Barriers for Development in Zambian Small- and Medium-Size Farms: Evidence from Micro-Data

Prepared for Presentation at the 2000 Annual Meetings of the American Agricultural Economics Association

by

Ayal Kimhi
Department of Agricultural Economics and Management
Hebrew University

and

Dennis Chiwele
Institute for Economic and Social Research
University of Zambia

Abstract

The objective of this paper is to identify factors which limit the ability of Zambian farmers to increase Maize productivity and/or diversify their crop mix. Both may enable wealth accumulation, investments, and further expansion. Specifically, we link variations in agricultural decisions, practices, and outcomes, to variations in the tightness of the different constraints. We model crop production decisions as having recursive structure. Initially, farmers decide on land allocation among the different crops, based on their information set at planting time. Then, as new information (weather, market conditions) is revealed, farmers can change output by influencing the yield. This recursive structure enables to separate the effects of the constraints on the different stages of production.

We therefore conduct estimation in two stages: we first estimate the fraction of land allocated to Maize as a dependent variable that is censored from below and from above, so that its predicted value is necessarily between zero and one. The yield of Maize is estimated in the second stage as a linear function of calculated land allotment (to avoid simultaneity bias) and the other state variables. Environmental and demographic variables also serve as explanatory variables in each stage.

The first-stage results indicate that crop diversification can be promoted by rural road construction, developing markets for agricultural products, increasing the availability of seeds, draught animals, and farm machines, increasing women's farm work participation, and increasing the size of landholdings. Specialization in Maize can be promoted by increasing the availability of credit, fertilizers, hired permanent workers, and irrigation knowledge, and improving the timeliness of input delivery.

The second-stage results show that the yield of Maize is inversely related to the area of Maize cultivated and to the operator's age, and is lower in female-headed farm households. Maize productivity can be improved by increasing the availability of seeds, fertilizers, labor, draught animals, machines, and credit.

Introduction

Zambia has vast natural and land resources that make her well positioned for agricultural development. Only 16% of an estimated 9 million hectares suitable for agriculture are regularly cultivated while only 6% of 2.5 to 3 million hectares of irrigable land are actually irrigated (Chipembe, 1990). Most parts of the country receive adequate rainfall for the production of arable crops despite reoccurrence of droughts in recent years. And yet Zambian agriculture suffers from constraints that make it fail to utilize its potential. This is shown in the low contribution of agriculture to GDP (about 20%) despite employing 60% of the labor force; in the sector growth rate averaging less than the population growth rate; and in the high food imports which make up as much as 11% of total imports during drought years. At the same time, the contribution of agriculture to exports has been far lower than its potential although significant improvements have occurred in the 1990s.

Since the early 1980s, there has been no evidence of increase in national average crop yields per hectare (Institute for African Studies, 1996). Yields have noticeably reduced in the 1990s due to exceptionally low levels of precipitation. The problem has been worsened by the sharp fall in the use of modern farm inputs such as fertilizer and hybrid seeds. These declined slightly between the early 1980s and the late 1980s. The decline was very marked thereafter and has been attributed to macroeconomic and agricultural sector policy changes, particularly the removal of subsidies on farm inputs which resulted in a sharp increase in their prices. A cut back on cheap credit made it difficult for farmers to purchase inputs as before. As a result of reliance on simple farm implements and poor farming practices by small and medium farmers, labor productivity is very low with only 0.52 hectares per farm worker being cultivated on average between 1983/84 and 1995/96 agricultural seasons. The situation in the 1990s has been worsened by the major declines in cattle (and hence oxen) population with the total number in 1993/94 at 0.74 million, less than half the 1983/84 peak of 1.88 million. Only 13% of farm households owned cattle in 1993/94, down from 20% in 1991/92.

There have been two main consequences of these declining and stagnating trends. First, farm incomes have declined, thereby worsening rural poverty. Farm incomes showed some signs of improvement in the 1980s as a result of price and subsidy policies but declined rapidly since the removal of these policies. Second, due to year-to-year variations in output for both maize and other food crops, the food security situation has deteriorated. It is clear, therefore, that a technological revolution in Zambian agriculture is required in order achieve two goals: (a) expand the land brought under cultivation and thereby maximize the advantage of Zambia as a land surplus country; and (b) raise land and labor productivity.

In order to achieve these goals, it is crucial to identify the major constraints facing Zambian farmers. Previous studies have analyzed several different constraints. For example, Holden (1993) cites the highly imperfect labor markets as the main problem in Zambian agriculture, while in the study of Jha and Hojjati (1993), credit seems to be the most limiting factor. Foster and Mwanaumo (1995) claim that "more emphasis is needed on support systems such as extension education, agricultural research, infrastructure, and marketing." They are all probably right, since all these limitations are interrelated. Modern inputs such as chemical fertilizers, hybrid seed varieties, and irrigation equipment, which are necessary for increasing labor productivity, require credit for their purchase, rural roads for their transport, and extension services for their implementation. The labor shortage can be overcome in part by increasing farm incomes so as to pull labor from other sectors, by changing the crop mix, or by shifting cropping activities to off-peak seasons (which in turn can only be achieved through irrigation).

