Journal of African Economies, Vol. 20, number 5, pp. 781–822
doi:10.1093/jae/ejr017 online date 3 June 2011

Smallholder Diversification and Income Growth in Zambia†

Arne Bigsten* and Sven Tengstam

Department of Economics, University of Gothenburg, Box 640, SE405 30 Gothenburg, Sweden

* Corresponding author: Arne Bigsten. Telephone: +46-31-7861358. Fax: +46-31-7861326. E-mail: arne.bigsten@economics.gu.se

Abstract

This paper uses nationally representative panel data to analyse the relationship between income diversification and incomes for Zambian smallholder households. We show that shifting to a higher degree of diversification is generally associated with higher incomes. We find that diversification is driven by endowments and access to markets and finance. Education opens up opportunities for well-paid non-agricultural activities, while land shortage forces distress diversification into agricultural wage work.

JEL classification: O13, O55, Q10, R11

1. Introduction

Structural change has been central in the theorising about economic development typified by the dual-economy model of Lewis (1954). Economic growth has been associated with a declining agricultural share of GDP and increasing shares for industry and services. This structural change can be seen as a macroeconomic phenomenon, but it occurs also within households. Smallholders in Africa (and elsewhere) were originally

†This paper was presented at the Nordic Conference in Development Economics in Stockholm and at the conference on development economics at Chr. Michelsen Institute in Bergen.

almost exclusively farmers, but over time they have shifted to non-agricultural activities as well. Some households have shifted completely out of agriculture, although the process is usually gradual with households maintaining a foothold in agriculture for an extended period of time. This paper analyses the process of diversification, and the income implications of diversification for smallholder households in Zambia. Our contribution to the literature on diversification in Africa is, first, that we use a large, nationally representative panel survey, which is uncommon in this literature. Second, since we can use panel methods, we can at least reduce the severity of the estimation bias, although it is hard to avoid endogeneity or omitted variables bias entirely.

We distinguish four different income-generating activities—farming, agricultural wage work, non-agricultural wage work and own business—and study selection into different combinations of those. We find that diversification is driven by endowments and access to markets and finance. Endowments in the form of education open up opportunities for non-agricultural wage work. Diversification into agricultural wage work is strongly associated with land shortage, suggesting that it is distress related. We find that local demand facilitates diversification into both wage work and own business. Access to transportation mainly benefits non-agricultural wage work.

The main result of our analysis is that a shift to a higher degree of diversification in rural Zambia is associated with higher income per labourer. Changing from being a full-time farmer household to a more diversified livelihood strategy raises the per labourer income in the range of 25–100%. Since rural income growth does not depend on agriculture alone, the government needs to pursue policies that build up the endowments of households and at the same time lowers barriers to income diversification.

The remainder of this paper is organised as follows. Section 2 presents a review of empirical work on diversification in Africa, while Section 3 presents our theoretical framework and approach. Section 4 describes the data and the income variable. Section 5 reports and discusses the descriptive results on income diversification, and Section 6 presents our econometric analysis. Finally, Section 7 concludes and discusses policy implications.

2. Review of African evidence

There has been considerable work done on income diversification and livelihoods in rural Africa. One empirical regularity that has been identified is

that there is a positive relation between non-farm income and household welfare (Barrett *et al.*, 2001b). A positive association between diversification and income levels is found by Lanjouw *et al.* (2001) for Tanzania, Barrett *et al.* (2001a) for the Ivory Coast and Canagarajah *et al.* (2001) for Uganda. They also show that increased non-farm earnings lead to more rapid growth in consumption. Households that have the endowments required to enter into higher return non-farm activities are thus able to improve their incomes.

Evans and Ngau (1991), Ellis and Bahiigwa (2003), Ellis and Freeman (2004) and Lay and Mahmoud (2008) find that access to external income increases agricultural productivity and higher overall incomes in studies of Uganda, Tanzania, Malawi and Kenya. They point out that improvement in living standard is a cumulative process and that it requires an ability to build assets and to diversify. They identify the generation of cash as crucial, since it makes it possible to enhance agricultural productivity and to diversify into lucrative non-agricultural activities. Barrett *et al.* (2001b) and Ellis and Freeman (2004) note that there is a positive feedback loop, where incomes earned from non-agriculture can be ploughed back into agriculture to help increase agricultural productivity. For example, cash generated by non-agricultural activities can release the credit constraint of households, making it possible for them to buy inputs for agriculture that can raise agricultural productivity.

