(SSF=1; MSF=2)	(0.0586)	(0.0586)	(0.0590)	(0.0589)	(0.0651)	(0.0649)	(0.0655)	(0.0654)
CV	Livestock	0.0626	0.0617	0.0784*	0.0738*	0.2493***	0.2485***	0.2509***	0.2510***
		(0.0417)	(0.0416)	(0.0421)	(0.0419)	(0.0479)	(0.0479)	(0.0479)	(0.0479)
CV	Rainfall District Level	0.0006***	0.0004**	0.0001	0.0000	0.0004**	0.0005**	-0.0000	-0.0000
		(0.0002)	(0.0002)	(0.0001)	(0.0001)	(0.0002)	(0.0002)	(0.0001)	(0.0002)
DV	**EPFRP Treatment**	-871.9633***		-0.1605**		-96.4543		-0.1629*	
	(T=1; NT=0)	(237.0590)		(0.0760)		(304.8437)		(0.0930)	
CV	Cotton Landfraction	-1.3831***	-1.3180***	-1.3408***	-1.3625***	1.0681***	1.0667***	0.1045	0.1050
		(0.1047)	(0.0998)	(0.1053)	(0.1049)	(0.1330)	(0.1326)	(0.0927)	(0.0929)
DV	D2	0.0000	0.0000	-0.0192	-0.0394	0.0000	0.0000	0.0000	-0.1690
		(0.0000)	(0.0000)	(0.0774)	(0.0777)	(0.0000)	(0.0000)	(0.0000)	(0.1296)
DV	D2*Treatment	871.7184***	-0.0699	0.0000	0.0598	96.2213	-0.3147**		0.0003
		(237.0540)	(0.1182)	(0.0000)	(0.1090)	(304.8399)	(0.1393)		(0.0043)
CV	**EPFRP Treatment**		-0.0077**		-0.0105***		0.0039		0.0003
	(share)		(0.0037)		(0.0035)		(0.0044)		(0.0043)
	Constant	0.4334***	0.3847***	7.0582***	7.2007***	-0.5039***	-0.5057***	4.5598***	4.5561***
		(0.0419)	(0.0368)	(0.2364)	(0.2423)	(0.0531)	(0.0515)	(0.2816)	(0.2853)
	Observations	2364	2364	2364	2364	2073	2073	2073	2073
	R^2	0,09	0,09	0,08	0,09	0,13	0,13	0,12	0,12
	F	18,43	19,06	16,87	16,75	25,21	25,72	23,05	21,28
	ll	-3248,77	-3251,69	-3269,71	-3264,97	-2965,14	-2964,77	-2984,92	-2984,91

Notes: Robust Standard Errors in parentheses: * $p<0.10$, ** $p<0.05$, *** $p<0.01$.
Source: Authors' estimations.

Further Specification

Table 11 reports the productivity results corrected for entry and exit, where we consider the treatment variable as respectively a dummy variable (**Model 1**) and as a percentage share of the feeder road network (**Model 2**). Column (1) reproduces the ***Difference-in-Differences*** estimates from column (3) of **Table 11**, which does not include controls for **φ**.

A method to overcome the inconsistency of OLS is to follow ***Tobin***, because the cross-section of PHSs reveal that a significant proportion ranging from 63 to 79 percent of the households with *zero cotton production* (i.e. censoring observations) and the rest with positive levels of cotton productivity. The sample is therefore a mixture of observations with zero and positive values. The ***censored regression model, or Tobit model***, is relevant because the dependent *'logarithm of cotton productivity'* variable is observed only over some interval of its support.[39] In fact this is a case of left-censored PHS data with the censoring point (L=0). Thus, in **columns (3)** we estimate a linear regression in the presence of *censoring* by using *a Tobit model*.

In the face of the heteroskedasticity, the *Tobit procedure* yields estimates that are as biased up as OLS is biased down. Deaton(1997:88) warns that "there is no general guarantee that the attempt to deal with censoring by replacing OLS with the Tobit MLE will give estimates that reduce the bias."

Table 12: Tobit Model Comparisons of Log Likelihood

	Log Likelihood				
	Tobit	**Two-limit(i)**	**Two-part(ii)**	**Type-2 Tobit (iii)**	**Type-2 Tobit (iv)**
Model 1	-2959,10	-2900,07	-4970,67	-4963,34	-4963,34
Model 2	-2962,87	-2903,92	-4984,34	-4978,10	-4977,71

Notes: (i) We exclude the 5.3% of the sample (111 observations) where logyield exceeds 8.23 and which doesn't correspond well with the in sample-fitted values. (iii) A bivariate Heckman sample-selection model without exclusion restrictions. (iv) A Heckman bivariate sample-selection model with exclusion restrictions.

By comparison (**table 12**), the log likelihood for *the two-limit tobit model* fits the data considerably better than both the *simple tobit model* and the *two-part model* as well as the *Heckman sample selection models*.

As a useful guide to failure of homoskedasticity or normality, comparing the *Tobit* estimates with *Powell's* estimator is done in **Columns (4) (table 14)**, where we calculates Powell's (1984) ***censored least absolute deviations estimator (CLAD)*** and bootstrap estimates of its sampling variance.[40] 100 bootstrap replications are performed.[41] Unlike the standard estimators of the censored regression model such as tobit or other maximum likelihood approaches, ***the CLAD estimator*** is robust to heteroscedasticity and is consistent and asymptotically normal for a wide class of error distributions. Due to insufficient observations and non achievement of convergence the final specification omits some of the covariates: Sex; stratum; livestock and rainfall. Consequently, the treatment in either model is not significant.

