

# Farm Productivity and Market Structure. Evidence From Cotton Reforms in Zambia\*

Irene Brambilla<sup>†</sup>  
Guido G. Porto<sup>‡</sup>

March 2005

## Abstract

This paper investigates the impacts of cotton marketing reforms on farm productivity, a key element for poverty alleviation, in rural Zambia. The reforms comprised the elimination of the Zambian cotton marketing board that was in place since 1977. Following liberalization, the sector adopted an outgrower scheme, whereby firms provided extension services to farmers and sold inputs on loans that were repaid at the time of harvest. There are two distinctive phases of the reforms: a failure of the outgrower scheme, and a subsequent period of success of the scheme. Our findings indicate that the reforms led to interesting dynamics in cotton farming. During the phase of failure, farmers were pushed back into subsistence and productivity in cotton declined. With the improvement of the outgrower scheme of later years, farmers devoted larger shares of land to cash crops, and farm productivity significantly increased.

JEL CODES: I32 Q12 Q17 Q18

Key Words: cotton marketing reforms, farm productivity

---

\*We wish to thank A. Khandelwal for excellent research assistance. T. Jayne, J. Nijhoff, and N. Bsemukila allowed us access to the Post Harvest Survey data collected by the Central Statistical Office in Lusaka, Zambia. We thank J. Altonji, R. Betancourt, M. Duggan, P. Goldberg, C. Udry, and seminar participants at Maryland and Yale for useful comments and discussion. Financial Assistance from the World Bank is greatly appreciated. All errors are our responsibility.

<sup>†</sup>Department of Economics, Yale University, 37 Hillhouse, New Haven, CT. 06541. email: irene.brambilla@yale.edu

<sup>‡</sup>Development Research Group, MailStop MC3-303, The World Bank, 1818 H Street, Washington DC 20433. email: gporto@worldbank.org

# 1 Introduction

In Africa, commodity markets were traditionally controlled by public marketing boards that connected domestic farmers with product and input markets. Typically, these marketing boards enjoyed monopsony power in purchases of agricultural products, and monopoly power in sales of agricultural inputs to farmers. Agricultural liberalization during the 1990s led to the elimination of the marketing boards found in many African countries. The Zambian cotton sector is a good example of this type of reforms. Until 1994, Lintco (Lint Company of Zambia), a parastatal organization, controlled the sector by selling inputs, buying cotton, giving credit, and facilitating access to technology, equipment and know-how. In 1994, the sector was liberalized. Lintco was privatized and entry into the market was encouraged. Sluggish initial entry gave rise to a phase of regional private monopolies. During this phase, the firms developed an outgrower scheme. These are vertical arrangements between firms and farmers, whereby cotton ginners (i.e., the firms) provided inputs on loans that were repaid at harvest time. As additional entry and competition ensued, the outgrower scheme failed because farmers increasingly fell in default. On the one hand, farmers would take loans from one firm (for instance, an incumbent ginner) and would sell to another (for instance, an entrant). On the other hand, credit prices increased and international cotton prices declined, making cotton production less profitable. In 1999, further entry led to more competition. Simultaneously, the outgrower scheme was highly perfected and contracts between farms and firms were mostly honored. At present, the market is relatively unregulated and several firms seem to compete for locally produced cotton.

This paper investigates the dynamic impacts of cotton marketing reforms on farm productivity and crop choices in Zambia. There are several channels through which the reforms impacted on Zambian farmers. First, profitability of cotton production was affected, mainly through changes in input and product prices. Second, the uncertainty associated with cash cropping was affected through changes in the transparency of cotton marketing caused by the provision of extension services and technical assistance. Third, the transfer of technology (new seeds) and cotton know-how may have driven farmers to more efficient method of cultivation, increasing productivity and profitability. Further, changes in the

credit availability affected the cost of financing fixed capital production costs. These changes in prices, in access to inputs, and in efficiency of advice on crop husbandry, led to changes in land allocation to cotton and in cotton yields. Our objective in this paper is to quantify these impacts.

The empirical analysis exploits unusual farm surveys, the Post Harvest Surveys (PHS) of the Zambian Central Statistical Office. These are repeated cross-sections of Zambian farmers covering the 1997-2002 period. Information on land allocation, yields, input use, and household characteristics across farmers in rural Zambia is collected. We use these data to set up an empirical model of cotton crop choice and cotton productivity. Our identification strategy relies on a modified difference-in-differences approach. First, we take differences of outcomes (i.e., cotton productivity) across the different phases of the reforms. Second, we use maize productivity to difference out unobserved household and aggregate agricultural year effects. Finally, since more productive cotton farmers are also more likely to allocate a larger fraction of their land to cotton production, we use cotton shares, purged of observed covariates, as a proxy for unobserved cotton-idiosyncratic productivity.

A simpler difference-in-differences model, without the correction for selection and thus without accounting for entry and exit into the agricultural cotton sector, would lead to biases in the estimates of aggregate productivity. Exit of low productivity farmers in the failure phase may bias productivity up, whereas entry of low productivity farmers in the success phase may lead to downward biases in measured productivity. Our dynamic approach and our modified difference-in-differences model takes care of all these effects.

Our analysis provides valuable lessons on the interaction between export crops and the adoption of domestic policies in Sub-Saharan Africa. Further, by affecting market agricultural participation and cotton yields, our results have important implications for household income and consumption. These are critical issues in rural Zambia, where poverty rates exceed 80 percent of the population.<sup>1</sup> Since the success of the reforms and the outgrower scheme varied from the initial phase to the final phase, we find rich dynamics in cotton markets. During the initial phase, the failure of the outgrower scheme led to a decline in

---

<sup>1</sup>Poverty is widespread and deep in Zambia. For a comprehensive description of poverty trends, see Balat and Porto (2005).

the participation of households in cotton production and a decline in farm productivity. In contrast, the later phase of success induced farmers to increase the fraction of land devoted to cotton, and caused yields per hectare to significantly increase.

The paper is organized as follows. In Section 2, we review the main reforms in cotton markets. In section 3, we discuss the theory behind crop choices and farm productivity in Zambia. Here, we derive guidelines for the estimation of the impacts of the cotton marketing reforms. In section 4, we estimate a joint model of crop choices and farm productivity using the Post Harvest Survey farm data. We discuss the results and assess the effects of the reforms. In Section 5, we conclude.

## 2 The Zambian Cotton Reforms

Zambia is a landlocked country located in Southern Central Africa. With a population of 10.7 million and a per capita GDP of only 302 US dollars, Zambia is one of the poorest countries in the world. In 1998, for instance, the national poverty rate was 69.6 percent, with rural poverty at 82.1 percent and urban poverty at 53.4 percent. Nationwide, only around 4 percent of the income of rural households comes from the sales of non-food crops. Given the characteristics of the soil, cotton can only be grown in three Zambian provinces (the Eastern, Central, and Southern provinces). Where it is grown, cotton is a major source of income. In fact, using data from the Living Conditions Monitoring Survey of 1998, we find that the share of cotton in income was 8.4 percent in the Central province, 9.5 percent in the Eastern province, and 2.8 percent in the Southern province. This makes cotton an important sector in rural Zambia.

The process of reform began in 1991, when the Movement for Multi-Party Democracy (MMD) was elected. Faced with a profound recession, the new government implemented economy-wide reforms such as macroeconomic stabilization, exchange rate liberalization, trade and industrial reforms, and maize subsidies deregulation. More importantly for our purposes, privatization of agricultural marketing in cotton was also pursued.<sup>2</sup>

---

<sup>2</sup>For more details on cotton reforms in Zambia, see Food Security Research Project (2000), and Cotton News (2002).

Traditionally, the Zambian cotton sector was heavily regulated. From 1977 to 1994, cotton marketing was controlled by the Lint Company of Zambia (Lintco), a parastatal organization. Lintco set the sale prices of certified cotton seeds, pesticides, and sprayers, as well as the purchase price of cotton lint. Lintco had monopsony power in cotton purchases and monopoly power in inputs sales and credit loans to farmers.

In 1994, comprehensive cotton reforms began to take place. Most interventions were eliminated when Lintco was sold to Lonrho Cotton. Initially, a domestic monopsony developed early after liberalization. Soon, however, expanded market opportunities induced entry of private ginners such as Swarp Textiles and Clark Cotton. Because these three major firms segmented the market geographically, the initial phase of liberalization did not succeed in introducing competition, giving rise, instead, to geographical monopsonies rather than national oligopsonies.

At that moment, Lonrho and Clark Cotton developed an outgrower scheme with the Zambian farmers. In these outgrower programs, firms provided seeds and inputs on loans, together with extension services to improve productivity. The value of the loan was deducted from the sales of cotton seeds to the ginners at picking time. Prices paid for the harvest supposedly reflected international prices. Initially, repayment rates were high (around roughly 86 percent) and cotton production significantly increased. We called this the *outgrower introductory phase*.

By 1998, the expansion of cotton farming attracted new entrants, such as Amaka Holdings and Continental Textiles. Instead of the localized monopsonies, entrants and incumbents started competing in many districts. As competition among ginners ensued, an excess demand for cotton seeds developed. Several concurrent factors explain why, however, farmers could not realize the full benefits of the competition phase. First, some firms that were not using outgrower schemes started offering higher prices for cotton to farmers who had already signed contracts with other firms as outgrower. This caused repayment problems and increased the rate of loan defaults. The relationship between ginners and farmers started to deteriorate. Second, world prices 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, farmers started to mistrust the ginners. As the relationship between farmers and firms deteriorated, default rates increased even further. In consequence, firms raised loan prices and farmers end up receiving a lower net price for their cotton production. We called this the *outgrower scheme failure phase*.

Partly as a result of this failure of the outgrower scheme, Lonrho exited the market in 2000. A new major player, Dunavant Zambia Limited, entered instead. Dunavant and competitors, Clark Cotton Limited, Amaka Holdings Limited, Continental Ginneries Limited, Zambia-China Mulungushi Textiles, and Mukuba Textiles, worked to improve the scheme. Farmers and firms understood the importance of honoring contracts and the benefits of maintaining a good reputation. The outgrower programs were perfected and there are now two systems utilized by different firms: the Farmer Group System and the Farmer Distributor System. In the latter, firms designate one individual or farmer as the distributor and provide inputs. The distributor prepares individual contracts with the farmers. He is also in charge of assessing reasons for loan defaults, being able, in principle, of condoning default in special cases. He is in charge of renegotiating contracts in incoming seasons. In the Farmer Group System, small scale producers deal with the ginneries directly, purchasing inputs on loan and repaying at the time of harvest. Both systems seem to work well. We call this the *outgrower scheme success phase*.