Another empirical regularity that has been identified is that there are substantial entry barriers to high-return activities, which tends to limit those to relatively well-endowed households. Block and Webb (2001) find in a study of rural Ethiopia in the early 1990s that wealthier households have more diversified incomes, and they also find that those that initially were more diversified subsequently experienced a more rapid income growth. They also find that households with a greater concentration of assets were least able to diversify their income portfolio and were more likely to experience a fall in their relative outcome ranking. Lay *et al.* (2009) compare three national household surveys for Burkina Faso for the period 1994–2003 and find that it is the richer households that can participate in better-paying non-farm activities.

Several studies have identified wealth-differentiated barriers to entry into non-farm activities, for example Reardon *et al.* (1992), Dercon and Krishnan (1996), Dercon (1998), Carter and May (1999), Woldenhanna and Oskam (2001) and Barrett *et al.* (2005), in Burkina Faso, the Ivory Coast, Ethiopia, Kenya, Rwanda, South Africa and Tanzania. Dercon and Krishnan (1996) find that diversification in Ethiopia and Tanzania

cannot be explained by risk behaviour, but is better explained by differences in ability, location and access to credit.[1] Reardon *et al.* (2000) show that diversification is held back by lack of agricultural assets such as land and nonfarm incomes, and this seems to be particularly pronounced in Africa. There seems to be factors specific to Africa that make it hard for the poorest households to break into higher-return activities. Reardon (1997) finds in his survey of the income-diversification literature that households with the highest farm income have the highest level and share of income from non-farm activities. The poor tend to rely more on farm wage labour, while richer households rely more on cash crops, livestock income and non-farm income. Doss *et al.* (2008) find that the poor are more likely to rely on income from their own farms. This suggests that diversification generally is a way up the income scale, but there is also an opposite pattern, distress diversification, where poor households seek to add to their meagre agricultural incomes (Barrett, 1998).

In this study, we are interested in finding out to what extent income diversification in the case of Zambia is of one or the other of these two types. Most households pursue strategies with several income components, and we will try to see whether some routes of diversification are more successful than others.

3. Theoretical framework and approach

In the process of economic development, the agricultural sector's share of output typically shrinks as income levels increase. We will investigate how this pattern plays out within smallholder households in Zambia, that is, analyse the dynamic process of activity switches within these households.

We use a farm household model as the framework for our analysis, and treat the household as one decision-making unit (Singh *et al.*, 1986). The choice situation of rural households in Zambia is well described by Barrett *et al.* (2005): 'Households choose an activity allocation vector for asset endowments that yield an uncertain income return from among a feasible set defined by the intersection of a non-tradable inputs availability constraint equal to one's endowment level of the input (e.g. land) and a budget constraint equal to one's current cash income plus access to liquid capital through savings or credit. Because income is a function of

[1] Ethiopia, with a very undifferentiated countryside, would be a case of distress diversification. There the households that diversify out of agriculture tend to be poorer than the non-diversified (Bigsten *et al.*, 2003).

activity choice, it is an endogenous function of the prevailing (shadow) price distributions for all factors, goods and services. So observed income patterns can be understood as a function of the constraints – including ex ante asset endowments – faced by the household and its preferences.'[2] Both poor and better-off households may also use the choice of activity combinations as a risk management tool.

Since the markets for factors, goods and services in rural Zambia are poorly integrated, households in different locations face different constraints. Thus, where a household is located influences its choice set. It is easier to diversify out of agriculture if a household has access to thriving off-farm activities, which means being located in a region with a diversified economy. This means that the shadow prices of identical factors will vary across households both off- and on-farm. There will also be differences in returns across household due to variations in the quality of the factors, transactions and search costs, risk premiums, locational and sector preferences in employment, etc. So there are differences in the feasible sets from which households choose their activity portfolios. Shortage of resources may make it impossible for many households to enter some lucrative activities, while in other situations it may force them to enter less attractive ones. For example, if they have little land and are poorly educated, they may be forced to sell labour cheaply.