The question is what are the most important constraints and how does each constraint affect production decisions, which in turn affect the area brought under cultivation and the yield levels that farmers obtain. In the 1995 Crop Forecast Survey conducted by the Central Statistical Office in Zambia, farmers were asked to indicate the major constraint for increasing production. 55% indicated shortage of funds for buying inputs, 27% indicated shortage of fertilizers, 14% indicated shortage of labor, 9% indicated shortage of improved seed varieties and 10% indicated late delivery of inputs.[1] It seems that farmers view shortage of credit as the major constraint, but it is not at all clear that its alleviation will have the greatest impact on production and productivity. It could be that many farmers are not even aware of the potential benefits of irrigation or improved seed varieties, for example. Therefore, a quantitative study of the impact of these and other constraints is needed.

The purpose of this study is to identify factors which could be influenced by policy makers in order to improve the situation of Zambian farmers on two fronts: diversifying the crop mix and increasing crop yields. Improvements in these fronts will naturally increase labor productivity and will lead to increasing land under cultivation. Our approach is to link variations in agricultural decisions, practices, and outcomes, to variations in the tightness of the different constraints. We specifically differentiate between the effect on planting decisions and the effect on the final outcome (yield). This is because of the high uncertainty prevailing in large parts of rural Zambia not only with regard to input availability and prices, but also with regard to rainfall. As more information is collected during the growing season, farmers may still change their cultivation practices accordingly.

Our analytical framework is based on the McGuirk and Mundlak (1992) framework, which relies on the recursive nature of decisions on a farm: *"...Initially, farmers decide, given* *information at planting time, how to allocate land among different crops. Farmers then can change output only by influencing yield*" (pp. 133). This recursive structure of crop production decisions is used to solve a common problem in standard estimation of production functions, that inputs of production cannot serve as explanatory variables in a yield regression. In our case, land allocation is instrumented by its determinants, given appropriate exclusion restrictions. The recursive structure also enables us to separate the effects of the various constraints on the different stages of production. For example, availability of improved seeds will likely affect the decision how much land to allocate to a certain crop, while availability of fertilizer, if not known in advance, will likely affect the yield.

Specifically, estimation will be conducted in two stages: we will first estimate equations describing the land allocated to each crop as a function of the state variables known to farmers at planting time. Between planting and harvest, farmers choose the levels of other inputs given the land allocation and perhaps other state variables that are revealed only after planting. Therefore, the yield of each crop will be estimated in the second stage as a function of calculated land allocations (to avoid the simultaneity bias) and the new vector of state variables. The estimated coefficients of the model can be used to evaluate the changes in the land allocation patterns and crop yields that can be attained by relaxing each of the constraints. Thereby the most important constraints can be identified. We start in the next section with a more detailed description of the structure of Zambian agriculture, which is important for understanding our choice of data and methods. The analytical framework is described in the subsequent section, followed by a description of the data and empirical specification, results, and conclusions.

The Structure of Agriculture in Zambia

Crops account for more than 60% of the total agricultural output in Zambia. The proportion of cultivated area devoted to different crops has varied little in the last decade, and is dominated by maize cultivation, which accounts for about 54.3%. Cereals other than maize (millet, sorghum and rice) account for 12.7%, oilseeds (groundnuts, sunflower and soybeans) - 13%, cassava - 13.1%, and other crops, mostly cash crops, 6.6%. The changed policies of recent years (removal of production and marketing subsidies) and the introduction of seasonal and regional pricing policy have made maize relatively unprofitable, particularly for farmers in remote areas. In addition, the climatic conditions have shown the susceptibility of maize to droughts. Hence, a challenge facing Zambian agriculture is the diversification into cash (preferably high value) crops and other food crops which tend to be more drought resistant.

The agricultural sector in Zambia is divided into three categories, the commercial, medium-, and small-scale sub-sectors, on the basis of the technologies applied (Government Republic of Zambia, 1994). Commercial farmers are characterized by extensive mechanization, use of modern technology and management, rear mostly exotic breeds of livestock and rely heavily on hired labor. They number less than 1,500 and are concentrated in the narrow corridor of the line-of-rail. Small-scale farmers, on the other hand, depend mostly on hand-hoe cultivation and unpaid family labor, and use little of modern farm inputs which, when used, consist mostly of chemical fertilizer and hybrid seeds on maize cultivation. There are about 600,000 farm households classified as small-scale farmers. Medium-scale farmers, also called emergent farmers, who number about 100,000 farm households, fall in between these two categories but are mostly distinguished by their use of animal power. This is a transitional phase prior to commercial farming.

Small- and medium-scale farmers contribute between 40% and 60% of agricultural output. Crops constitute about 80% of their production, while livestock, which contributed around 30% in the mid-eighties, has significantly declined due to the animal losses of the 1990s. They produce most of the food crops, i.e. maize, sorghum, millet, cassava, groundnuts, and mixed beans. The commercial sub-sector produces almost all of the wheat, 80%-85% of the soybeans, up to 75% of the Virginia tobacco, almost all of the coffee and all the horticultural crops for export. The crop yields in the small- and medium-scale sub-sectors are low and usually about half of those in the commercial sub-sector. As a result of the differences in the crop mix and yields, the difference in the value of output between commercial and small- and medium-scale farmers is remarkable.

The farming systems applied in the small- and medium-scale farming vary from location to location and have been historically shaped by agro-ecological zones grouped into three main ones with 36 sub-zones. Zone I includes the main valleys of Zambia such as the Luangwa in the east and the Gwembe in the south. It is characterized by low rainfall, short growing season, high temperatures during the growing season, and a high risk of drought. Zone II, whose climatic conditions fall in between Zones I and III, covers the central parts of the country. Zone III is in the north and encompasses Northern, Luapula, Copperbelt and Northwestern Provinces. It is a high-rainfall area with a long growing season, low probability of drought, and cooler temperatures during the growing season. There are great variations in the agronomic features within and between the three zones which makes it possible to grow a wide range of crops.