[39] Cotton productivity in levels is very heavily skewed and has considerable nonnormal kurtosis. The logarithmic transformation reduced both the skewness and nonnormal kurtosis significantly.
[40] We choose the *bootstrap estimate* which assumes that the sample was selected in *two-stages* and which replicates the design by bootstrapping in two stages. An advantage of *the two-stage bootstrap estimates* is that if the sample was collected using a two-stage process, then the estimated standard errors will be robust to this *design effect*.
Kish (1995) and Cochran (1997) show the importance of correcting mean values for design effects.
Scott and Holt (1982) show that the magnitude of the bias for the estimated variance-covariance matrix for *OLS estimates* can be quite large when it is erroneously assumed that the data were collected using a simple random sample, if in fact *a two-stage design* had been used.
[41] Rogers (1993) shows that these standard errors are not robust to violations of homoscedasticity or independence of the residuals and proposes a bootstrap alternative.

Table 13 Cotton Yields: Impacts of the EPFRP Baseline Regressions

	Model 1			Model 2		
	Simple Model	Controlling for I and η		Simple Model	Controlling for I and η	
	(1)	(2)	(3)	(4)	(5)	(6)
	logyield	lnY	lnY_CM	logyield	lnY	lnY_CM
Age of Head of HH	0.0070	0.0187	0.0201	0.0057	0.0193	0.0200
	(0.0086)	(0.0131)	(0.0132)	(0.0086)	(0.0131)	(0.0132)
Age Squared	-0.0001	-0.0002	-0.0002*	-0.0001	-0.0002*	-0.0002*
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Sex	0.1019*	0.0163	0.0479	0.1024*	0.0035	0.0489
(M=1; F=0)	(0.0568)	(0.0835)	(0.0855)	(0.0564)	(0.0841)	(0.0856)
Household Size	0.0027	-0.0070	-0.0059	0.0021	-0.0059	-0.0060
	(0.0077)	(0.0104)	(0.0109)	(0.0077)	(0.0105)	(0.0109)
Share of Males in HH	-0.0847	-0.1635	-0.4234**	-0.0975	-0.1487	-0.4241**
	(0.1193)	(0.1842)	(0.1847)	(0.1192)	(0.1849)	(0.1848)
Farm Type (Stratum)	-0.1097*	-0.1694*	-0.1522*	-0.1104*	-0.1668*	-0.1522*
(SSF=1; MSF=2)	(0.0590)	(0.0898)	(0.0848)	(0.0590)	(0.0899)	(0.0848)
Livestock	0.0790*	-0.0044	-0.0763	0.0760*	0.0087	-0.0773
	(0.0420)	(0.0637)	(0.0618)	(0.0419)	(0.0642)	(0.0620)
Rainfall District Level	0.0001	-0.0003*	-0.0006***	0.0000	-0.0002	-0.0006***
	(0.0001)	(0.0002)	(0.0002)	(0.0001)	(0.0002)	(0.0002)
EPFRP Treatment	-0.1759***	0.2887***	-0.0185			
(T=1; NT=0)	(0.0417)	(0.0639)	(0.0585)			
Cotton Landfraction	-1.3377***			-1.3571***		
	(0.1032)			(0.1026)		
EPFRP Treatment				-0.0094***	0.0085***	-0.0005
(share)				(0.0018)	(0.0027)	(0.0026)
Constant	7.0440***	-0.4795	-0.1622	7.1622***	-0.5260	-0.1607
	(0.2359)	(0.3556)	(0.3645)	(0.2381)	(0.3573)	(0.3645)
Observations	2364	2221	1909	2364	2221	1909
R^2	0.0818	0.0148	0.0173	0.0854	0.0096	0.0173

Source: Authors' estimations.

In **Table 13** reports our benchmark results. Columns (1) and (4) report estimates of equation (**3)**, that is, **a simple model of cotton yields per hectare** that does not control for the unobservables (It, ηt, ϕt). There is evidence in both models in favor of decreasing returns to scale in cotton since there is *a negative association* between the size of land allocated to cotton and cotton yields. There is also negative association between farm size and cotton productivity and production. Age; household size and rainfall don't seem to matter. The dynamics of cotton yields are closely linked to the market accessibility through the EPFRP treatment. In both the simple models of cotton yields, the estimated magnitudes are both significant and larger than the other covariates except cotton land share.

Columns (2) and (5) report productivity results and columns (3) and (6) the production results both from equation **(5)**, controlling for agricultural effects ($\mathbf{I_t}$) and unobserved heterogeneity ($\boldsymbol{\eta_t}$). In **model 1** the estimated magnitude of the impact of the EPFRP treatment is likewise significant but the coefficient is both larger and its sign is correct in **column 2**. In **model 2** it is only the sign that is correct **column 5**. Using the logarithm of cotton production instead of productivity as the dependent variable doesn't seem to improve the results in both model specifications.