We assume that households do not choose their location. The mobility of whole households is limited in rural Zambia, although there is migration of individual household members. This rules out potential endogeneity due to the possibility that households are choosing their location. There are large spatial variations in transaction costs, local public goods, market prices, etc. 'Geographic capital' may thus alter returns to private investment (Jalan and Ravallion, 2002).

Constraints differ across households also in terms of their access to credit or other forms of liquidity. There are considerable start-up costs involved in some activities; a household sometimes has to enter at a reasonably large scale to be able to enter at all. Barrett *et al.* (2005) note that 'entry into lower-return niches (e.g., petty commerce at weekly rural markets) is low cost and widespread, but movement within the sub-sector in the higher-return niches requiring partially irreversible investment in fixed

[2] For further discussion, see Ellis (1998, 2000) and Barrett and Reardon (2000). In the poverty trap literature, households are assumed to be 'trapped' in a livelihood strategy that pays little (Carter and Barrett, 2006). To be able to make the leap to a better strategy, the households need to accumulate assets (Zimmerman and Carter, 2003), so also in this type of model the allocation choices of households are constrained.

capital is sharply limited by liquidity constraints, social networks necessary to establish, monitor, and enforce contracts, etc.' This means that households that do not possess sufficient human and financial resources do not have access to some potentially lucrative activities.

To be full-time farmers, they need access to land, and the larger the labour force of the household, the more land is required. Consequently, the land/labourer ratio of the household is a key determinant of its desire to move into off-farm activities. The human-capital endowment (education) of household members also determines activity choices as does endowments of physical capital.[3]

So we assume that our households choose a utility-maximising activity vector subject to constraints that may vary considerably among households. Since households are different in many respects, income patterns may vary considerably according to assets and constraints. The main factors behind allocation choices are differences in endowments, differences in access to markets and differences in access to finance.

We suggested in the introduction that there is a link between diversification and income levels or growth. The main mechanism that matters in our context is that incomes increase if households are able to diversify into better remunerated activities. Furthermore, households will be able to achieve a more efficient allocation of its resources, and higher incomes, if the constraints and market distortions that hold them back are reduced. And even with given constraints, they would be able to increase their incomes if they can build up human, financial or other relevant assets that facilitate entry into better paying activities. These are the key growth mechanisms focused in this analysis.

The challenge of establishing causality in the analysis of diversification requires the use of panel data. Although we may not be able to avoid endogeneity or omitted variables bias entirely, we can at least reduce the estimation biases significantly. We have panel data for Zambia for 2 years, 2001 and 2004, which we use to analyse changes over time for individual households. Our empirical approach to the analysis is as follows: how and if endowments, access to markets and finance influence selection into activity combinations will be analysed with a multinomial logit regression. Variables for endowments are land per labourer, maximum level of education in the family, cropping advice and agricultural commodity information. Access to markets is captured by distance to vehicular transport,

[3] Assets are, of course, endogenous variables, and to fully understand the dynamics one would also need to understand the process of factor accumulation.

income level in district and wage level in district. Access to finance is captured by loans in district and wealth. Province dummies are included to capture effects that vary between provinces but have not been captured by other variables. A dummy for the household head being female, and age and age squared are included as well, since it is well known that they might matter. In the selection regression also the size of labour force is included, since it is easier for large households to diversify.

To analyse how and if diversification affects income, two different methods are used. Log income per labourer is regressed on dummies for activity combinations (and controls) for the full sample, both pooled and as a panel. The same exercise is carried out using the observations for 2004, with separate regressions by activity combination in 2001. Finally, we investigate how different factors affect income growth by the activity combination of the household in the initial year.

4. Data and the income variable

The data come from the first and second supplemental surveys to the nationally representative 1999/2000 Post-Harvest Survey (PHS). The PHS is also known as the Agricultural and Pastoral Production Survey. These supplemental surveys, carried out by the Central Statistical Office in conjunction with the Ministry of Agriculture, Food, and Fisheries and commissioned by the Food Security Research Project (FSRP), cover incomes and livelihoods of small and medium scale rural holdings.[4]

The surveys[5] carried out in April/May 2001 and June/July 2004 collected data for the 1999/2000 and 2002/03 cropping seasons and for the 2000/01 and 2003/04 marketing seasons and covered the same sample of roughly 7,000 households as the 1999/2000 PHS. A sampling frame of smallholder farmers (cultivating less than 20 ha) in the rural areas of Zambia was used, and a household has to have at least some land to be included in the PHS.