### **3 Overview of Strategy**

We define productivity as yields per hectare in physical units. Hence, our productivity definition differs from the standard definition used in industrial analysis (usually value added at constant prices). A physical definition of productivity is economically more meaningful because it only reflects technology, while value added depends on market conditions via prices.

We begin by discussing some of the determinants of agricultural productivity that have been identified in the literature. In a model with decreasing returns to fixed factors of production, a key determinants of cotton yields per hectare is the size of the plot allocated

to cotton. A family farm, for instance, may obtain higher yields per unit of land if the scarce labor resources are utilized in smaller plots. Major determinants of the size of land allocated to cotton can be found in the literature on crop choice. There are different factors determining this selection process, and sometimes separate strands of literature explore the different dimensions of the problem. The theoretical problem is straightforward: endowed with a fixed amount of land, the farmer must choose the fraction of resources to be allocated to food crops (mainly maize) and cash crop (mainly cotton).

A key factor is the trade-off between profitability and risk, as in a standard portfolio allocation choice (Rosenzweig and Binswanger, 1993). Thus, relative product prices (cotton, maize) and input prices (seeds, fertilizers, pesticides) affect the choice of crops. It may be argued that cash crops show higher returns but are riskier than food crops, so that different attitudes towards risks (degree of risk aversion) can help explain the selection (Binswanger and Sillers, 1983; Dercon, 1996; Shahabuddin et al., 1982). Since direct measures of risk aversion are not available, we need to proxy them with relevant household characteristics. In particular, the attitude towards risk can be affected by variables such as household wealth, household size, and household composition.

Often times, growing cash crops requires a start-up lumpy investment that may constrain the allocation of resources (Eswaran and Kotwal, 1986; Dercon, 1996). Sometimes this takes the form of capital investment in machines, tractors, animals. In addition, there might be large initial costs of input purchases such as new seeds or expensive pesticides or sprayers. In the presence of well-developed credit markets, this fixed costs could be easily covered. When credit constraints are binding, however, the ability to borrow and the availability of collateral can be determinants of the choice of crops.

An additional important argument claims that the allocation of resources is affected by missing markets (de Janvry, Fafchamps, and Sadoulet, 1991). In fact, whereas cash crops must be sold at the market price, maize can be consumed in the family to provide food security. In many less developed countries, concerns for food security are of the utmost importance. Families will want to secure the food needs of the family first, and then move to cash cropping. If food markets were well-developed, then food risk would not be an issue

because households could grow cash crops, sell them at the market, and use the proceeds to purchase food. If food markets are missing or are thin and isolated (so that ex-post food prices are high and volatile), then a strategy of self-sufficiency in food production may be optimal (Fafchamps, 1992; Jayne, 1994). This suggests two sets of empirical determinants of cotton choice. First, regional characteristics, such as the availability of food markets, the number of food producers in the area, regional infrastructure, and other variables that may affect how thin local food markets are, may be relevant. Second, in the presence of food security issues, the determinants of food needs may be important. Examples include household size, household composition (so that, for instance, households with a larger fraction of children would have larger food needs), land tenure, and non-farm income.

The switch from subsistence to cotton can be interpreted as technology adoption in agriculture. There is a large literature that explores the determinants of adoption (Besley and Case, 1994; Foster and Rosenzweig, 1995; Conley and Udry, 2004). This literature identifies human capital (measured by education, gender, and age) as a major determinant of technology adoption. In addition, these authors argue that social capital, learning by doing, and learning externalities are important determinants, too. In this setting, technology adoption depends on the fraction of the neighbor farmers that have already adopted.

Many of the factors mentioned above also affect productivity directly (and not only through the cotton land allocation). For instance, the human capital of the farmer, as measured by his age, gender, and education, surely affects yields. Technology, in the form of crop know-how, high yield seeds, and efficient agricultural tools (like tractors or sprayers) may also lead to a better use of resources and to higher productivity. Similarly, if the production of the cash crops involves the initial purchase of inputs, capital goods, and machines, lack of credit and collateral (determined by land size, assets, wealth, savings, off-farm income, and remittances) may hinder access to more efficient resources like better seeds, sprayers or animals (i.e., oxens). By the same token, access to local infrastructure and public goods and capital can increase yields per hectare. Further, there is an important role played by agricultural extension services and technical advice on crop husbandry, land use, and general agricultural assistance, that allow farmers to achieve higher productivity.

A similar role can be attached to social capital and learning externalities.

We are not only interested in the direct determinants of agricultural productivity, but also in the impacts of the marketing reforms. There are several channels that can be advanced. The provision of credit and of inputs on loan, which may allow farmers to better combine factors of production, may depend on the phases of the reform. During the collapse of the outgrower scheme, credit became more expensive to farmers thus hindering productivity. When the scheme improved, credit became cheaper, probably causing productivity to increase. Similarly, the privatization of the ginning industry may cause firms to adopt better cotton seeds (with higher yields) and more efficient pesticides and fertilizers. This would work as technological advances that firms pass-through to farmers, leading to increases in farm productivity. Finally, the outgrower scheme involved an improvement in the provision of extension services, particularly in terms of information about marketing. This could help eliminate some uncertainty about the crop. In addition, more efficient extension services, providing advice on crop husbandry and know-how, can help farmers increase yields.

Our empirical analysis relies on a modified difference-in-differences approach. The simple difference between cotton productivity in any two of the reform phases (i.e., the introductory, the failure, and the success phases) is not a consistent estimate of the impacts. It does not take into account the general trend or time-varying aggregate effects in agriculture. In addition, there are unobserved idiosyncratic farm effects that can affect inputs choices. To account for these household and agricultural effects, we use a model of maize productivity. This second differencing works because maize is a major food crop that is produced by virtually all cotton producers.<sup>3</sup> Moreover, we also show that maize productivity is a good indicator of aggregate productivity in agriculture. Hence, we can use it to control for time-varying effects.

In most applications, such a difference-in-differences approach would be enough. In the present case, there may be additional unobserved factors in the productivity equation that may bias our estimates. These unobservables are cotton-specific farm effects, such as ability

---

<sup>3</sup>We will provide strong evidence in favor of this claim below. A key characteristic of cotton farming in Zambia is its scale: cotton is grown by smallholders, family farms endowed with small farms, usually smaller than four hectares and with an average size of around 2 hectares.

(in weeding or planting cotton, for instance) and land quality. In what follows, we further elaborate on this idea.

If there are entry costs into cotton, then cotton production will only be profitable if productivity is high enough. This means that there is a cut-off  $\bar{y}^c$  (which depends on prices, market conditions, infrastructure) such that farmers with productivity above this cut-off will enter the market and farmer below the cut-off will not enter (or exit, if they were in the market already). When the reforms increase the profitability of cotton, for instance, lower productivity farmers may enter the market. Failure to control for this may lead to inconsistently lower estimates of aggregate productivity (thus leading to a downward bias in the estimates of any productivity increases). In contrast, in periods of induced exit, farmers with lower unobserved productivity will be more likely to abandon cotton production. In consequence, measures of average productivity that do not control for these dynamic effects may be artificially high (thus leading to downward biases in the estimates of productivity declines).<sup>4</sup>

Figure 1 clarifies these dynamics. The graph shows cotton productivity  $y^c$  as a function of unobserved cotton-specific effects (which we denote by  $\phi$ ). Productivity is increasing in  $\phi$  since better land quality or higher cotton skills will lead to higher output (for a given usage of other inputs). The horizontal line at  $\bar{y}^c$  denotes the cut-off; for simplicity, we assume here that it does not vary with the reforms. It follows that we can determine a cut-off for the unobservables, denoted  $\bar{\phi}$ . The line denoted  $y_0^c$  represents the cotton productivity function before the reform. Average productivity is, say,  $E(y_0^c)$ , the average of  $y^c$  for levels of  $\phi > \bar{\phi}$ .

Consider the effects of the failure of the outgrower scheme. If cotton productivity is negatively affected, the productivity curve shifts down to  $y_1^c$ . Assuming a fixed cut-off  $\bar{y}^c$ , the cut-off for the unobservables increases to  $\bar{\phi}'$ .<sup>5</sup> This induces the “exit” of those farmers with relatively low levels of  $\phi$ , between  $\bar{\phi}'$  and  $\bar{\phi}$ . Average productivity drops to  $E(y_1^c)$ . But the decline in individual productivity is larger. The right quantity is the average productivity, computed along the curve  $y_1^c$ , and integrating over values of  $\phi$  above the cutoff before the

---

<sup>4</sup>See Olley and Pakes (1996), Levinsohn and Petrin (2003), and Pavcnik (2003) for models of industrial productivity, entry and exit.

<sup>5</sup>It is not necessary to assume that  $\bar{y}^c$  remains fixed after the reform. Our intuition remains unchanged.

reform  $\bar{\phi}$ . This is given by  $E(y_r^c)$ .

The difference in differences model described so far estimates changes in average productivity given by  $E(y_0^c) - E(y_1^c)$ . To estimate the true effect at the farm level,  $E(y_0^c) - E(y_r^c)$ , we need to account for the role of unobserved cotton effects. To do this, we make the argument that more productive farmers are also more likely to cultivate more cotton and to allocate larger fractions of land to this crop. This means that we can use the crop cotton shares as proxies for cotton-specific productivity. To reduce biases in the estimation, these shares are “purged” of the effects of standard determinants of crop choice.

## 4 Estimation

In this section, we estimate the impacts of the cotton marketing reforms on farm productivity. We describe the data, we develop the empirical model, and we provide the basic productivity results. Finally, we extend our results to control for selection issues, we report robustness results, and we address some additional econometric concerns.

### 4.1 The Post Harvest Survey

We use farm surveys called the Post Harvest Survey (PHS). These data are collected by the Zambian Central Statistical Office (CSO). The surveys are not panel data but rather a set of repeated cross-sections. We have annual data available for the period 1997-2002. The survey is representative at the national level, but in this paper we only use the data pertaining to cotton producing regions: the Central, Eastern, Southern and Lusaka provinces. CSO gathers information on land tenure, land usage (allocation), output in physical units, and households characteristics such as demographic composition, age of head, and housing infrastructure. There are also limited data on farm assets and inputs.