Table 14: Comparison of Cotton Productivity Estimation Models

	Model 1				Model 2			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
	DID	OLS	Tobit	CLAD	DID	OLS	Tobit	CLAD
Age of Head of HH	0.0070	0.0230	0.0340	-0.1091***	0.0054	0.0225	0.0322	0.1921***
	(0.0085)	(0.0179)	(0.0295)	(0.0247)	(0.0086)	(0.0178)	(0.0297)	(0.0719)
Age Squared	-0.0001	-0.0003	-0.0003	0.0011***	-0.0001	-0.0003	-0.0003	-0.0020***
	(0.0001)	(0.0002)	(0.0003)	(0.0003)	(0.0001)	(0.0002)	(0.0003)	(0.0007)
Sex	0.1017*	-0.1005	-0.1475		0.1026*	-0.1140	-0.1945	
(M=1; F=0)	(0.0568)	(0.1227)	(0.1968)		(0.0565)	(0.1232)	(0.1973)	
Household Size	0.0027	-0.0269*	-0.0145	-0.0254*	0.0020	-0.0254*	-0.0096	-0.0935*
	(0.0077)	(0.0143)	(0.0199)	(0.0137)	(0.0077)	(0.0143)	(0.0200)	(0.0514)
Share of Males in HH	-0.0846	-0.6342**	-0.5892	-1.1998***	-0.0989	-0.6285**	-0.5734	-0.1354
	(0.1193)	(0.2645)	(0.3855)	(0.3475)	(0.1190)	(0.2651)	(0.3880)	(0.8756)
Farm Type (Stratum)	-0.1108*	-0.3650***	-0.2274		-0.1126*	-0.3646***	-0.2301	
(SSF=1; MSF=2)	(0.0590)	(0.1183)	(0.1640)		(0.0589)	(0.1183)	(0.1651)	
Livestock	0.0784*	-0.0821	0.0932		0.0738*	-0.0719	0.1264	
	(0.0421)	(0.0856)	(0.1432)		(0.0419)	(0.0860)	(0.1436)	
Rainfall District Level	0.0001	-0.0010***	-0.0012**	-0.0008*	0.0000	-0.0010***	-0.0011**	0.0005
	(0.0001)	(0.0003)	(0.0005)	(0.0004)	(0.0001)	(0.0004)	(0.0005)	(0.0011)
EPFRP Treatment	-0.1605**	0.2317***	0.6506***	0.1105				
(T=1; NT=0)	(0.0760)	(0.0895)	(0.1350)	(0.1241)				
Cotton Landfraction	-1.3408***				-1.3625***			
	(0.1053)				(0.1049)			
D2	-0.0192				-0.0394			
	(0.0774)				(0.0777)			
D2*Treatment	0.0000				0.0598			
	(0.0000)				(0.1090)			
Cotton Landfraction		-3.8872***	-5.1098***	-7.5985***		-3.8924***	-5.1328***	-14.9180***
(Residuals: ϕ)		(0.2281)	(0.4106)	(0.8439)		(0.2289)	(0.4136)	(2.0747)
EPFRP Treatment					-0.0105***	0.0090**	0.0239***	0.0094
(share)					(0.0035)	(0.0045)	(0.0068)	(0.0133)
Constant	7.0582***	1.6479***	0.7817	4.4438***	7.2007***	1.6462***	0.8296	-3.4324**
	(0.2364)	(0.5302)	(0.8603)	(0.7830)	(0.2423)	(0.5378)	(0.8819)	(1.6898)
Observations	2364	1357	1357	219	2364	1357	1357	170
R^2	0.0818	0.1962			0.0855	0.1945		

Notes: (1) Difference-in-Differences Estimator; (2) OLS: Log-linear models for which the parameters need to be interpreted as semielasticities. (3) Tobit regression. (4) Clad calculates Powell's (1984) censored least absolute deviations estimator (CLAD) and bootstrap estimates of its sampling variance.
Source: Authors' estimations.

In **table 14**, we report the results with the entry and exit correction, thereby taking into account the *compositional effects* induced by entry and exit into cotton farming (Brambilla and Porto, 2007).

Column (1) reproduces the *DID estimates* from **column (3) of Table 11**, which does not include controls for ϕ. Columns (2) use a linear OLS model; columns (3) use a Tobit model to estimate *the selection equation*, and columns (4) use a CLAD model. **Model 1** and **Model 2** are the same as before. We find that in both models, the ***Tobit model*** outperforms the other models with regards to the magnitude and sign of the EPFRP coefficient (α). In all our specifications in columns (2) to (4) of **Table 14**, the estimates of $\mathbf{b_0}$ are similar to those from the DID model that does not correct for **ϕ** column (1).

The ***Tobit Model*** also outperforms two other non-linear models: The Partial Linear regression model with Yatchew's weighting matrix and the Stochastic Frontier model.

5.3. Robustness Checks

Specification tests and model diagnostic

Given the fact that the EPFRP wasn't implemented simultaneously in all five treatment districts, it may be difficult to justify assigning 1998 as the year where the entire project was implemented. To examine the robustness of the results, we re-estimate the model by first redefining the post-implementation phase to exclude the year 1998 and thus move it to the pre-implementation phase. Thus, we assign the first three years and the three years to two different phases of the EPFRP, where the first period is considered the *pre-implementation phase*, whereas the second period is considered the *(post-) implementation phase*.