The FSRP reports that rural poverty has been falling (Jayne *et al.*, 2007). Agricultural growth has been positive, and real staple-food consumer prices have declined by 20% over the past decade. The total gross value of agricultural output rose by over 50% from the mid-1990s to 2001–04. Neither 2001 nor 2004 was exceptional in terms of the conditions for agricultural production (Jayne *et al.*, 2007). We can therefore be reasonably confident that our data sets are representative of the trend in rural incomes.

[4] Policy-makers in Zambia have access to the Crop Forecast and the annual PHS when deciding how to promote small-farmer welfare (Zulu *et al.*, 2007).

[5] Details about the surveys are presented in Republic of Zambia (2001, 2004).

Smallholder income is broadly made up of on-farm (agricultural) income and off-farm income. While the latter is well measured, the former lacks some components on the income side, and also lacks some costs. The ideal income concept includes all current household income (revenues minus costs) plus asset-valuation changes. The latter component is difficult to gather, but for a smallholder household one would like to know stock-valuation changes (changes in the value of livestock assets). Since we do not have this information, we are confined to looking at current income in a year. However, these data also have some shortcomings, which are discussed below.

Our data on farm income have the following four components:

(a) own consumption of crops—the value of crops produced less crops sold;
(b) crop sales—the value of the part of gross production that is sold;
(c) vegetable and fruit sales—the value of vegetables and fruit sold; this income is underestimated to the extent that the household itself consumes vegetables and fruit;
(d) livestock and fish income—this is total income from livestock, i.e., the value of sales of animals (live and slaughtered), production of milk and eggs, plus production of fish; we underestimate this income by ignoring own consumption of meat.

The income from these categories are overestimated to the extent that there were input costs related to the production of crops sold that were not deducted.

Our data on off-farm income also have four components:

(a) own-business income—net income, i.e., gross income less costs; there are no conceptual problems here, but it is difficult for people to remember all costs and revenues for a whole year; to compute annual income, the questionnaire therefore asks for data for a good month and data for a bad month, and then about the numbers of such months;
(b) agricultural wage income—the value of agricultural wage income;
(c) non-agricultural wage income—the value of non-farm labour wage income;
(d) remittances—remittances received by the household from non-household members or organisations; households may also remit out, but we consider that to be part of household expenditures and it is therefore not deducted here.

All income variables are expressed in 2004 Zambian Kwacha (ZMK). See the appendix for further details about the variables and summary statistics.

5. Description of the pattern of income diversification

The question discussed in this paper is how patterns of diversification relate to incomes. We start by presenting our data in some descriptive tables designed to show how income diversification among Zambian smallholders changed from 2001 to 2004. We report estimates for the whole aggregate and by quintile. What is reported in these tables can be compared to some basic figures: In 2004, GDP per capita was 2.29 million ZMK [1,133 purchasing power parity adjusted US$ (PPP-$) in 2005 prices], and the food poverty line was approximately 900,000 ZMK per adult equivalent. The average smallholder per adult equivalent income was thus below the food poverty line (Table 1). Even if incomes may be