Variations in rainfall from one year to another has been an important determinant of the year-to-year changes in output. In the 1990s, there has been a reoccurrence of droughts, more than in any other decade this century, which has devastated harvests in some years and exposed the vulnerability of small- and medium-scale farmers to variations in rainfall patterns. During the 1991/92 season, which had the worst drought in many decades, rainfall averaged 375.5 millimeters and 615.3 millimeters in Zones I and II respectively. Zone III recorded 971.5 millimeters. If this pattern persists into the future, Zone III, which includes the majority of agricultural land reserves in Zambia, will become the most reliable for agricultural production. However, soils in Zone III tend to be highly acidic and hence less fertile than in Zone II. This problem can be rectified by the use of lime, but this requires an expenditure which few small farmers are able to meet. Unless an aggressive liming program is introduced, the relatively better performance of Zone III cannot be sustained.

Analytical framework

Assume that the production function for the j-th crop is $y_j = F_j (x_j , k_j ,e)$, where y is quantity of output, x is a vector of quantities of variable inputs, k is a vector of quantities of quasi-fixed inputs, and e is a vector of environmental variables (constraints, policy, weather, soil type, location, etc.). The short-run objective of the farmer is to maximize the net cash flow of the farm: $R = \Sigma p_j y_j - w\Sigma x_j$, over choices of x_j and k_j , subject to the constraints on the availability of the quasi-fixed inputs $\Sigma k_j = K$, where p is the vector of output prices, w is a vector of variable input prices, and K is a vector of total farm-level quantities of quasi-fixed inputs. "Short run" means that K is given in a certain year. If all relevant information is known prior to planting, the results of the optimization are vectors of input demands $x_j(p,w,e,K)$, $k_j(p,w,e,K)$ and output supplies $y_j(p,w,e,K)$.

Alternatively, assume that at the time of planting, the farmer has a partial information set s_1, including last year prices, early weather conditions and forecasts, other environment variables, and K or parts of it. Hence, planting decisions, i.e. land allocated to each crop (land is assumed to be one of the quasi-fixed inputs), are made according to this information set only, and the land allocated to crop j can be specified as $a_j(s_1)$. After planting, more information is revealed, such as current prices and weather, additional elements of e and K, and of course the vector of land shares a. All these in addition to the previous information, are included in the new information set, s_2. Based on the new information set, farmers choose levels of other inputs: $x_j(s_2)$, $k_j(s_2)$.

If production exhibits constant returns to scale in all inputs, then output can be expressed as a product of land input and yield: $y_j = a_jY_j(x_j,k_j)$, where Y (yield) is crop output per unit of land. Since the allocation of inputs to crops is not observed, we can substitute optimal inputs into the yield functions: $Y_j\ (s_2) = Y_j[x_j(s_2),\ k_j(s_2)]$. This specification leads to the following two-stage estimation procedure. In the first stage one estimates the land allocation equations $a_j(s_1)$. The yield equations $Y_j(s_2)$ are estimated in the second stage, after substituting the calculated value (from the first stage) of a_j in s_2, to avoid simultaneity bias. We will return to the empirical specification of these equations after the presentation of the available data in the next section.

Data and descriptive statistics

In the survey mentioned in section 1 above, farmers were asked to indicate the major constraint for increasing production. They were also asked about their access to particular services such as extension, credit and marketing channels, and about their irrigation practices. The data set includes 7269 observations, 87% of which are defined as small-scale farmers and the other 13% are defined as medium-scale farmers. The survey was matched to the 1993/4 post-harvest survey in which input-output data were collected in detail, and from which knowledge of and access to modern production techniques such as improved seed varieties and chemical fertilizers can be inferred. The post-harvest survey included 6469 farms. This data set was checked for consistency of the cropping information by checking whether a farmer who indicated that he grows a certain crop also reports a positive amount of land allocated to that crop. 5903 farms (91%) passed this test for all crops reported. The two data sets were then merged, resulting in 5329 matched observations (90% of the consistent observations in the post-harvest survey). Some other observations were excluded due to missing explanatory variables. The estimation procedure eventually used 5280 observations. Table 1 includes definitions of variables used in the analysis and their sample means.

The major crop in these farms is maize, which is grown by 84% of the farmers in the merged data set, and accounts for 78% of the cultivated land in the farms that do grow maize, and 65% overall. Hence, we have decided to concentrate on maize. We have tried to repeat the analysis for several other crops, but the results were not satisfactory and will not be reported here.