Table 15: Sensitivity to the Definition of EPFRP and Reassignment of Implementation Year

	Tobit						
	Model 1		Model 2		Model 3		Model 4
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
EPFRP (T=1; NT=0)	0.6506***						
(Primo: 1998)	(0.1350)						
Cotton Landfraction	-5.1098***	-5.2231***	-5.1328***	-5.1719***	-5.3043***	-5.3917***	-5.1071***
(Residuals: ϕ)	(0.4106)	(0.3981)	(0.4136)	(0.3989)	(0.3984)	(0.4179)	(0.4105)
epfrp (T=1; NT=0)		1.6017***					
(Primo: 1999)		(0.1697)					
EPFRPpct (share) (i)			0.0239***				
(Primo: 1998)			(0.0068)				
epfrppct (share)				0.0751***			
(Primo: 1999)				(0.0085)			
D1998					0.0904		
(1998 year effect)					(0.1499)		
D1999					1.7143***		
(1999 year effect)					(0.1716)		
D301						0.1515	
(District 301 effect)						(0.8124)	
D302						0.1127	
(District 302 effect)						(0.8463)	
D303						0.5846	
(District 303 effect)						(0.7658)	
D304						0.6730	
(District 304 effect)						(0.7660)	
D305						-0.5912	
(District 305 effect)						(0.8016)	
D306						0.4568	
(District 306 effect)						(0.7995)	
D308						-0.1548	
(District 308 effect)						(0.7800)	
EPFRPPCT (share) (ii)							0.0225***
(Primo: 1998)							(0.0047)
Constant	0.7817	-0.3277	0.8296	-0.1387	-0.7530	2.2258*	0.6497
	(0.8603)	(0.8287)	(0.8819)	(0.8326)	(0.8378)	(1.2171)	(0.8679)
Observations	1357	1357	1357	1357	1357	1357	1357
ll	-1234.1097	-1201.8225	-1239.4838	-1207.9393	-1191.7112	-1225.7072	-1234.2839

Notes: Standard errors in parentheses * p<0.1, ** p<0.05, *** p<0.01.

Thus in **column 1** in **table 15** we insert the estimates from the ***Tobit model*** in **column 3** from **table 14** above. In **columns (2)** and **(4)** we show the Tobit results for the same definition of the treatment variable but with a readjusted implementation phase starting in 1999. This reassignment of the start year raises the magnitude of the treatment variable's coefficient (**α**) to an even higher level in both models. When looking at the district and year effects in **model 3** we only find that the 1999 year effect is significant at the 1% level, whereas year 2000 and 2001 are dropped due to collinearity. Finally, by looking at the EPFRP's percentage share of the primary feeder roads (**column 7**) instead of as a share of the entire feeder road network (**column 3**) we get more or less the same results, although a bit inferior with regards to the magnitude of the coefficients.

6. Conclusions and Policy Implications

This paper has investigated the dynamic impacts at the district level of the implementation of the EPFRP on farm cotton yield and production in Zambia's rural Eastern Province exclusively, contrary to Brambilla and Porto(2005, 2007) who cover all four cotton growing provinces in Zambia at the provincial level.

The key development objective of the EPFRP in our context was *'to improve access to the productive areas of the Province to ensure the introduction of farm inputs as well as the timely evacuation of harvested agricultural product.'* In addition to the *short-run effects* such as the creation of direct employment because of the use of labour-based technology, the EPFRP also had *medium to long-run effects* through reallocation of incumbent firms, entry and exit of firms, market creation and destruction (for inputs, outputs and credit), and contract enforcement mechanisms (i.e. the out-grower scheme).

To explore these market dynamics of the EPFRP we identify two phases of the EPFRP. Starting with a baseline period from July 1996 to 1997/1998, which in our study is the pre-implementation period, where the project focused on building the capacity of the districts and the private contractors to rehabilitate and maintain rural feeder roads using labour-based techniques.[42] The second (post-) implementation phase ran from 1998/1999 to 2001/2002. Only five districts (Chadiza, Chipata, Katete, Lundazi and Petauke) received capital assistance to carry out rehabilitation works. These five districts are considered our *treatment districts*. Moreover, at the mid-term evaluation carried out in July 1998 only 76 km of feeder road of the final 404 km had actually been rehabilitated to an excellent standard and had been maintained for one year, and another 196 km had received separate routine maintenance interventions.[43]

By following the Brambilla and Porto(2005, 2007) approach we have estimated the dynamic impacts of the implementation of the EPFRP by building *a model of cotton yields and crop choices*. We have compared average cotton yields across two phases, the pre-implementation phase and the post-implementation phase of the EPFRP, conditional on the aggregate trend in agricultural, observed covariates at the farm level, and unobserved farm effects like land quality and cropping ability.

To correct for compositional effects associated with entry and exit into cotton farming and with cotton-specific unobserved heterogeneity, we have introduced *a model of selection into cotton* that provides proxies for unobserved cotton productivity. Adapting techniques from the industrial organization literature, these proxies are given by *land cotton shares* (i.e., the shares of total land allocated to cotton) purged of the effects of observed covariates.

Our model thus provides an overall consistent estimator of the impacts of the EPFRP in the presence of confounding observed and unobserved effects and in the presence of compositional effects in cotton farming.

Unlike Brambilla and Porto(2007) who besides Eastern Province also include Central, Southern and Lusaka Provinces in their sample, we don't find an increase in cotton

[42] The project succeeded in training seven private, labor-based rehabilitation contractors; 15 private, labor-based maintenance contractors and 19 public officials—14 district supervisors.