Table 1 provides an overview of the relevant sample sizes, by year and by province. Around 600-700 households are interviewed in the Central province, around 1,200, in the Eastern province, around 800 in the Southern province, and around 200 in Lusaka. Table

2, which reports the fraction of farmers involved in cotton production, confirms that these are the major cotton producing areas. Significant percentages of cotton farmers are only observed in the Central, the Eastern, and the Southern provinces. Cotton participation is largest in the Eastern province (39 percent in 2002, for instance), then in the Central province (20 percent in 2002), and finally in the Southern province (12.6 percent). There are some, but not many, cotton producers in the Lusaka region, too. In the remaining provinces, the percentage of households involved in cotton is virtually zero.

Table 2 reveals interesting dynamic patterns that we explore below. During the introduction phase, which spans the years 1997 and 1998, cotton participation is relatively stable in all provinces (although a declining pattern may be discernible). The failure phase, which spans the 1999-2000 period, shows lower participation rates. This is particularly evident in 2000: in the Central province, for instance, cotton participation drops from 22.6 percent in 1998 to 10.3 percent in 2000. Similarly, participation declines from 32.7 to 20.4 percent in the Eastern province, from 10.7 percent to 4.3 percent in the Southern provinces, and from 3.3 percent to 0.4 percent in Lusaka. The success phase correlated with entry into cotton: the percentage of cotton growers increases significantly in all provinces.

An important aspect of our strategy is the use of maize productivity to control for some unobserved determinants of productivity in cotton. For this strategy to work well, maize has to be produced by most farmers. Table 3, which reports the percentage of households that grow maize, provides evidence supporting this. We find that in the cotton provinces, maize is grown by virtually all households. Participation in maize production is always above 90 percent in the relevant regions. In the Eastern and Lusaka provinces, the percentage of maize producers is nearly 100 percent. Table 4 reports additional evidence that further supports our differencing strategy. We report the percentage of farmers that grow maize, conditional on being cotton growers. These shares are nearly 100 percent in the three main cotton-growing provinces.

In Table 5, we report data on the fraction of land allocated to cotton. In 2002, for instance, an average farmer in the Eastern province allocated around 17.2 percent of his land to cotton; in the Central province, the fraction is 10.5 percent. Instead, cotton adoption is

less widespread in the Southern province, where only an average of 5.6 percent on land is allocated to cotton. The dynamics of cotton adoption are also revealed in Table 5. The fraction of land allocated to cotton sharply declines in 1999 and 2000 and then increases in 2001 and 2002.

Finally, Table 6 describes the evolution of cotton yields per hectare. The figures are in logarithms, so that changes from one year to the other can be interpreted as growth rates. At the national level, cotton yields increased from 1997 to 1998, and then decline during the failure phase. In fact, productivity dropped by 24 percent from 1998 to 2000. However, productivity in 2000 is comparable to productivity in 1997. The success phase brought productivity up in and 2001 and 2002. Notice that there are interesting differences in regional dynamics. In the Eastern and Southern provinces, for instance, productivity changes track those observed at the national level. However, in the Central province, productivity increased steadily from 1997 to 2000 and then declined in the success phase of 2001 and 2002. We examine these differences below.

## 4.2 The Empirical Model

Productivity is defined in physical units. Let  $y_{ht}^c$  denote the volume of cotton production per hectare produced by household  $h$  in period  $t$ . The log of output (in kilograms) per hectare is given by

$$(1) \quad \ln y_{ht}^c = \mathbf{x}_{ht}^c \beta_c + \alpha_1 F_t^1 + \alpha_2 F_t^2 + I_t + \eta_{ht} + \epsilon_{ht}^c.$$

Here,  $\mathbf{x}_{ht}^c$  is a vector of household determinants of cotton yields such as the age of the household head, his education, the size of the household, household demographics, input use, assets, the size of the land allocated to cotton, farm size, and district dummies.

We model the productivity effects of the reforms with two variables,  $F_t^1$  and  $F_t^2$ .  $F_t^1$  is a dummy variable that captures the second period of the reform, the *outgrower scheme failure phase* of 1999-2000, and  $F_t^2$  is another dummy that captures the third period of the reform, the *outgrower scheme success phase* of 2001-2002. The impacts of these phases of the reform

are measured relatively to the excluded category, which is the *outgrower scheme introductory phase* of 1997-1998.

The model includes a number of fixed effects, such as an agricultural effect  $I_t$ , district (regional) effects –included in the  $x$ –, and idiosyncratic farm fixed effects,  $\eta_{ht}$ . The  $I_t$  effect captures agricultural effects and other shocks that are common to all farmers in a given period  $t$ . The regional effects, which include market access, local infrastructure, local knowledge, access to credit, etc., can be captured with district dummies. The idiosyncratic productivity captures all sorts of household effects that affect productivity, including land quality, know-how, and other specific fixed effects.

The fixed effects are unobserved by the econometrician. A panel data set would allow us to account for both year and household effects. The Post Harvest Survey, however, is a repeated cross section of farmers. We thus need some control equation to deal with the fixed effects. As argued above, we propose to model agricultural productivity in maize to difference out the household and year effects. Yields per hectare in maize,  $y_{ht}^m$ , is given by

$$(2) \quad \ln y_{ht}^m = \mathbf{x}_{ht}^m \beta_m + I_t + \eta_{ht} + \epsilon_{ht}^m,$$

Maize productivity depends on covariates  $\mathbf{x}_{ht}^m$ . In principle, we allow the regional effects in cotton and maize to differ due to varying marketing, information, local infrastructure, or agricultural extension effects. For example, to the extent that  $\delta_r$  captures local market access effects, we allow marketing conditions to affect cotton (a cash crop activity) and maize (a mostly subsistence crop) differently. The agricultural effects,  $I_t$ , and the farm effects  $\eta_{ht}$  are also determinants of maize productivity.

By taking first differences, we get

$$(3) \quad \ln y_{ht} = \ln(y_{ht}^c/y_{ht}^m) = \mathbf{x}_{ht}'\beta + \alpha_1 F_t^1 + \alpha_2 F_t^2 + \epsilon_{ht}.$$

This is an estimable difference-in-differences equation. The observed household covariates  $\mathbf{x}_{ht}$  included in the estimation are based on the determinants of productivity discussed in the previous section, such as household demographics, human capital, determinants of household

collateral, determinants of food needs, etc. Equation (3) includes district effects that are equivalent to the difference of the district effects in cotton and maize. These district effects capture local infrastructure and a myriad of village effects, such as credit schemes, food markets, and prices.

The coefficients  $\alpha_1$  and  $\alpha_2$  measure the impacts of the different phases of the reforms on cotton productivity. There are two important identification assumptions. First, we assume that the agricultural effects,  $I_t$  affect cotton and maize productivity proportionately. This is a consequence of the logarithmic specification that we adopt. In other words, the agricultural effects are assumed to have the same effect on growth of cotton and maize output per hectare.<sup>6</sup> This is an instance of the parallel trend assumption of difference-in-differences models. It means that we can use the trend in maize productivity to predict the counterfactual productivity in cotton in the absence of the reforms. Although this assumption cannot be tested directly, we can provide indirect evidence supporting it. In particular, 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.

Figure 2 provides evidence that supports this. Each panel compares the trend in maize productivity (solid line) with the trend in alternative crops (broken line). These are sorghum, millet, sunflower, groundnuts, and mixed beans. We observe that, perhaps with the sole exception of groundnuts in 2001, the trends in all these crops are very similar. In the regression analysis, we use maize as control because, as opposed to the other crops, virtually all household produce it.

A similar identifying assumption is adopted as regards the farm effects  $\eta_{ht}$ . That is, we assume in (3) that the unobservable farm effects affect cotton and maize productivity proportionately. In section 4.4 below, we relax this assumption.

The second critical assumption of our difference-in-differences model is that the cotton reforms did not affect maize productivity. Theoretically, agricultural reforms of the type studied here are likely affect productivity in all crops through resource allocation (i.e., labor,

---

<sup>6</sup>Of course, the level effect is going to be different. This is reasonable, since physical units of cotton and maize are not comparable.

fertilizers, pesticides) and capital accumulation. In addition, there may be indirect channels, through, for example, access to credit. If the reforms affect farms by providing cotton inputs on loan, household resources to purchase seeds or fertilizers in maize may be released. To the extent that the regression includes these variables in the observed covariates  $\mathbf{x}$ , these effects will be accounted for. In the regressions, we include measures of labor, agricultural tools, and fertilizers, land allocation.

Notice, however, that for some determinants, such as labor allocation, we only have household level data (as opposed to crop level data). This raises the possibility that the reforms affected maize productivity and that  $\alpha_1$  and  $\alpha_2$  are measures of the impacts of the reforms on cotton productivity relative to maize productivity. We could rule out this possibility by providing additional evidence of the trends in maize productivity in those provinces that were not affected by the cotton reforms. These trends are plotted in Figure 3. The solid line corresponds to the trend in maize productivity in reform provinces. Instead, the broken line displays the trend in non-reform provinces. It can be seen that the parallel trend holds in this case, except in 2002. Overall, this indicates that the differencing will identify the impact of the reform on cotton productivity only.

### 4.3 Main Results

Our benchmark productivity results are reported in Table 7. Columns (1) and (2) report estimates of equation (1), that is, a productivity model that does not control for unobservables such as  $I_t$  and  $\eta_h t$ . In these regression, we use data from the three main cotton provinces, the Central, the Eastern, and the Southern provinces. The main findings indicate that small farms are more productive; there is also evidence 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 per hectare. The negative association between farm size and household agricultural productivity has long been established in the literature (Feder, 1985; Benjamin, 1994). In addition, households with male heads are more productive in cotton, as are larger households. Assets (such as ploughs or livestock) are positively associated with yields. The effects of inputs such as basal and top-dressing fertilizers are not as strong as

expected.<sup>7</sup>

The dynamics of cotton productivity are closely linked to the dynamics in market structure: compared to the *introductory phase*, productivity is lower in the *failure phase* and higher in the *success phase*. The estimated magnitudes are important: in the failure phase, productivity declines by around 15.1 percent (column 2) and increases by 11.9 percent in the success phase.