Quantitative variables include age, land used for field crops and for maize, yield of maize, number of permanent farm workers, number of male and female family members employed on the farm, number of draught animals, fraction of loans approved and received, total credit received, amount of chemical fertilizer and seeds, total wages paid to hired workers, and number of animal-drawn implements. The amounts of land, credit, fertilizer, seeds, implements, and family and hired workers are considered as quasi-fixed inputs, whose quantity is given in the short run. Fraction of loans approved and received serve as proxies for credit constraints.[2] Indexes of maize suitability and land acidity, average annual rainfall, monthly rainfall as a fraction of the average amount of rainfall in each month of the planting season, and actual monthly rainfall in the months following the planting season are also included. All these are district-level variables.[3]

Qualitative variables include sex and levels of schooling, distance to nearest road, access and distance to output market, exposure to extension services through direct and indirect channels, major constraints on increasing farm production, an irrigation dummy and the reasons for not irrigating, dummies for loan applications from agricultural financial institutions or other sources, and land accessibility. The distance variables enable us to measure the effect of infrastructure. The extension variables show that 41% of the farmers listen to agricultural programs on the radio, 27% received advice directly from an extension worker, 13% are members of a village extension group, and another 13% attended an extension demonstration, either as participants or as observers. The importance of extension services to farm productivity can be examined by looking at the effects of these variables on the yield of maize. The single most quoted constraint on increasing production is lack of funds, and only second comes inavailability of fertilizer. The third place is shared by late delivery of inputs and inavailability of labor, and last comes inavailability of seeds. Many farmers indicated that there is a different major constraint but did not indicate what it was. These variables tell us which farmers can benefit from the alleviation of each constraint. The irrigation variables show that less than 12% of the farmers irrigate their fields. There is no single major reason for not irrigating, but the reasons can tell us which farmers can benefit from increased knowledge, increased loan availability, or water projects in general.

Empirical specification and estimation procedure

The share of land devoted to maize out of the total land used for seasonal field crops is used as the dependent variable in the first stage regression. As any share variable, it is restricted to be between zero and one. Figure 1 shows the distribution of this variable. We see that many observations are concentrated in the limits of the distributions. These are farms which do not grow maize at all, and those who do not grow any field crops other than maize. As a result, we cannot treat the dependent variable as a continuous variable. Alternatively, we assume the existence of a latent continuous variable describing the amount of land that the farmer would have liked to devote for maize, divided by the total land which is available for field crops. Denote this unobserved variable as S^*, and denote the observed dependent variable as S. Then S is derived from S^* as a double-censored transformation according to

$$S = S^* - S^* \mathbf{1}(S^*<0) + (1 - S^*) \mathbf{1}(S^*>1) \quad (1)$$

where **1**() is the indicator function. Assuming that S^* is distributed normally conditional on a set of explanatory variables enables to use the double-censored Tobit model for estimation. Assuming further that $E(S^*) = X\beta$ where X is the matrix of explanatory variables and β is a corresponding vector of coefficients, the likelihood function of the model is

$$\prod_{S=0}[1-\Phi(X\beta/\sigma_S)] \prod_{0<S<1}\varphi[(Y-X\beta)/\sigma_S]/\sigma_S \prod_{S=1}[1-\Phi[(1-X\beta)/\sigma_S]] \quad (2)$$

where Φ and φ are the cumulative distribution function and the probability density function, respectively, of a standard normal random variable, and σ_S is the conditional standard deviation of S^*.

An alternative to the estimation procedure above would be to use a nonlinear regression specification that will force the calculated values to be between zero and one. Such an alternative was proposed by Papke and Wooldridge (1996). They suggest that the expectation of the fractional dependent variable S, conditional on the vector of explanatory variables X, be specified as a nonlinear function G(Xβ), where β is a vector of coefficients. As natural candidates for the function G they suggest simple cumulative distribution functions such as the logistic function

$G(X\beta)=\exp(X\beta)/[1+\exp(X\beta)]$. Whereas the nonlinear least squares method seems to be the obvious estimation procedure for this case, it could be problematic in the likely case of heteroscedasticity. As an alternative, they propose a quasi-likelihood method, in which the log-likelihood function is

$$S \cdot \log[G(X\beta)]+(1-S)\log[1-G(X\beta)]. \quad (3)$$

The resulting quasi maximum likelihood estimator is consistent and asymptotically normal regardless of the true distribution function of S conditional on X. The two likelihood functions (2) and (3) were maximized using procedures included in the Gauss software package. The results are reported in the next section.

In the second stage we want to estimate an equation describing the yield of Maize. Clearly, we can only use observations on households who were actually involved in Maize cultivation. To the extent that the decision to grow Maize and the yield of Maize are correlated due to, say, unobserved farm characteristics, one cannot estimate the yield equation without correcting for selectivity. We correct for selectivity using the Heckman (1979) procedure. First we specify the expected value of the yield Y as a linear function of land and other explanatory variables Z, $E(Y) = SL + W\alpha$ where L is the total land devoted to field crops, and α is a vector of coefficients. Then we assume that Y and S^* are jointly normally distributed, conditional on the explanatory variables. As a result, it can be shown that conditional on $S>0$, the expected value of the yield is

$$E(Y) = SL + W\alpha + \sigma_{sy}/\sigma_y \lambda \quad (4)$$

where σ_{sy} is the covariance between Y and S^*, σ_y is the standard deviation of Y, and λ is equal to the Inverse Mills Ratio $\phi(Z)/[1-\Phi(Z)]$, where $Z = X\beta/\sigma_s$. Equation (4) can be estimated by Ordinary Least Squares in the sub-sample including observations in which S>0, after S and Z are calculated using the first-stage coefficients of the double-Tobit model.