[43] Considering that not all essential inputs have been in place at any one time (national staff, the finance company, office accommodation, supporting projects).

productivity in the last two years 2000/2001 to 2001/2002, but rather a stagnation after a sharp fall from 1999/2000 to 2000/2001. This is because as they likewise acknowledge that there are significant differences across these four provinces. These provincial differences could be due to differences in the out-grower schemes offered by different firms. This piece of information unfortunately isn't captured by the PHSs. We also find some evidence indicating that cotton yields followed different patterns in the treatment districts compared with the control districts.

Finally, notwithstanding the sensitivity of our results to the choice of model specification and the limitation of the PHS database, our findings tentatively seem to suggest that the EPFRP treatment had an effect that was significant. In other words, the EPFRP's improvement and maintenance of the feeder road network in Zambia's Eastern Province apparently may have contributed to the stimulation of the production of cotton (MT) and the yield (MT/HA). The concern is that the yield improvements weren't sustained beyond the short-term perhaps due to idiosyncratic factors in Eastern Province not captured by the models used, such as the incomplete implementation of the cotton out-grower scheme (cf. issue of side-selling); the non-competitive structure of the cotton market (Chauvin and Porto, 2011) etc.

In line with Brambilla and Porto(2007) we derive a number of lessons from our empirical analysis.

First, in 2005, most cotton production in Zambia was carried out under *the out-grower scheme*. Farmers and firms have understood the importance of honouring contracts and the benefits of maintaining a good reputation. The out-grower programs have been perfected and there are now two systems utilized by different firms: *the Farmer Group System* and *the Farmer Distributor System*. Both systems seem to work well (Balat and Porto, 2005a; Brambilla and Porto, 2005, 2012; Poulton et al., 2004; Tschirley et al., 2006).

Second, our results generate some evidence on how labour-based technology rural transport infrastructure projects at the regional level in combination with other effects such as the persistent effects from the prior liberalization of the agricultural sector affects yields at the farm level via input and output prices, credit, input use, technical advice, information and technology, and efficiency. In other words, unlike Brambilla and Porto(2007) who exclusively associate the impact with the economic reform effect, we believe that the agricultural trade liberalization and the privatization of the parastatals wouldn't have had a significant impact on the crop choice and the crop productivity if the EPFRP hadn't subsequently been implemented in Zambia's Eastern Province as an indispensable complementary policy for rural development.

Finally, Schulz and Bentall(1998) already noted in the mid-term evaluation of the EPFRP that the rate of implementation of the decentralization policy had a negative impact on the sustainability of the project results. Budget constraints forced the District Councils in the Eastern Province not to continue to carry out rehabilitation tasks after the project was completed and instead to concentrate on maintenance unfortunately not always on a routinely basis. Moreover, in the end only a total of 404km of feeder roads of the targeted 580km were completed within the project budget, which most likely diminished the EPFRP's impact on the cotton yield and production.

References

ABADIE, A., DRUKKER, D., HERR, J. L. & IMBENS, G. W. 2004. Implementing matching estimators for average treatment effects in Stata. *Stata Journal*, 4**,** 290-311.

BALAT, J. F., BRAMBILLA, I. & PORTO, G. G. 2004. An Analysis of the WTO Development Round on Poverty in Rural and Urban Zambia.

BALAT, J. F. & PORTO, G. G. 2005a. The WTO Doha Round, Cotton Sector Dynamics and Poverty Trends in Zambia. *Policy Research Working Paper Series,* 3697.

BALAT, J. F. & PORTO, G. G. 2005b. Globalization and complementary policies: poverty impacts in rural Zambia. Working Paper 11175**,** 42.

BALAT, J. F. & PORTO, G. G. 2006. The WTO Doha Round, Cotton Sector Dynamics, and Poverty Trends in Zambia. *In:* WINTERS, T. W. H. A. L. A. (ed.) *Poverty and the WTO. Impacts of the Doha Development Agenda.*: Palgrave MacMillan.

BALTAGI, B. H. 2001. *Econometric Analysis of Panel Data*, Chichester, John Wiley and Sons, Chichester, 1995, (2nd Edition, 2001),. .

BANISTER, D. & BERECHMAN, J. 2000. *Transport Investment and Economic Development*, Routledge (UK).

BELLEMARE, M. F. 2009. The Productivity Impacts of de Jure and de Facto Land Rights. *Duke University Working Paper*.

BLUNDELL, R. & MACURDY, T. 1999. Labor supply: A review of alternative approaches. *In:* ASHENFELTER, O. & CARD, D. (eds.) *In: Handbook of Labor Economics.* Elsevier

Brambilla, I. and Guido G. Porto. 2012. Market structure, outgrower contracts, and farm output. Evidence from cotton reforms in Zambia. Oxford Economics Papers. January 2012 64(1).

BRAMBILLA, I. & PORTO, G. G. 2005. Farm Productivity and Market Structure. Evidence From Cotton Reforms in Zambia. *Yale University Economic Growth Center Discussion Paper,* 919**,** 42.

BRAUTIGAM, D. 2009. *The Dragon's Gift. The Real Story of China in Africa*, Oxford University Press.

CAMERON, A. C. & TRIVEDI, P. K. 2005. *Supplement to microeconometrics: Methods and applications,* New York, Cambridge University Press.

Chauvin, N.D., and Guido Porto. 2011. Market Competition in Export Cash Crops and Farm Income. African Center for Economic Transformation and Universidad de La Plata.