Columns (3) and (4) report results of a the difference-in-differences model. The estimated impacts of the marketing reforms are significantly higher. On the one hand, productivity during the failure phase declines by 45.1 percent (column 4) instead of by 15.1 percent (column 2). This is because there is a positive trend in yields (net of the effects of covariates) from the introductory to the failure phase. On the other hand, the increase in productivity during the success phase is of around 20.8 percent (column 4) as opposed to 11.9 percent (column 2). This suggests a declining trend in yields from the introductory to the success phase. Interestingly, this means that, when comparing the failure and success phases, productivity in fact increases by a whopping 65.9 percent.

#### 4.4 Entry and Exit in Cotton Farming

As argued above, the dynamics of entry and exit in cotton farming may lead to biases in our productivity estimates. When the reforms induce entry, these entrants are likely to show lower levels of cotton productivity (otherwise, they would have entered the sector before). Omitting controls for the fact that entrants are low productivity farmers lay lead to downward biased in the estimates of the impacts of the reforms. During periods of exit, in contrast, those farmers that are leaving the market are those with lower productivity. Omitting these controls would lead to upward biases in the impacts of the reforms.

One way to deal with these biases is by introducing additional unobserved cotton-specific effects in the regression model. These effects include factors like unobserved ability in cotton husbandry or land quality that affect both the adoption of cotton and the level of cotton

---

<sup>7</sup>One reason for these result is that both basal and top-dressing fertilizers are actually used in maize more than in cotton. See the discussion below for more details.

productivity. Formally, we model entry, exit, and unobserved crop-specific productivity effects by rewriting the difference-in-differences model as

$$(4) \quad \ln y_{ht} = \ln(y_{ht}^c/y_{ht}^m) = \mathbf{x}'_{ht}\beta + \alpha_1 F_t^1 + \alpha_2 F_t^2 + b_0\phi_{ht} + \epsilon_{ht}^a.$$

This model is similar to (3) but includes an additional productivity factor,  $\phi_{ht}$ . In one interpretation of this model, there are unobserved cotton and maize effects, and  $\phi_{ht}$  captures their relative importance. An alternative model argues that  $\phi_{ht}$  measures cotton-specific effects only, such as cotton husbandry and land quality in cotton. In fact, all maize effects have been taken care of by the agricultural effect,  $I_t$ , and the farm effect,  $\eta_{ht}$ , so that it is unlikely to have additional unobserved maize effects. Zambian farmers have grown maize for many years and maize production for food consumption is a tradition in Zambian agriculture. Hence, it is reasonable to expect small changes in these unobservables from year to year.<sup>8</sup>

Our solution to this problem is to construct proxies for the unobserved productivity parameter. Our method exploits the idea that since households with high  $\phi$  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  $\phi_{ht}$  in (4). In practice, consistent estimation requires that we purge these shares of the part explained by observed determinants of cotton choice.

Let  $a_{ht}^c$  be the fraction of land allocated to cotton. A general model of these shares is

$$(5) \quad a_{ht}^c = m(\mathbf{z}_{ht}, \phi_{ht}),$$

where  $\mathbf{z}$  is a vector of regressors which includes district effects  $\delta_{rt}^a$  that affect selection into cotton production. For instance, we use  $\delta_{rt}^a$  to capture access to market and local infrastructure that facilitates farmers participation in market cash agriculture. The function  $m$  allows regressors  $\mathbf{z}$  and unobservables  $\phi$  to affect the shares  $a$  non-linearly.

---

<sup>8</sup>Notice that omitting  $\phi$  may also induces correlation between some variables in the vector  $\mathbf{x}$  and the error term in the difference-in-differences model. For example, the choice of inputs, such as labor or pesticide use, will depend on  $\phi$  (so that higher levels of unobserved productivity may be positively correlated with input use). The model in (4) takes care of these biases.

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

$$(6) \quad a_{ht}^c = \mathbf{z}'_{ht} \gamma_t + \phi_{ht},$$

Estimation of (6) 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 (6) by using a CLAD (censored least absolute deviation) model. Notice that, provided the right specification for the model is used, consistency follows because the regressors  $\mathbf{z}$  are exogenous to  $\phi$ . This requires that fertility, family composition, or farm size do not depend on unobservables such as cotton-specific ability or land quality. Importantly, since we use data on all households to estimate (6), this equation does not suffer from a selection problem like the one we are attempting to control for in the productivity model.

The allocation of land to cotton depends on several key factors that we need to account for. Importantly, the selection into cotton depends on the reform. This means that we should include  $R^1$  and  $R^2$  in (6). 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 region, however, we can account for them with year or regional dummies. In practice, we estimate a different model like (6) in each of the six years from 1997 to 2002 (notice that  $\mathbf{z}_t$  is indexed by  $t$  in (6)). This means that we will not be able to separate the effects of the reforms from the effects of changes in international prices on land allocation. But at least we will be able to control effectively for  $\phi$  in the productivity model.<sup>9</sup>

Finally, note that the identification assumption is that the selection into cotton is affected by the same unobservables affecting cotton productivity. In principle, it would be possible to argue that there are additional unobservable factors that affect the selection into cotton. We extend our results to the case where these unobservables differ in section 4.6.

---

<sup>9</sup>We also consider the possibility of estimating different selection models in different years and in different provinces. This would control for idiosyncratic provincial effects in cotton adoption. We report results in the next section.

Plugging in the estimates of  $\phi$  in (4), the productivity model is

$$(7) \quad \ln y_{ht} = \mathbf{x}'_{ht}\beta + \alpha_1 F_{ht}^1 + \alpha_2 F_{ht}^2 + b_0 \widehat{\phi}_{ht} + \widetilde{\epsilon}_{ht}.$$

This modified difference-in-differences approach is consistent with entry and exit into cotton farming.

Table 8 reports the productivity results corrected for entry and exit.<sup>10</sup> Column (1) reproduces the estimates from column (4) of Table 7, which does not include controls for  $\phi$ . Columns (2) and (5) use a Tobit model to estimate the selection equation, columns (3) and (6) use a linear model, and columns (4) and columns (4) and (7) use a CLAD model. Model 1 and Model 2 in Table 8 differ in the list of covariates: both models share the same regressors, but Model 1 measures assets (harrows and ploughs) in monetary units and Model 2 measures them in physical units. Notice that since the regression includes an estimated regressor,  $\widehat{\phi}$ , the standard errors should be corrected. We estimate them with a bootstrap model with 100 repetitions.

We confirm that productivity declines during the failure phase (i.e.,  $\alpha_1$  is negative and significant), and increases during the success phase (i.e.,  $\alpha_2$  is positive and significant). The results are robust to the selection model used to build the proxy for  $\phi$ , i.e., the linear model, the Tobit model or the CLAD model. The decline of cotton yields per hectare during the failure phase ranges from 49.6 to 53.6 percent. The coefficients of  $F^2$  range from 0.202 to 0.209.

Failure to control for  $\phi$  can damage the estimated impacts of the reforms on average productivity, particularly during the failure phase. In column (1), we report a decline in average productivity of 45.1 percent from the introductory phase to the failure phase. When exit is accounted for, the decline in productivity is, instead, of around 51.9 percent. This means that although the average aggregate productivity in the economy declined by 45.1 percent, the average productivity of a typical cotton farm declined by 51.9 percent. In other words, average productivity is 6.8 percent higher than what it would hadn't the most unproductive farmers (in terms of  $\phi$ ) exited the market.

---

<sup>10</sup>The results of different selection models are discussed in Appendix 1.

It is interesting to notice that, during the success phase, the reforms increase yields by around 20.2 percent, comparable to the findings in column (1). That is, the estimated  $\alpha_2$  do not depend on whether the regression controls for  $\phi$  or not. This means that entry is not affecting the estimated changes in average productivity by much. One explanation of this finding is that entry is much more costly than exit. When unobservables  $\phi_{ht}$  are such that cotton becomes unprofitable, farmers may exit at no significant cost. Instead, when cotton becomes profitable, there might still be impediments to entry.

So far, we have assumed that  $\phi$  enters additively in the land cotton shares equation (6). However, there are reasons to believe that the residuals from (6) are a non-linear function of the unobservables  $\phi$ . We can write

$$(8) \quad a_{ht}^c = \mathbf{z}'_{ht}\gamma_t + \rho_{ht},$$

where  $\rho_{ht} = f_t(\phi_{ht})$ . The productivity model is

$$(9) \quad \ln y_{ht} = \mathbf{x}'_{ht}\beta + \alpha_1 F_{ht}^1 + \alpha_2 F_{ht}^2 + g_t(\rho_{ht}) + \tilde{\epsilon}_{ht},$$

where  $g_t(\rho_{ht}) = f_t^{-1}(\rho_{ht})$ . This model can be estimated using a partially linear model (Robinson, 1988). In the first stage, both  $\ln y$  and all of the covariates  $\mathbf{x}$  are regressed on  $\rho$  non-parametrically. This is done using a locally weighted linear regression (Pagan and Ullah, 1999). In the second stage, we estimate residuals for all these variables using the non-parametric estimates. Finally, a linear OLS regression between residuals is run. This procedure recovers the linear part of the model,  $\beta$ ,  $\alpha_1$ , and  $\alpha_2$ .<sup>11</sup>

The results of the partially linear model are reported in Table 9. In all our specification, we find that the non-parametric correction does not affect the estimates of the impacts of the reforms. Concretely, the failure phase leads to a decline in productivity of 52-53 percent, whereas the success phase leads to increases of productivity of around 20 percent.

---

<sup>11</sup>The non-linear function  $g_t(\cdot)$  can be estimated with a non-parametric regression of  $\ln y$ , purged of the observed covariates  $\mathbf{x}$ ,  $F_1$ , and  $F_2$ , on  $\rho$ .

## 4.5 Robustness

Our robustness analysis follows along three lines: sensitivity to the definition of the reforms, sensitivity to the inclusion of Lusaka growers in the sample, and differences in regional analysis.