Results of the share of land devoted to Maize

The results of the Maize share equation are reported in Table 2. We compare the distributions of the predicted values of the two alternative estimation procedures to that of the actual values in order to asses the quality of the fit. We can see that the double-Tobit model is not able to correctly predict the concentration of observations in the extremes of the distribution, while the logistic transformation does so in a satisfactory way. On the other hand, the logistic transformation does not account for the peaks on the right hand side of the distribution, while the double-Tobit model does so to some extent. Hence, although the logistic transformation seems to do a somewhat better predictive job, the comparison does not result in an absolute advantage to one of the methods. Also, Table 2 reveals that the statistical significance of the coefficients of the two methods is not extremely different. Therefore, we present both sets of results and focus on signs of coefficients and their statistical significance rather than on magnitudes.[4]

Maize share of the land is decreasing in the total amount of land used for field crops. It seems like diversification into other crops has fixed costs, so the tendency to diversify is positively related to the amount of land. Maize share in smaller in female-headed households, and is positively related to the age of the farmer and his/her level of schooling. Maize share of land is lower in farms which are more than 5 kilometers from the nearest road. It is lower in farms which have access to markets, but is negatively related to the distance from the market. These two effects contradict each other and we do not have a satisfactory explanation for that.

Extension services do not seem to have significant impacts on the share of land devoted to Maize. The only significant extension variable (at the 5% level) is the dummy for attending an extension demonstration as an observer, which has a negative effect. Maize share of the land is lower in farms in which the major constraint is inavailability of seeds, and higher in farms in which the major constraint is inavailability of fertilizer or late delivery of inputs. Irrigation variables are not significant at the 5% level except for the dummy for not irrigating due to lack of knowledge, which is negative. Therefore knowledge of irrigation methods will likely increase the share of land devoted to Maize.

The number of permanent hired farm workers has a positive effect on the share of land devoted to Maize, while the number of female household members has a negative effect, and the number of male household members does not have a significant effect. The number of draught animals as well as the number of farm machines have negative coefficients. The share of land devoted to Maize is higher among farmers who applied for loans from agricultural financial institutions and lower among farmers who applied for loans from other sources. The percent of loans approved, which proxies for lack of credit constraints, has a positive coefficient. Hence credit constraints tend to decrease the share of land devoted to Maize.

The coefficients of land accessibility imply a positive effect of accessibility on the share of land devoted to Maize. The same is true for Maize suitability. Land acidity did not have a significant effect. The average annual rainfall has a negative effect on the share of land devoted to Maize. Actual rainfall in the early months of the season has a mixed effect - rainfall in July and September has a positive effect while rainfall in August has a negative effect - on the share of land devoted to Maize. Rainfall in November and December has a positive effect, with a much larger coefficient in December than in all other months combined.

Results of the yield equation

Two versions of the results of the yield regression are reported in Table 3.[5] The one on the left includes the district-specific explanatory variables. In the second version, district dummy variables replace the district-specific explanatory variables. The reason for estimating the second version is that the district-specific explanatory variables do not capture all the variation in Maize yield across districts, and to the extent that this variation is correlated with other explanatory variables, other estimated coefficients might be inconsistent.[6]

Not many coefficients of the yield equation turned out statistically significant. Some of the statistically significant coefficients are sensitive to the inclusion of district-specific explanatory variables (version 1) or district dummies (version 2). These cases will be indicated in the following discussion. We first observe that the yield of maize is inversely related to the area of maize cultivated, which implies that small Maize growers are more efficient or employ more intensive cultivation techniques, other things equal. The yield also declines with the age of the household head (version 1 only).

Farms in which lack of funds or lack of seeds is the major constraint are associated with lower yields of Maize. The same is true in farms in which there exists a major constraint different than those specified explicitly (version 1 only). Farms in which late delivery of inputs is the major constraint are associated with higher yields of Maize (version 2 only). Farmers who do not irrigate due to lack of knowledge enjoy higher yields of Maize, surprisingly (version 1 only). The number of male family workers increases the yield of Maize, and the same is true for the number of hired workers (significant at the 5% level only in version 2). The amount of credit received and the amount of chemical fertilizer used by the household have positive effects on yield, and the same is true for the number of draught animals (version 2 only). Farmers who applied for loans from sources other than agricultural financial institutions had lower yields of Maize.

The model with district-specific variables (version 1) shows that better land accessibility improves the yield of Maize, and the same is true for Maize suitability and land acidity. Among the rainfall variables, only March rainfall has a significantly positive effect on the yield. The statistically significant coefficient of lambda indicates a negative correlation between Maize share of land and the yield of Maize, so the correction for selectivity is indeed important. In version 2, however, selectivity does not seem to be important.[7]

Concluding comments

In this paper we have used an empirical framework designed for estimating production relationships in two stages, to identify the factors which affect land allocation among Maize and other crops in Zambia, and those which affect the yield of Maize. The results could be useful for the planning authorities for designing policies that could either promote farmers to diversify their crop mix or simply help them to increase the yield of Maize.

For example, according to the results, rural road construction is likely to increase field-crop diversification, and the same is true for developing markets for agricultural products, and promoting an increase of average farm land-holdings. Increasing availability of seeds will also have a similar effect. On the other hand, increased availability of fertilizers, a more timely delivery of inputs, and advancement of irrigation knowledge are likely to decrease field-crop diversification. Changes in the availability of labor, draught animals, and machinery, are also likely to affect the crop mix.

Increased availability of hired permanent workers may decrease diversification, while increased availability of draught animals and machines may increase it. Crop-mix diversification may also be achieved by increasing the participation of women in farm work. Relaxing credit constraints is likely to increase specialization in Maize. Extension services seem to affect crop diversification to some extent, but the direction of the effect is not clear.