CHIWELE, D. K., MUYATWA-SIPULA, P. & KALINDA, H. 1998. Private Sector Response to Agricultural Marketing Liberalisation in Zambia. A Case Study of Eastern Province Maize Markets. *In:* OLUKOSHI, A. (ed.) *Research report.* Uppsala: Nordiska Afrikainstitutet.

CLIFTON, J. 1998. Preparation of Enhanced Training Programme for District Staff Forming Key - Staff of Contract Management Units. *In:* HOUSING, M. O. L. G. A. (ed.). Lusaka: Consultancy Report.

CSO 2001. 2000 Census of Population and Housing. Preliminary Report. *In:* CSO, R.O.Z. (ed.) Preliminary report for the 2000 Census of Population and Housing Population count ed.

CSO 2011. 2010 Census of Population and Housing. Preliminary Report. Lusaka: Central Statistical Office.

DEHEJIA, R. H. & WAHBA, S. 2002. Propensity score-matching methods for Non-experimental causal studies. *The Review of Economics and Statistics,,* 84**,** 151-161.

DEICHMANN, U., FAY, M., KOO, J. & LALL, S. V. 2002. Economic structure, productivity, and infrastructure quality in southern Mexico.

DERCON, S. & HODDINOTT, J. 2005. Livelihoods, growth, and links to market towns in 15 Ethiopian villages *FCND Discussion Paper 194*.

DORWARD, A., KYDD, J., MORRISON, J. & UREY, I. 2004. A Policy Agenda for Pro-Poor Agricultural Growth. *World Development,* 32**,** 73–89.

DORWARD, A., KYDD, J. & POULTON, C. (eds.) 1998. *Smallholder Cash Crop Production under Market Liberalisation. A New Institutional Economics Perspective*: Cab International.

EASTERN PROVINCE CHAMBER OF COMMERCE AND INDUSTRY. 2005. *Investment Profile* [Online]. Chipata. Available: http://www.easternchamber.com/Investment/socio-economics.htm [Accessed 9th of June 2005].

ESCOBAL, J. & PONCE, C. 2003. The benefits of rural roads. Enhancing income opportunities for the rural poor. *Documentos de Trabajo dt40b, Grupo de Análisis para el Desarrollo (GRADE)***,** 56.

FAN, S. & CHAN-KANG, C. 2004. Road development, economic growth, and poverty reduction in China. *IFPRI DSGD Discussion Paper,* 12**,** 86.

GOLDSTEIN, M. & UDRY, C. 2008. The Profits of Power: Land Rights and Agricultural Investment in Ghana. *Journal of Political Economy,,* 116**,** 981-1022.

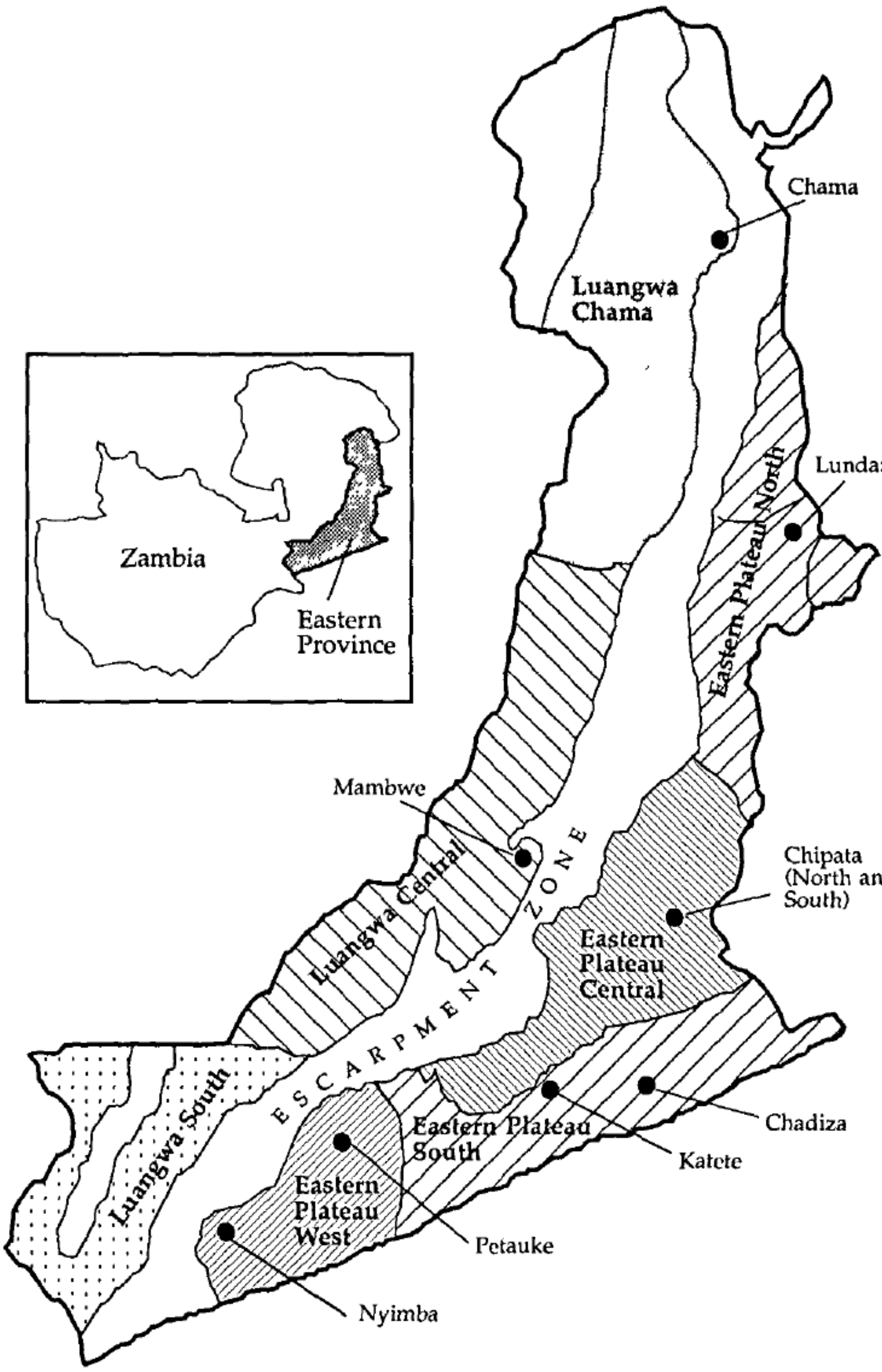
HECKMAN, J., ICHIMURA, H. & TODD, P. 1997. Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme. *Review of Economic Studies,* 64**,** 605-54.