Table 10 reports estimates for different definitions of the reforms. The dynamics generated by the elimination of the marketing board are generally complex, and it may be difficult to assign different years to the different phases of the reforms. Our estimates can thus be sensitive to the definition that is being used. To examine the robustness of our results, we re-estimate the model using two additional definitions of the reforms. First, we redefine the failure phase as including only the year 2000 (dummy denoted  $R^1$ ) and including 1999 in the introductory phase. As shown in section 4.1 (Tables 2, 5 and 6), the drop in the share of land allocated to cotton declines much more markedly in 2000 than in 1999. Similar observations characterize the trends in cotton yields. The success phase still includes 2001 and 2002 (with dummy defined by  $R^2$ ). In our second redefinition, we measure the impacts of the reforms by including year dummies, thus allowing the effects of the reforms to vary year by year. In this model, there are six phases in the dynamics that we estimate.

We estimate two different models in Table 10. Columns (1) to (3) use a Tobit procedure and Model 1 of Table 8 (measuring assets in monetary units) for the estimation of  $a_{ht}^c$ ; columns (4) to (6) also use a Tobit model, but adopt Model 2 (assets in physical units) of Table 8. Our qualitative conclusions remain unaffected. There is a decline in productivity in 2000 of around 30 percent in both specification. Also, there is an increase in productivity in the success phase of 16.2 percent. More detailed patterns can be discerned when we use year dummies to measure the different phases of the reforms. Compared to 1997, we find that productivity first increases in 1998 and declines in 1999 to 1997 levels. We still find a large decline in productivity in 2000, of around 41 percent. During the success phase, productivity follows an increasing trend: output per hectare is 20 percent higher in 2001 than in 1997, and 40 percent higher in 2002.

In Table 11, we reproduce Table 10 but we include Lusaka in the estimation. There are fewer cotton growers in Lusaka, but enough to allow us to check if results are sensitive to the

inclusion of those farmers in our model. Table 11 confirms that the estimated impacts are essentially unchanged. For Model 1, for example, the coefficients of  $F^1$  and  $F^2$  in column (1) are  $-0.522$  and  $0.202$ , respectively, close to what we found before ( $-0.519$  and  $0.202$ ). Similarly, the coefficients of  $R^1$  and  $R^2$  (column 2) are  $-0.287$  and  $0.167$  (similar to  $-0.298$  and  $0.162$  in Table 8). Finally, the pattern of year phases are also similar to those estimated before (column 3): there is an increase in productivity in 1998, a decline in 1999 and a sharper decline in 2000, and finally significant increases in 2001 and 2002.<sup>12</sup>

We have shown evidence indicating that cotton productivity followed different patterns in different regions of the country. In Table 12, we report estimates of the model that account for these differences. Concretely, we estimate a separate model for each of the three main cotton producing provinces. The first three columns of the table reproduce the benchmark results at the national level. When the original definition of the phases of the reforms is used,  $F^1$  and  $F^2$ , the estimated regional coefficients track the national coefficients: they are negative and significant during the failure phase and positive and significant during the success phase. Notice, however, that the magnitudes are very different. In particular, much more pronounced changes are observed in the Southern province. For example, whereas the decline in productivity during the failure phase is of around 42 and 44.5 percent in the Central and Eastern provinces, respectively, it is of 96.5 percent in the Southern province. This means that the coefficient of  $F^1$  almost double (in absolute value) in the Southern province. The coefficient of  $F^2$  varies from region to region as well, from 0.303 and 0.106 in the Central and Eastern provinces, to 0.554 percent in the Southern provinces.

Some interesting differences are also observed when we use alternative definitions of the phases of the reforms. This is specially so under  $R^1$  and  $R^2$ .<sup>13</sup> In the Central province, for example, there is a large increase in productivity during the success phase, but no statistically significant changes during the failure phase of 2000. In contrast, in the Eastern province there is a significant decline in the failure phase of 2000 ( $R^1$ ), but there is not any significant change during the success phase. Finally, the Southern province shown a sharp decline (of

---

<sup>12</sup>These results are robust to the specification and model used in the cotton land share ( $a_{ht}^c$ ) equation (Model 1 or Model 2, for instance).

<sup>13</sup> $R^1$  includes only 2000, and  $R^2$  includes 2001 and 2002 (as does  $F^2$ ). The difference is that the introductory phase now includes 1999.

34.8 percent) in failure phase, and a sharp increase (of 55.7 percent) in the success phase.

## 4.6 Additional Unobservables in Selection

Let us assume that the cotton land share model is given by

$$(10) \quad a_{ht}^c = \mathbf{z}_{ht}' \gamma_t + \delta_{rt}^a + \phi_{ht} + u_{ht}.$$

This equation includes  $u_{ht}$ , together with  $\phi_{ht}$ , in the error term to capture potential additional unobservables that affect the selection into cotton but not the productivity equation. The implication of this model for our purposes is that our proxy of unobserved productivity is now estimated with error (see Altonji, 1986).

The problem resembles estimation under measurement error. In principle, these problems are corrected with instrumental variables. Notice that, in our case, we need to instrument  $\widehat{\phi}_{ht} + \widehat{u}_{ht}$ . Since we do not have instruments for this variable, but we follow the procedures used in Monte Carlo analysis of measurement error. If we knew the variance of the measurement error—that is, the variance of  $u$ —, then it would be in principle possible to correct the OLS estimates to get consistent estimates. The problem is precisely that the variance of  $u$  is not known. In Monte Carlo studies, the model is estimated under many different assumption about the variance. If the estimated coefficients do not change much with  $\sigma_u^2$ , then there is evidence that the measurement error is not generating significant inconsistencies.

Our results are reported in Table 13. We report the estimates of  $\alpha_1$ ,  $\alpha_2$  and  $a_0$  under eleven different assumptions about  $\sigma_u^2$ . We confirm that the coefficients of the phases of the reform, and the unobserved productivity remain relatively unaffected by the potential measurement error. We believe that this is evidence that the problem can be safely discarded and that our results are not sensitive to it.

## 5 Conclusions

This paper has investigated the relationship between market structure in cotton and farm productivity. We have used unique farm surveys for rural Zambia, the Post Harvest Survey, spanning the 1997-2002 period. We have exploited a marketing reform whereby the Zambian government eliminated the cotton marketing board in 1994. Entry and exit into the market and the development of the outgrower scheme gave rise to interesting dynamics in market organization. Starting with a baseline period in 1997-1998, there was a subsequent failure of the outgrower scheme in 1999 and 2000. Further entry and competition into the sector led to an improvement in the outgrower scheme in 2001-2002.

We have estimated the impacts of the different phases of the reforms by building a modified difference-in-differences estimator. The first differences are taken across the different phases of the reforms. An equation of maize productivity, a major staple produced by virtually all households, provides the second difference. In the presence of entry and exit into cotton farming, and in the presence of cotton-specific farm unobservables, the estimated average productivity can be biased. To correct for these dynamics effects, we introduce a model of selection into cotton that provides proxies for unobserved productivity. 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. This modified difference-in-differences model delivers consistent estimate of the impacts of the reforms on farm productivity.

We find interesting dynamic effects of the marketing reforms. Compared to the introductory phase of 1997-1998, the failure of the outgrower scheme caused farmers to move back to subsistence and led to significant reductions in farm productivity. The improvement of the outgrower scheme in 2001-2002, reverted these trends: farmers allocated more land to cotton, and productivity (i.e., yields per hectare) significantly increased.

## Appendix 1: Cotton Selection Models

Table A.1 reports a set of results of the selection equation. These estimates are obtained from a Tobit model. Qualitatively similar results are estimated with OLS or the CLAD models. We find that household assets are positively linked to land cotton shares. Total land and whether the household raise livestock can work as collateral perhaps allowing the

household to obtain cheaper credit and to purchase inputs or to afford any initial investment. In addition, household assets may allow farms to adopt riskier (but also more profitable) agricultural activities.

The size of the family also affects cotton allocation positively. One explanation is that bigger households can take care of own-consumption needs (food security) and have additional resources needed to embark in cash agriculture. A related finding in Table A.1 is that household with higher proportion of males tend to allocated higher shares of land to cotton. This is consistent with the notion that the availability of labor supply matters in the choice of crops.

Finally, there is some evidence that male-headed households tend to grown more cotton than female-headed families.

## References

Altonji, J. (1986). "Intertemporal Substitution in Labor Supply: Evidence From Micro Data," *Journal of Political Economy* vol. 94, no 3, part 2, pp. S176-S215.

Balat, J. and G. Porto (2005). "Globalization and Complementary Policies. Poverty Impacts in Rural Zambia," in A. Harrison ed. (2005).

Benjamin, D. (1994). "Can Unobserved Land Quality Explain the Inverse Productivity Relationship?," *Journal of Development Economics* 46, 51-84.

Besley, T. and A. Case (1994). "Diffusion as a Learning Process: Evidence From HYV Cotton," mimeo, Princeton University.

Binswanger, H., and D. Sillers (1983). "Risk Aversion and Credit Constraints in Farmers' Decision-Making: A Reinterpretation," *Journal of Development Studies*, 20, pp. 133-140.

Conley, T. and C. Udry (2004). "Learning About a New Technology: Pineapple in Ghana," mimeo, Yale University.

Cotton News (2002). *Cotton Development Trust*, Zambia.

de Janvry, A., M. Fafchamps, and E. Sadoulet (1991). "Peasant Household Behaviours with Missing Markets: Some Paradoxes Explained," *Economic Journal*, 101, pp. 1400-1417.

- Deaton, A. (1997). *The Analysis of Household Surveys. A Microeconometric Approach to Development Policy*, John Hopkins University Press for the World Bank.
- Dercon, S. (1996). "Risk, Crop Choice, and Savings: Evidence from Tanzania," *Economic Development and Cultural Change*, XXX, pp. 485-513.
- Edmonds, E. and N. Pavcnik (2004). "The Effects of Trade Liberalization on Child Labor," forthcoming *Journal of International Economics*.
- Eswaran, M. and A. Kotwal (1986). "Access to Capital and Agrarian Production Organization," *Economic Journal*, 96, pp. 482-498.
- Fafchamps, M. (1992). "Cash Crop Production, Food Price Volatility, and Rural Market Integration in the Third World," *American Journal of Agricultural Economics*, XX, pp. 90-99.
- Fan, J. (1992). "Design-adaptive nonparametric regression," *Journal of the American Statistical Association*, vol. 87, No 420, December, pp. 998-1004.
- Feder G. (1980). "Farm Size, Risk Aversion, and the Adoption of New Technologies Under Uncertainty," *Oxford Economic Papers*, 32, pp. 263-283.
- Feder G. (1985). "The Relationship Between Farm Size and Farm Productivity," *Journal of Development Economics*, 18, pp. 297-313.
- Food Security Research Project (2000). "Improving Smallholder and Agribusiness Opportunities in Zambia's Cotton Sector: Key Challenges and Options," Working Paper No 1, Lusaka, Zambia.
- Foster, A. and M. Rosenzweig (1995). "Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture," *Journal of Political Economy*, vol. 103, no. 6, pp. 1176-1209.
- Harrison, A. (2005). *Globalization and Poverty*, National Bureau of Economic Research, Boston, Massachusetts.