The yield of Maize can be increased by improving the availability of seeds, fertilizers, labor, draught animals, machines, and credit. Roads, markets, extension services, and irrigation, do not seem to have significant effects on the yield of Maize.

This research can be improved if more environmental variables, such as more detailed weather conditions, become available. Another possible extension is to look at data from other periods as well. We plan to pursue these possibilities in future research.

Notes

[1] Note that farmers were allowed to indicate more than one major constraint.

[2] The problem is that loan applications may depend on credit constraints if farmers are aware of those in advance.

[3] Rainfall data were not available for all districts. Especially problematic were the 1993/94 rainfall statistics, which were available for about 57% of the farms only. We used dummy variables to control, at least in part, for the missing data.

[4] The magnitudes of coefficients are not comparable across the two estimation procedures anyway.

[5] We had to exclude a number of observations which apparently devoted land to Maize but did not report the yield.

[6] There is also a way to find the effects of district-specific variables on the yield when estimating the model with district dummies. This is accomplished by running a linear regression of the estimated district dummies on the set of district-specific variables. The use of this method was suggested by Borjas and Sueyoshi (1995). We were not able to get interesting results from this last regression and hence it is not reported here. The reason is that we had very few observations due to the missing rainfall data.

[7] Selectivity correction terms were calculated using the results of the double-tobit maize share model. The Heckman (1979) procedure is known to be vulnerable to collinearity between W and λ, yet informal tests revealed little if any collinearity in this case.

References

Chipeleme, A. C., 1990, Potential for Irrigation Development in Zambia, in: Proceedings of an International Seminar on Policies for Irrigation Development in Zambia (Ministry of Agriculture, Lusaka).

Borjas, G.J. and G.T. Sueyoshi, 1994, A Two-Stage Estimator for Probit Models with Structural Group Effects, Journal of Econometrics 64, 165-182.

Foster, K.A. and A. Mwanaumo, 1995, Estimation of Dynamic Maize Supply Response in Zambia, Agricultural Economics 12, 99-107.

Government Republic of Zambia, 1994, National Census of Agriculture (1990/92): Census Report Part 1 (Central Statistical Office, Lusaka).

Heckman, J.J., 1979, Sample Selection Bias as a Specification Error, Econometrica 47, 153-161.

Holden, S.T., 1993, Peasant Household Modelling: Farming Systems Evolution and Sustainability in Northern Zambia, Agricultural Economics 9, 241-267.

Institute for African Studies, 1996, Agricultural Sector Performance Analysis, Vol. 1 (University of Zambia, Lusaka).

Jha, D. and B. Hojjati, 1993, Fertilizer Use on Smallholder Farms in Eastern Province, Zambia, Research Perort 94, IFPRI, Washington, DC.

McGuirk, A. and Y. Mundlak, 1992, The Transformation of Punjab Agriculture: A Choice of Technique Approach, American Journal of Agricultural Economics 74, 132-143.

Papke, L.E. and J.M. Wooldridge, 1996, Econometric Methods for Fractional Response Variables with an Application to 401(K) Plan Participation Rates, Journal of Applied Econometrics 11, 619-632.