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

JACOBY, H. G. & MINTEN, B. 2009. On Measuring the Benefits of Lower Transport Costs. *Journal of Development Economics,* 89**,** 28-38.
JALAN, J. & RAVALLION, M. 1997. Spatial Poverty Traps. *In:* GROUP, D. R. (ed.). World Bank.
JALAN, J. & RAVALLION, M. 2003. Estimating the Benefit Incidence of an Antipoverty Program by Propensity-Score Matching. *Journal of Business & Economic Statistics,* 21**,** 19-30.
KARLSSON, J. & BERKELEY, H. (eds.) 2005. *Handbook on Rural Household, Livelihood and Well-Being,* Wye: UNECE.
KASALI, E. 2002. Report On The Consultancy To Measure Distances From The 1999-2000 Post Harvest Survey Sample Areas. To The Nearest Tarred Or Main Roads And District Towns. .
KHANDKER, S. R., BAKHT, Z. & KOOLWAL, G. B. 2006. The Poverty Impact of Rural Roads: Evidence from Bangladesh. *World Bank Policy Research Working Paper,* 3875.
Kingombe, C.K.M., and D.W. te Velde. 2012. Structural Transformation and Employment Creation: The role of growth facilitation policies in sub-Saharan Africa. Background for the World Development Report 2013.
KYDD, J. & DORWARD, A. 2004. Implications of Market and Coordination Failures for Rural Development in Least Developed Countries. *Journal of international development,* 16**,** 951-972.
KYDD, J., DORWARD, A., MORRISON, J. & CADISCH, G. 2004. Agricultural Development and Pro-poor Economic Growth in Sub-Saharan Africa: Potential and Policy. *Oxford Development Studies, ,* 32.
LUNGU, A. 2006. Chipata headed for economic boom.
MCCORD, A. & WILKINSON, K. 2009. Assessing the Incidence of Public Works Programmes: Using Propensity Score Matching Techniques to Assess the Poverty Targeting of Employment in Two Public Works Programmes in South Africa. *SALDRU Working Papers.*
MCCULLOCH, N., BAULCH, B. & CHEREL-ROBSON, M. 2001. Poverty, Inequality and Growth in Zambia during the 1990s. *Discussion Paper No. 2001/123***,** 47.
MEGILL, D. J. 2000. Review of Sample Design for Zambia Post Harvest Survey (1997/98) and the Recommendations for Improving the Sampling Strategy and Estimation Procedures. *FSRP Working Paper No. 2.*
MEGILL, D. J. 2004. Recommendations on Sample Design for Post-Harvest Surveys in Zambia Based on the 2000 Census.**,** 38.
MINTEN, B. & KYLE, S. 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.MU, R. & VAN DE WALLE, D. 2007. Rural roads and poor area development in Vietnam. *Policy Research Working Paper Series,* 4340.
PLETCHER, J. 2000. The Politics of Liberalizing Zambia's Maize Markets. World Development Vol. 28, No. 1, pp. 129-142. .
POULTON, C., GIBBON, P., HANYANI-MLAMBO, B., KYDD, J., MARO, W., OSORIO, A., TCHIRLEY, D. & ZULU, B. 2004. Competition and Coordination in Liberalized Africa Cotton Market Systems. *World Development,* 32**,** 519-536.
RAVALLION, M. & JALAN, J. 1996. Growth divergence due to spatial externalities. *Economics Letters,* 53**,** 227-232.
ROSENBAUM, P. R. & RUBIN, D. B. 1983. The central role of the propensity score in observational studies for causal effects *Biometrika,* 70**,** 41-55.
RWAMPORORO, R. K., TOURNÉE, J. M. & CHITEMBO, A. 2002. Joint Final Evaluation. District Development Planning and Implementation in Eastern Province (ZAM/93/CO1) And Rehabilitation and Maintenance of Feeder Roads (ZAM/93/CO2 & ZAM/93/009). *In:* UNCDF (ed.).
SCHULZ, R. & BENTALL, P. 1998. Rehabilitation and Maintenance of Feeder Roads (Eastern Province). Project Evaluation Summaries – 1998. Prepared by the Policy, Planning and Evaluation Unit (PPEU). *Project Number: ZAM/93/C02*.
SIANESI, B. 2001. Implementing Propensity Score Matching Estimators with STATA. University College London and Institute for Fiscal Studies. .
THURLOW, J. & WOBST, P. 2005. Operationalizing Pro-Poor Growth. The Department for International Development.
TSCHIRLEY, D., POULTON, C. & BOUGHTON, D. 2006. The many paths of cotton sector reform in Eastern and Southern Africa: lessons from a decade of experience. *Policy Synthesis,* 80.
Ulimwengu, John; Funes, José; Headey, Derek; You, Liang. 2010. Paver le chemin du développement? L'impact des infrastructures de transports sur la production agricole et la réduction de la pauvreté en République démocratique du Congo. Washington, D.C. International Food Policy Research Institute (IFPRI).
VAN DE WALLE, D. 2009. Impact evaluation of rural road projects. *Journal of Development Effectiveness,* 1**,** 15-36.
WOOLDRIDGE, J. 2002. *Econometric Analysis of Cross-Section and Panel Data,,* Cambridge, Mass., MIT Press.