- Jayne, T. (1994). "Do High Food Marketing Costs Constrain Cash Crop Production? Evidence from Zimbabwe," *Economic Development and Cultural Change*, XXX, pp. 387-402.
- Levinsohn, J. and A. Petrin (2003). "Estimating Production Functions Using Inputs to Control for Unobservables," *Review of Economic Studies*, 70, pp. 317-341.
- Olley, G. and A. Pakes (1996). "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, vol. 64, no 6, pp. 1263-1297.
- Pagan, A. and A. Ullah (1999). *Nonparametric Econometrics*. Cambridge University Press, New York.
- Pavcnik, N. (2003). "Trade Liberalization, Entry and Productivity Improvements: Evidence From Chilean Plants," *Review of Economic Studies*, vol. 69, No. 1, pp. 245-276.
- Robinson, P.M. (1988). "Root- $N$ -Consistent Semi-Parametric Model," *Econometrics*, vol. 56, No. 4, pp. 931-954.
- Rosenzweig, M. and H. Binswanger (1993). "Wealth, Weather Risk and the Composition and Profitability of Agricultural Investments," *The Economic Journal*, vol. 103, No. 416, pp. 56-78.
- Shahabuddin, Q., S. Mestelman, and D. Feeny (1986). "Peasant Behaviour Towards Risk and Socio-Economic and Structural Characteristics of Farm Households in Bangladesh," *Oxford Economic Papers*, 38, pp. 135-152.

Table 1  
Post Harvest Survey  
(sample sizes)

Province	1997	1998	1999	2000	2001	2002
Central	654	674	648	795	663	701
Eastern	1,225	1,197	1,255	1,437	1,248	1,292
Southern	895	828	835	961	835	850
Lusaka	246	252	243	244	185	182
Copperbelt	370	349	379	464	367	372
Luapula	803	775	799	869	760	761
Northern	1,211	1,190	1,348	1,551	1,293	1,376
Nwestern	409	423	429	543	435	431
Western	706	648	725	835	699	733
Total	6,519	6,336	6,661	7,699	6,485	6,698

Note: Own calculations based on the Post Harvest Surveys 1997-2002.

Table 2  
Percentage of Farmers Growing Cotton  
1997 - 2002

Province	1997	1998	1999	2000	2001	2002
Central	24.6	22.6	16.6	10.3	14.7	20.2
Eastern	35.2	32.7	31.7	20.4	32.1	39.0
Southern	9.9	10.7	11.7	4.3	8.8	12.8
Lusaka	5.4	3.3	4.7	0.4	5.1	8.2
Copperbelt	0.8	0.6	0.3	0.0	0.0	0.0
Luapula	0.0	0.0	0.0	0.0	0.0	0.0
Northern	0.0	0.1	0.0	0.0	0.0	0.0
NWwestern	0.3	0.0	0.0	0.0	0.0	0.2
Western	1.3	0.6	0.4	0.1	0.1	0.0
Total	11.0	10.4	9.4	5.4	9.0	11.6

Note: Own calculations based on the Post Harvest Surveys 1997-2002.

Table 3  
Percentage of Households that Grow Maize  
1997 - 2002

Province	1997	1998	1999	2000	2001	2002
Central	90.2	92.2	93.5	94.3	94.0	93.2
Eastern	99.9	99.5	99.2	99.5	99.7	99.6
Southern	93.5	92.0	94.4	96.3	97.3	97.6
Lusaka	100.0	99.5	98.9	100.0	98.9	99.4
Copperbelt	96.7	94.3	90.9	93.5	93.5	93.6
Luapula	28.5	24.3	31.6	35.1	31.8	41.1
Northern	45.1	35.9	48.8	46.9	46.9	59.3
NWestern	75.7	65.9	71.9	66.7	80.7	77.7
Western	89.9	82.2	88.5	82.6	90.1	87.2
Total	76.2	72.1	76.1	76.2	77.7	80.7

Note: Own calculations based on the Post Harvest Surveys 1997-2002.

Table 4  
Percentage of Households that Grow Maize  
Conditional on Growing Cotton  
1997 - 2002

Province	1997	1998	1999	2000	2001	2002
Central	96.7	95.7	95.1	100.0	97.9	99.3
Eastern	100.0	98.4	98.7	99.7	99.5	100.0
Southern	97.6	90.5	96.8	97.4	93.1	92.4
Lusaka	100.0	100.0	88.9	100.0	100.0	100.0
Total	98.8	96.6	97.7	99.5	98.4	98.8

Note: Own calculations based on the Post Harvest Surveys 1997-2002.

Table 5  
Fraction of Land Allocated to Cotton  
1997 - 2002

Province	1997	1998	1999	2000	2001	2002
Total	9.2	9.3	8.1	4.3	7.6	9.9
Central	12.1	10.7	6.7	3.5	6.3	8.5
Eastern	12.4	13.0	12.3	7.2	11.9	14.6
Southern	4.1	4.2	3.7	1.3	3.2	5.1
Lusaka	2.4	1.4	1.7	0.1	1.9	3.3

Note: Own calculations based on the Post Harvest Surveys 1997-2002.

Table 6  
Yields per Hectare in Cotton  
1997 - 2002

Province	1997	1998	1999	2000	2001	2002
Total	6.18	6.53	6.38	6.21	6.44	6.39
Central	6.33	6.67	6.72	7.04	6.98	6.73
Eastern	6.14	6.45	6.28	6.07	6.32	6.32
Southern	6.09	6.65	6.40	5.56	6.57	6.23
Lusaka	6.00	6.40	6.43	6.40	5.51	6.57

Note: Own calculations based on the Post Harvest Surveys 1997-2002.

Table 7  
Basic Productivity Regression

	Simple Difference		Difference-in-Differences	
	(1)	(2)	(3)	(4)
age	0.005	0.005	0.014	0.01
	0.006	0.006	0.008*	0.007
age squared	0	0	0	0
	0	0	0.000**	0.000*
farm size	-0.109	-0.061	0.12	0.057
	0.038***	0.039	0.043***	0.044
head male?	0.121	0.124	0.093	0.088
	0.040***	0.040***	0.047**	0.047*
family size	0.103	0.079	-0.037	0.007
	0.031***	0.031**	0.036	0.036
share of males	-0.014	-0.033	-0.035	-0.038
	0.087	0.087	0.106	0.105
livestock	0.132	0.091	-0.063	-0.06
	0.032***	0.032***	0.038*	0.04
size cotton plot	-0.292	-0.307		
	0.023***	0.023***		
harrows		-0.076		-0.064
		0.046*		0.053
ploughs		0.105		0.034
		0.019***		0.027
basal		0.088		-0.713
		0.109		0.257***
top-dressing		0.09		-0.621
		0.107		0.283**
relative plot size			-0.343	-0.376
			0.024***	0.024***
$F^1$	-0.157	-0.151	-0.484	-0.451
	0.053***	0.053***	0.057***	0.057***
$F^2$	0.084	0.119	0.188	0.208
	0.028***	0.030***	0.036***	0.036***
Constant	6.172	6.119	-0.882	-0.703
	0.167***	0.166***	0.204***	0.202***
Observations	3476	3476	3418	3418
$R^2$	0.11	0.12	0.11	0.14

Note: Robust standard errors in parentheses: \* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%.

Table 8  
Cotton Productivity  
Entry and Exit in Cotton Farming

	Model 1			Model 2			
	(1)	Tobit (2)	OLS (3)	CLAD (4)	Tobit (5)	OLS (6)	CLAD (7)
age	0.01 0.007	0.01 0.008	0.009 0.007	0.009 0.008	0.009 0.008	0.009 0.007	0.01 0.008
age squared	0 0.000*	0 0.000*	0 0.000*	0 0.000	0 0.000*	0 0.000*	0 0
farm size	0.057 0.044	0.027 0.045	0.019 0.044	0.039 0.044	0.024 0.045	0.016 0.044	0.032 0.044
head male?	0.088 0.047*	0.114 0.048**	0.103 0.047**	0.111 0.048**	0.116 0.048**	0.104 0.047**	0.113 0.048**
family size	0.007 0.036	0.027 0.037	0.027 0.037	0.026 0.037	0.029 0.037	0.028 0.037	0.025 0.037
share of males	-0.038 0.105	-0.023 0.105	-0.031 0.105	-0.033 0.105	-0.022 0.105	-0.031 0.105	-0.041 0.104
livestock	-0.06 0.04	-0.035 0.041	-0.034 0.04	-0.044 0.040	-0.032 0.041	-0.033 0.04	-0.038 0.041
relative plot size	-0.376 0.024***	-0.42 0.030***	-0.562 0.045***	-0.43 0.032***	-0.424 0.029***	-0.565 0.044***	-0.436 0.033***
harrows	-0.064 0.053	-0.055 0.053	-0.056 0.053	-0.057 0.053	-0.057 0.053	-0.065 0.053	-0.063 0.053
ploughs	0.034 0.027	0.041 0.027	0.04 0.027	0.035 0.027	0.04 0.027	0.039 0.027	0.037 0.027
basal	-0.713 0.257***	-0.693 0.252***	-0.672 0.250***	-0.695 0.252***	-0.691 0.252***	-0.669 0.250***	-0.696 0.252**
top-dressing	-0.621 0.283**	-0.598 0.277**	-0.6 0.274**	-0.597 0.277***	-0.596 0.276**	-0.601 0.274**	-0.596 0.276**
$F^1$	-0.451 0.057***	-0.519 0.061***	-0.535 0.058***	-0.496 0.058***	-0.523 0.061***	-0.536 0.058***	-0.508 0.059***
$F^2$	0.208 0.036***	0.202 0.036***	0.209 0.036***	0.203 0.036***	0.202 0.036***	0.208 0.036***	0.202 0.036***
$\phi$		0.002 0.001**	0.01 0.002***	0.003 0.001**	0.003 0.001***	0.01 0.002***	0.003 0.001**
Constant	-0.703 0.202***	-0.902 0.212***	-1.031 0.208***	-0.704 0.201***	-0.912 0.213***	-1.027 0.209***	-0.717 0.202***
Observations	3418	3418	3418	3418	3418	3418	3418
$R^2$	0.14	0.15	0.15	0.15	0.15	0.15	0.15

Robust standard errors in parentheses: \* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%.