Figure 1. Distributions of Actual and Predicted Values

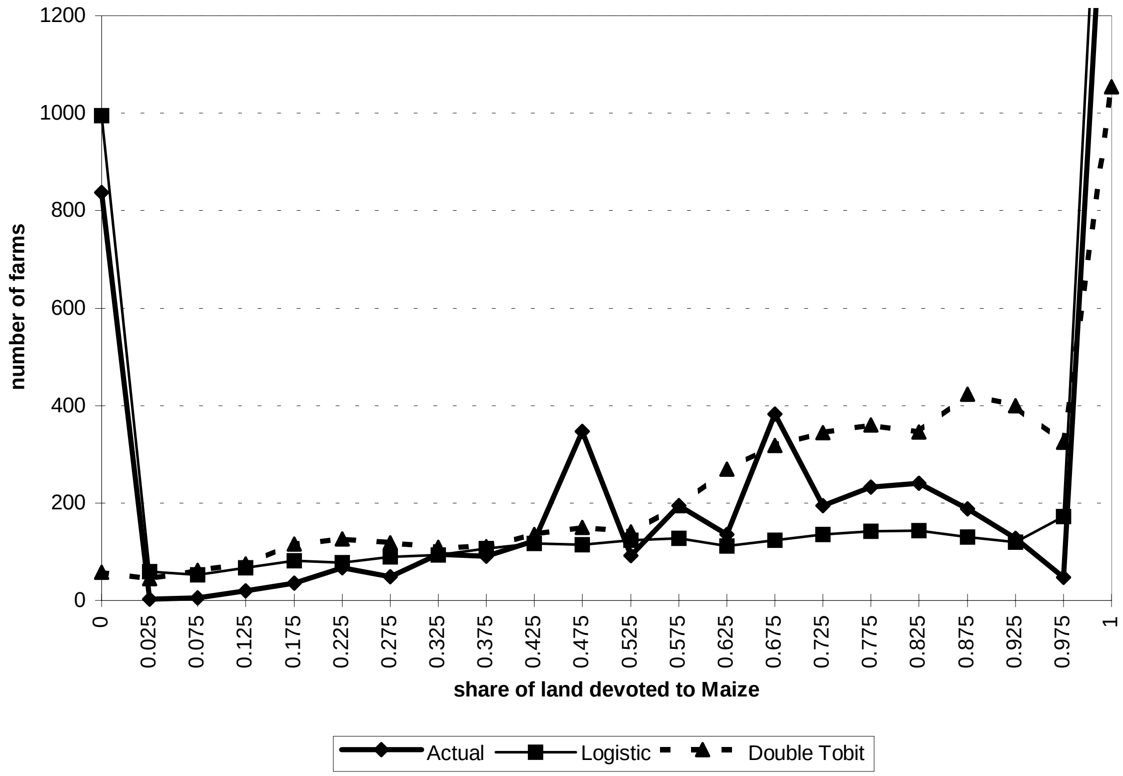

Table 1. Variables used in the estimation

Name	Description	Mean
land	total land used for seasonal field crops (Hectares)	1.78
maizelan	fraction of land used for maize	0.654
maizeyld[a]	yield of maize (100 kg/Hectare)	1.42
sex	1=female head of household	0.212
age	age of head of household (years)	44.89
primary	1=head of household has primary education	0.600
higher	1=head of household with higher than primary education	0.188
distrd5	1=nearest road is more than 5 kilometers away	0.179
outputd	1=household has access to output markets	0.726
output20+	1=distance to nearest output market is more than 20 kilometers	0.116
advice (Q8)	1=received advice directly from an extension worker	0.271
group (Q9)	1=member of a village extension group	0.135
demonstA	1=attended an extension demonstration as a participant	0.064
demonstB	1=attended an extension demonstration as an observer	0.067
radio	1=listen to agricultural programs on the radio	0.412
consfund[b]	1=major constraint on farm production is lack of funds	0.566
consseed[b]	1=major constraint on farm production is inavailability of seeds	0.083
consfert[b]	1=major constraint on farm production is inavailability of fertilizer	0.257
conslab[b]	1=major constraint on farm production is inavailability of labor	0.129
constinp[b]	1=major constraint on farm production is late delivery of inputs	0.135
consoth[b]	1=major constraint on farm production is "other"	0.208
irrigat	1=some of the land is irrigated	0.117
noirknow	1=not irrigating more due to lack of knowledge	0.294
noirfund	1=not irrigating more due to lack of funds for equipment	0.263
noirwatr	1=not irrigating more due to lack of water sources	0.277
hired[c]	number of permanent hired workers	0.032
familyM	number of male family members employed on the farm	1.77
familyF	number of female family members employed on the farm	1.89
draught	number of draught animals used on the farm	0.544
loans2a	1=applied for loans from agricultural financial institutions	0.182
loans2b+	1=applied for loans from non-agricultural financial institutions	0.040
approv%	fraction of loans approved	0.167
receiv%[a]	fraction of loans received	0.203
credit[a]	amount of credit received (10000 Kwacha)	3.41
tchem[a]	total amount of chemical fertilizers used (100 kg)	2.55
tseed[a]	total amount of seeds used (100 kg)	0.20
tvwage[a]	total amount of wages paid (1000 Kwacha)	6.02
machines	number of animal-drawn implements	0.596

continued on next page

Table 1 *(continued)*

Name	Description	Mean
access2	1=land accessibility is at degree 2 (second highest)	0.119
access34	1=land accessibility is at lowest degrees	0.063
msuit	maize suitability index	2.98
acidity	land acidity level	3.63
totrain	average annual rainfall (in mm, last six years)	889
missrain	1=missing average annual rainfall data	0.171
rrain7	rainfall in July 1993 relative to average July rainfall	0.067
rrain8	rainfall in August 1993 relative to average August rainfall	0.450
rrain9	rainfall in September 1993 relative to average September rainfall	0.936
rrain10	rainfall in October 1993 relative to average October rainfall	0.108
rrain11	rainfall in November 1993 relative to average November rainfall	1.54
rrain12	rainfall in December 1993 relative to average December rainfall	0.748
rain1	rainfall in January 1994	245
rain2	rainfall in February 1994	149
rain3	rainfall in March 1994	46.7
rain4	rainfall in April 1994	27.9
rain5	rainfall in May 1994	0.30
missr93/4	1=missing 1993/94 rainfall data	0.427

b. farmers could indicate more than one major constraint on farm production.
c. only 76 farms had permanent hired farm workers, among them the average was 1.64.