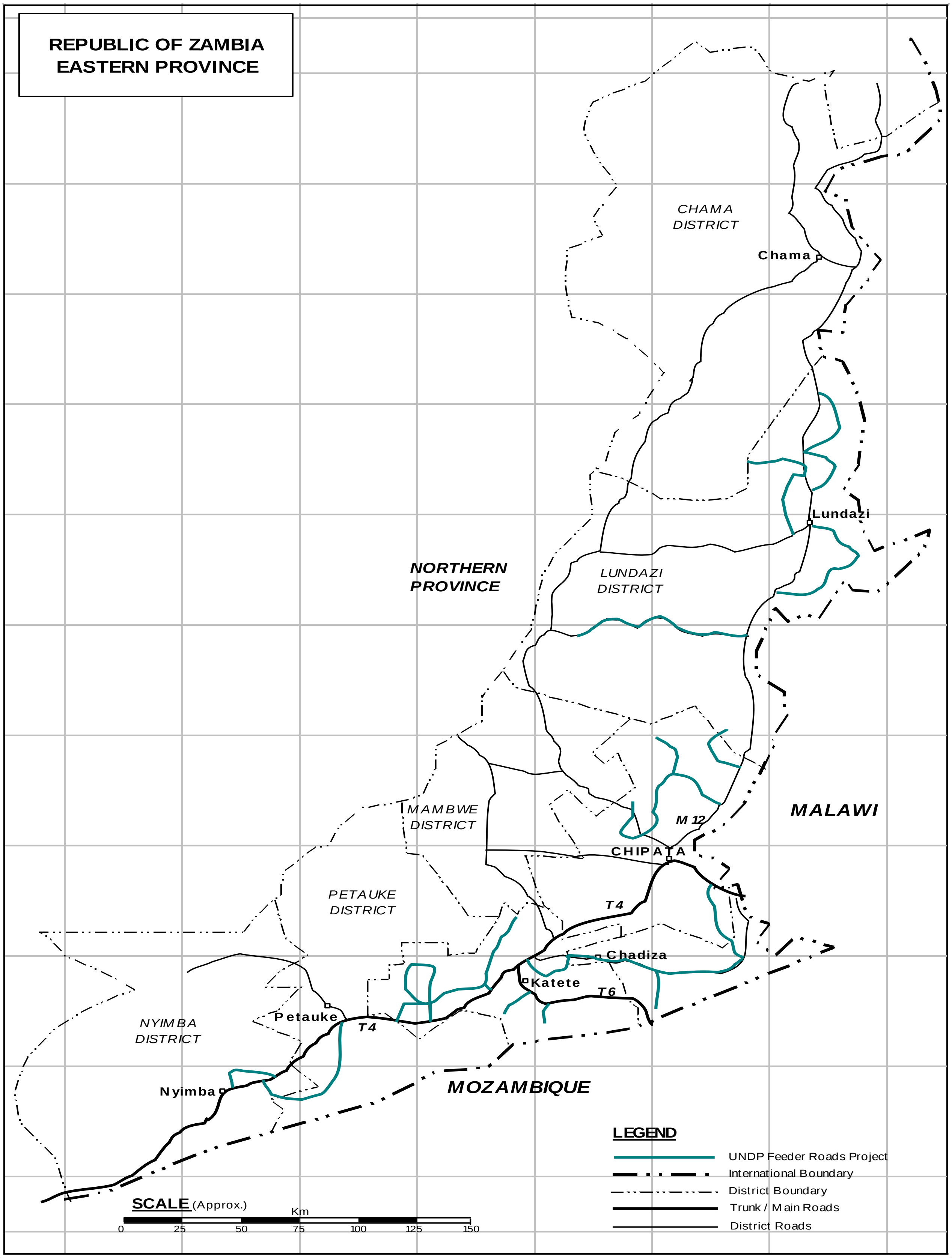
Appendix

Map A1: Agro-ecological zones and agricultural districts, Eastern Province, Zambia



Source: ARPT (Adaptive Research Planning Team), Eastern Province Agricultural Development Project, "Annual Report, 1985-86" (Chipata, Zambia, 1986, mimeographed).

Map A2: Illustration of the Eastern Province Feeder Roads



Source: UNDP EPFRP Document.