Column (1) does not include  $\phi$ .

Tobit, OLS and CLAD refer to different models used to estimate  $\phi$ . See Appendix 1.

Model 1: includes total land tenure, family size, age, age squared, farm size, a dummy for male-headed farms, the proportion of males in the family, a dummy for livestock raising households, and assets (harrows, ploughs) in monetary units.

Model 2: includes total land tenure, family size, age, age squared, farm size, a dummy for male-headed farms, the proportion of males in the family, a dummy for livestock raising households, and assets (harrows, ploughs) in physical units.

Table 9  
Cotton Productivity  
Non-Linearity of Unobserved Productivity

	Model 1		Model 2	
	Tobit (1)	Robinson (2)	Tobit (3)	Robinson (4)
age	0.01 0.008	0.0099 0.0075	0.009 0.008	0.0098 0.0075
age squared	0 0.000*	-0.0001 0.00007*	0 0.000*	-0.0001 0.00007*
farm size	0.027 0.045	0.035 0.043	0.024 0.045	0.032 0.043
head male?	0.114 0.048**	0.12 0.053**	0.116 0.048**	0.117 0.053**
family size	0.027 0.037	0.023 0.036	0.029 0.037	0.024 0.037
share of males	-0.023 0.105	-0.024 0.101	-0.022 0.105	-0.024 0.101
livestock	-0.035 0.041	-0.035 0.041	-0.032 0.041	-0.032 0.041
relative plot size	-0.42 0.030***	-0.421 0.028***	-0.424 0.029***	-0.424 0.028***
harrows	-0.055 0.053	-0.054 0.048	-0.057 0.053	-0.055 0.048
ploughs	0.041 0.027	0.041 0.021**	0.04 0.027	0.041 0.021**
basal	-0.693 0.252***	-0.0007 0.0002***	-0.691 0.252***	-0.0007 0.0001***
top-dressing	-0.598 0.277**	-0.0006 0.0001***	-0.596 0.276**	-0.0006 0.0001***
$F^1$	-0.519 0.061***	-0.523 0.058***	-0.523 0.061***	-0.527 0.059
$F^2$	0.202 0.036***	0.202 0.037**	0.202 0.036***	0.201 0.037***
$\phi$	0.002 0.001**	yes -	0.003 0.001***	yes -
Constant	-0.902 0.212***	-0.006 0.016	-0.912 0.213***	-0.006 0.016
Observations	3418	3418	3418	3418
$R^2$	0.15	0.13	0.15	0.13

Robust standard errors in parentheses: \* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%.

For a description of Models 1 and 2, see note in Table 8.

Columns (1) and (3) use a Tobit procedure in the selection model and OLS in the productivity model. Columns (2) and (4) use a Tobit procedure in the selection model and a partially linear, Robinson procedure in the productivity model.

Table 10  
Cotton Productivity  
Sensitivity to the Definition of the Reform

	Model 1			Model 2		
	(1)	(2)	(3)	(4)	(5)	(6)
$\phi$	0.002	0.001	0.003	0.003	0.001	0.003
	0.001**	0.001	0.001***	0.001***	0.001	0.001***
$F^1$	-0.519			-0.523		
	0.061***			0.061***		
$F^2$	0.202			0.202		
	0.036***			0.036***		
$R^1$		-0.298			-0.3	
		0.045***			0.045***	
$R^2$		0.162			0.162	
		0.039***			0.039***	
dummy 1998			0.337			0.336
			0.054***			0.054***
dummy 1999			0.023			0.023
			0.059			0.059
dummy 2000			-0.408			-0.412
			0.068***			0.069***
dummy 2001			0.203			0.201
			0.055***			0.055***
dummy 2002			0.402			0.401
			0.054***			0.054***
Constant	-0.902	-0.691	-0.982	-0.912	-0.701	-0.989
	0.212***	0.212***	0.213***	0.213***	0.213***	0.213***
Observations	3418	3418	3418	3418	3418	3418
$R^2$	0.15	0.14	0.16	0.15	0.14	0.16

Robust standard errors in parentheses: \* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%.  
 Model 1: includes total land tenure, family size, age, age squared, farm size, a dummy for male-headed farms, the proportion of males in the family, a dummy for livestock raising households, and assets (harrows, ploughs) in monetary units.  
 Model 2: includes total land tenure, family size, age, age squared, farm size, a dummy for male-headed farms, the proportion of males in the family, a dummy for livestock raising households, and assets (harrows, ploughs) in physical units.

Table 11  
Cotton Productivity  
Sensitivity to the Sample

	Model 1			Model 2		
	(1)	(2)	(3)	(4)	(5)	(6)
$\phi$	0.003	0.001	0.003	0.003	0.001	0.003
	0.001***	0.001	0.001***	0.001***	0.001	0.001***
$F^1$	-0.522			-0.527		
	0.061***			0.061***		
$F^2$	0.202			0.202		
	0.036***			0.036***		
$R^1$		-0.287			-0.289	
		0.045***			0.045***	
$R^2$		0.167			0.167	
		0.039***			0.039***	
dummy 1998			0.338			0.336
			0.054***			0.054***
dummy 1999			0.038			0.037
			0.059			0.059
dummy 2000			-0.407			-0.411
			0.068***			0.069***
dummy 2001			0.196			0.194
			0.055***			0.055***
dummy 2002			0.415			0.414
			0.054***			0.054***
Constant	-0.857	-0.638	-0.942	-0.872	-0.652	-0.953
	0.215***	0.215***	0.215***	0.215***	0.215***	0.215***
Observations	3462	3462	3462	3462	3462	3462
$R^2$	0.15	0.14	0.16	0.15	0.14	0.16

Robust standard errors in parentheses: \* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%.  
 Model 1: includes total land tenure, family size, age, age squared, farm size, a dummy for male-headed farms, the proportion of males in the family, a dummy for livestock raising households, and assets (harrows, ploughs) in monetary units.  
 Model 2: includes total land tenure, family size, age, age squared, farm size, a dummy for male-headed farms, the proportion of males in the family, a dummy for livestock raising households, and assets (harrows, ploughs) in physical units.

Table 12  
Cotton Productivity  
Regional Analysis

	All Provinces			Central			Eastern			Southern		
$\phi$	0.002	0.001	0.002	0.008	0.006	0.008	0	-0.001	0	0.002	-0.001	0.003
$F^1$	0.001**	0.001	0.001***	0.002***	0.002***	0.002***	0.001	0.001	0.001	0.003	0.002	0.003
	-0.513			-0.42			-0.445			-0.965		
$F^2$	0.060***			0.138***			0.066***			0.287***		
	0.2			0.303			0.106			0.554		
$R^1$	0.036***			0.092***			0.040***			0.144***		
		-0.298			-0.111			-0.333			-0.348	
$R^2$		0.045***			0.105			0.052***			0.156**	
		0.162			0.351			0.033			0.557	
$y_{97}$		0.039***			0.097***			0.043			0.158***	
			0.339			0.213			0.302			0.859
$y_{98}$			0.054***			0.112*			0.062***			0.211***
			0.025			0.187			-0.076			0.37
$y_{99}$			0.059			0.133			0.07			0.190*
			-0.401			-0.327			-0.373			-0.588
$y_{00}$			0.068***			0.159**			0.074***			0.314*
			0.203			0.143			0.123			0.754
$y_{01}$			0.055***			0.134			0.060**			0.251***
			0.399			0.585			0.221			1.066
			0.054***			0.128***			0.060***			0.213***
Obs	3418	3418	3418	669	669	669	2339	2339	2339	410	410	410
$R^2$	0.15	0.14	0.16	0.19	0.18	0.2	0.13	0.13	0.15	0.2	0.18	0.24

Robust standard errors in parentheses: \* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%.  
Tobit selection equations are run by province including: logtotarea logtotpop age age2 smallshmale raisistik valueharrows valueploughs (note that headmale is not included, Tobit did not converge).

In the first 3 columns (all provinces), Tobit regressions are run by province.

Table 13  
 Additional Unobservables in the Selection Model  
 Sensitivity to Measurement Error

	0	1	4	9	16	25	36	49	64	81	100	200
$\sigma_u^2$	0											
$F^1$	-0.519	-0.519	-0.519	-0.520	-0.521	-0.523	-0.525	-0.528	-0.531	-0.536	-0.541	-0.587
$F^2$	0.202	0.202	0.202	0.202	0.202	0.202	0.202	0.202	0.201	0.201	0.201	0.197
$\phi$	0.0025	0.0025	0.0025	0.0026	0.0026	0.0027	0.0027	0.0029	0.0030	0.0031	0.0033	0.0051

Estimates impacts of the reforms under different assumptions about the variance of the measurement error,  $\sigma_u^2$ .

Table A.1  
Determinants of Land Cotton Shares

	1997	1998	1999	2000	2001	2002
total land	29.30	33.07	34.51	34.20	31.69	23.53
	2.11	2.40	2.51	3.35	2.43	1.94
family size	2.53	4.77	0.01	0.26	-4.87	2.37
	2.42	2.72	2.62	3.11	2.70	2.49
age	-1.00	-0.21	-0.86	-0.45	-0.08	-0.68
	0.51	0.57	0.51	0.70	0.56	0.49
age squared	0.00	0.00	0.00	0.00	0.00	0.00
	0.01	0.01	0.01	0.01	0.01	0.00
farm size	1.26	-0.63	3.06	23.41	4.32	-0.27
	3.03	3.82	3.80	5.41	3.45	3.10
head male?	9.79	10.70	6.59	5.87	11.54	7.90
	3.56	3.88	3.69	4.45	3.76	3.04
share of males	18.02	11.72	12.87	-0.11	6.13	-9.04
	6.82	7.47	7.41	8.76	7.02	6.42
livestock	-1.67	3.61	6.64	9.33	3.80	3.18
	2.71	2.99	3.07	3.59	3.05	2.80
value harrows	0.36	3.09	2.72	1.00	-	1.73
	2.51	2.40	1.65	2.53	-	1.39
value ploughs	-2.04	-4.32	-3.15	0.21	-	0.43
	1.11	1.05	0.84	0.74	-	0.57
constant	-28.43	-59.26	-51.40	-82.93	-51.00	-21.36
	14.01	15.52	14.50	18.43	15.43	13.57

Tobit estimates of land cotton shares. Includes dummy district variables. A separate regression is run in each year to account to macro shocks, prices, and the reforms. Since no information on assets was collected in 2001, the Tobit specification for that year does not include the value of harrows and the value of ploughs.