Table 2. Results of land share of Maize

	Double Tobit		Logistic Transformation	
Variable	Coefficient	T-Value	Coefficient	T-Value
intercept	0.6709	6.857	0.6791	2.694
land	-0.0163	-3.405	-0.0143	-1.218
sex	-0.0508	-2.179	-0.1306	-2.152
age	0.1500	2.325	0.3552	2.112
primary	0.0232	0.973	0.0511	0.850
higher	0.1747	5.591	0.4592	5.646
distrd5	-0.0596	-2.634	-0.1580	-2.734
outputd	-0.0743	-3.236	-0.1664	-2.825
output20+	-0.1143	-4.057	-0.2693	-3.723
advice	-0.0193	-0.834	-0.0544	-0.921
group	-0.0365	-1.168	-0.1157	-1.473
demonstA	0.0662	1.643	0.1575	1.590
demonstB	-0.0890	-2.423	-0.1961	-2.089
radio	0.0258	1.348	0.0698	1.399
consfund	0.0160	0.770	0.0729	1.350
consseed	-0.0760	-2.307	-0.2549	-2.995
consfert	0.0635	2.737	0.1914	3.227
conslab	-0.0178	-0.653	-0.0017	-0.024
constinp	0.0498	1.814	0.2117	2.976
consoth	0.0153	0.612	0.0885	1.415
irrigat	-0.0426	-1.522	-0.0958	-1.327
noirknow	-0.0465	-2.089	-0.1210	-2.120
noirfund	0.0353	1.574	0.0585	1.025
hired	0.0601	2.175	0.2088	2.206
familyM	0.0004	0.058	0.0056	0.285
familyF	-0.0162	-2.178	-0.0339	-1.829
draught	-0.0127	-1.613	-0.0355	-2.013
loans2a	0.0962	2.374	0.2595	2.549
loans2b+	-0.0994	-1.784	-0.3837	-2.786
approv%	0.1319	2.953	0.5572	4.882
machines	-0.0162	-1.896	-0.0289	-1.545
access2	-0.0598	-1.921	-0.1254	-1.579
access34	-0.1630	-4.178	-0.4279	-4.179
msuit	-0.0720	-7.800	-0.1838	-7.832

(continued on next page)

Table 2. *(continued)*

	Double Tobit		Logistic Transformation	
Variable	Coefficient	T-Value	Coefficient	T-Value
acidity	0.0052	0.607	0.0240	1.006
totrain	-0.1900	-2.122	-0.6796	-2.855
missrain	-0.0754	-2.373	-0.2149	-2.521
rrain7	0.0706	3.466	0.1258	2.408
rrain8	-0.0345	-5.269	-0.0664	-4.006
rrain9	0.0094	1.793	0.0204	1.485
rrain10	0.0037	0.082	-0.1785	-1.457
rrain11	0.0808	5.132	0.0748	1.833
rrain12	0.4909	12.656	1.3188	12.893
missr93/4	-0.0706	-2.754	-0.1626	-2.372
sigma	0.5666[a]			
# of cases	5280		5280	
log-likelihood	-4498		-3081	

a. The standard deviation coefficient was transformed in the estimation, so the standard error of the untransformed estimate is not reported.

Table 3. Results of Maize yield

	Without District Dummies		With District Dummies	
Variable	Coefficient	T-Value	Coefficient	T-Value
intercept	16.9952	12.4251	14.0025	9.9411
landmaiz	-0.7881	-6.2890	-0.8982	-6.5936
sex	-0.3319	-0.7226	-0.7578	-1.6015
age	-3.0566	-2.4267	-1.5072	-1.1440
primary	-0.4912	-1.0555	-0.1362	-0.2910
higher	0.0143	0.0228	1.5214	1.9938
distrd5	0.3076	0.6727	-0.5666	-1.1495
outputd	0.3345	0.7482	-0.4560	-0.9291
output20+	0.7401	1.1584	-0.1430	-0.1995
advice	0.5809	1.3029	0.2771	0.6116
group	-0.2318	-0.3796	-0.6885	-1.1247
demonstA	0.5453	0.7095	1.0507	1.3524
demonstB	0.8914	1.1666	0.3196	0.4005
radio	0.3467	0.9354	0.6666	1.7666
consfund	-0.9826	-2.3909	-0.8579	-2.0130
consseed	-2.4491	-3.7140	-2.2156	-3.1314
consfert	-0.0097	-0.0214	0.2185	0.4469
conslab	0.4759	0.8863	0.7169	1.2974
constinp	0.6613	1.2129	0.9153	1.6101
consoth	-1.3415	-2.7354	-0.7333	-1.4695
irrigat	0.6374	0.9389	-0.5407	-0.7700
noirknow	0.9839	2.2242	-0.2028	-0.4204
noirfund	-0.1152	-0.2672	-0.2383	-0.5275
hired	0.8157	1.3011	1.1282	1.7718
familyM	0.7972	5.4241	0.7570	5.1815
familyF	0.0152	0.1067	-0.0875	-0.5859
draught	0.3106	2.1401	0.2055	1.3835
loans2a	-0.9111	-1.1197	-0.1665	-0.1943
loans2b+	-2.0575	-1.9041	-2.7423	-2.4630
approv%	1.2057	0.8574	2.0651	1.4419
receiv%	0.3978	0.2902	0.6240	0.4583
credit	0.0928	5.9384	0.0828	5.3234
tchem	0.0561	3.0894	0.0534	2.9677
tseed	0.0887	0.4740	0.1358	0.7352
tvwage	0.0093	1.1555	0.0077	0.9733

continued on next page

Table 3 *(continued)*

	Without District Dummies		With District Dummies	
Variable	Coefficient	T-Value	Coefficient	T-Value
machines	0.2375	1.5057	0.2045	1.2516
access2	-3.3481	-5.5267		
access34	-2.3585	-3.1852		
msuit	-0.6853	-4.5589		
acidity	0.6744	3.6078		
rain712	-4.0582	-1.3737		
rain1	-2.6009	-0.7972		
rain2	1.4240	0.3987		
rain3	26.5902	7.4384		
rain4	0.9311	0.2510		
rain5	370.723	-1.4136		
missr93/4	-0.1676	-0.3754		
lambda	-7.6646	-3.1847	6.4220	1.1389
sigma	10.30		10.13	
r-squared	0.130		0.167	
adjusted R^2	0.120		0.148	
# of cases	3973		3973	
F (47,3925)	12.52		8.828	