Figure 1  
Average Productivity  
Entry and Exit into Cotton Farming

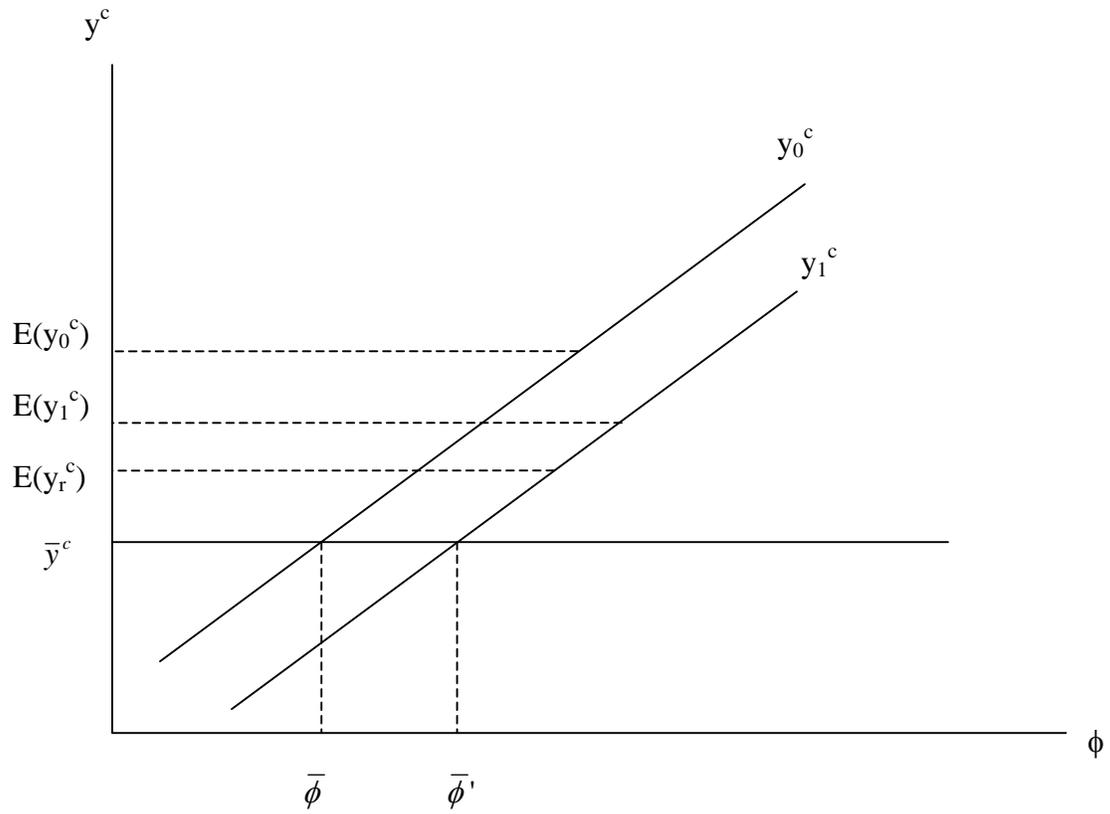
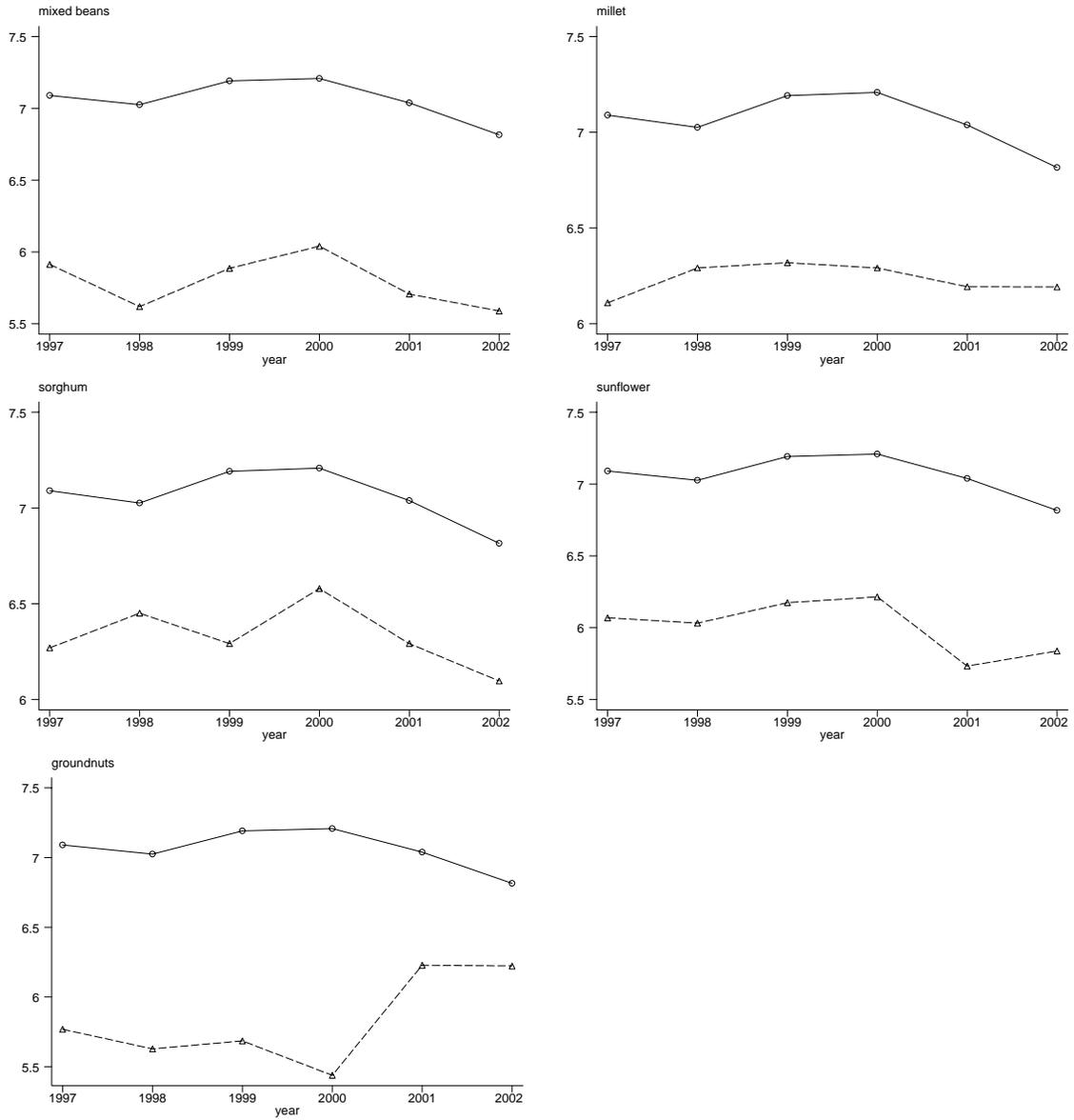
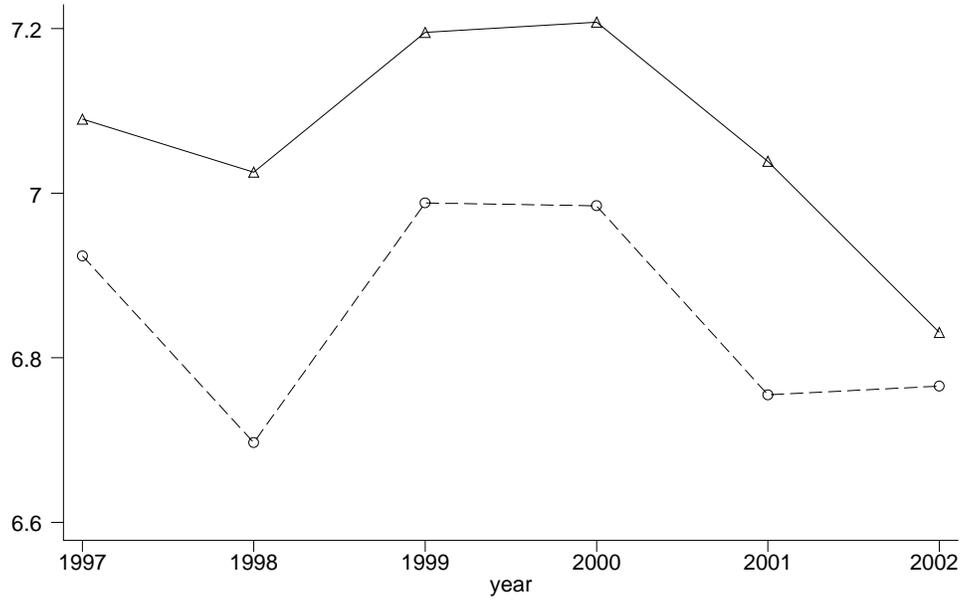


Figure 2  
Trends in Agricultural Productivity  
Maize, Mixed Beans, Millet, Sorghum, Sunflower, and Groundnuts



Note: The graphs compare the trend in maize productivity with the trends in productivity in alternative crops. Starting at the top-left, the panels represent the cases of Mixed Beans, Millet, Sorghum, Sunflower, and Groundnuts, respectively. In each panel, the solid line represents the trends in maize productivity and the broken line, the trend in the productivity in the alternative crops.

Figure 3  
Trends in Maize Productivity  
Reform Provinces versus Non-Reform Provinces



Note: The graph reports the trends in maize productivity in reform provinces (solid line) and non-reform provinces (broken line). Estimates based on the Post Harvest Survey.