Table 1: Overall Income Diversification 2001 and 2004, in 2004 ZMK

	Percent		Per adult equivalent (in thousands)		Total (in billions)	
	2001	2004	2001	2004	2001	2004
Farm income	**51.9**	**56.9**	**233.1**	**302.3**	**1,205.8**	**1,868.2**
Crop production	*42.8*	*45.7*	*192.0*	*242.7*	*992.9*	*1,499.8*
Cereal production	24.6	22.9	110.4	121.6	571.0	751.6
Tubers production	9.4	8.5	42.1	45.2	217.8	279.1
Bean and oilseed produced	5.7	6.1	25.7	32.2	132.9	199.0
High-value crops produced	3.1	8.2	13.8	43.4	71.3	268.1
Other crop production	0.0	0.1	0.0	0.3	0.0	1.8
Vegetable and fruit sales	*5.2*	*5.2*	*23.4*	*27.5*	*121.2*	*169.9*
Livestock and fish income	*4.0*	*6.0*	*17.7*	*32.1*	*91.7*	*198.5*
Animal sales	1.9	4.0	8.7	21.0	45.0	129.7
Egg and milk production	1.9	2.1	8.4	10.9	43.6	67.6
Fish sales	0.1	0.0	0.4	0.1	1.8	0.7
Fish consumption	0.1	0.0	0.3	0.1	1.3	0.5
Farm work	**2.5**	**2.3**	**11.2**	**12.0**	**57.9**	**74.4**
On large-scale farm	*1.8*	*1.7*	*8.1*	*8.8*	*42.1*	*54.7*
On small-scale farm	*0.7*	*0.6*	*3.1*	*3.2*	*15.8*	*19.7*
Non-farm work	**18.6**	**16.2**	**83.4**	**86.3**	**431.4**	**533.5**
Own business income	**25.0**	**23.5**	**112.4**	**125.0**	**581.2**	**772.5**
Remittances	**1.9**	**1.1**	**8.7**	**5.6**	**45.1**	**34.5**
Sum	**100**	**100**	**448.8**	**531.3**	**2,321.4**	**3,283.1**

Note: Weights are used. Italic values are the total of the figures below, and bold values are the total of the italic values.

underestimated as the poverty line may be too high (Bigsten and Tengstam, 2010), this suggests that poverty is widespread among Zambian smallholders.[6]

Although incomes were exceedingly low, all income categories except remittances increased in absolute terms over the studied period. Incomes from high-value crops and animal sales developed favourable starting from low levels, but also incomes from cereal production and own business increased considerably. The percentage coming from farm income increased, while the off-farm percentage decreased. The dependence on subsistence income (not shown) declined slightly.

Table 2 shows how income diversification varied by quintile.[7] In general, the higher the quintile, the lower the farm-income share. Higher quintiles had a higher share of wage incomes (mostly for non-farm labour), and a higher share of own-business income, but a lower share of remittances. We also note that the lower quintiles had strikingly low incomes, but this does not necessarily mean that consumption levels were that low. Income per adult equivalent for the lowest quintile fell by 6.7% over the period covered. Crops harvested for this quintile fell, and also own-business income and remittances fell. Wage incomes increased, which is in line with the notion that the wage-income option is mainly used by the poorest to supplement their income when other sources yield too little. This is distress diversification. The three middle quintiles saw their total income grow by 2–6%. Compared to the overall figures in Table 1, these households had a less favourable development of farm income and own business income. They did not benefit from the positive trend in cereal production and own business. On high-value crops and animal sales, they did benefit, but the growth rate was below the general trend. Finally, in quintile 5,

[6] Even if incomes might be underestimated to some extent, they are mainly in line with the Living Conditions Monitoring Survey IV. The Central Statistical Office (2005, pp. 86 and 91) reports that total smallholder farm income in 2004 was 2,158 billion ZMK, whereas Table 1 shows that total smallholder farm income was 1,868 billion ZMK. Both income levels are expressed in June/July ZMK. The smallholder farm income reported by Central Statistical Office (2005) is reasonable in relation to the total smallholder income, smallholder consumption, household consumption and GDP reported by Central Statistical Office (2005, pp. 86 and 99) and World Bank (2007).

[7] There are the same number of persons in each quintile, so for 2004 the poorest 1,500,000 persons (not adult equivalents) are in quintile 1. 'Poor' means belonging to a household with low income per adult equivalent (not per capita).

Table 2: Income Diversification Shares by Quintile, 2001 and 2004 (Percentage Shares)

Quintile	**1**		**2**		**3**		**4**		**5**	
Year	**2001**	**2004**	**2001**	**2004**	**2001**	**2004**	**2001**	**2004**	**2001**	**2004**
Farm income	**85.5**	**89.6**	**83.0**	**85.5**	**80.0**	**80.7**	**67.5**	**74.2**	**36.6**	**45.0**
Crop production	*78.8*	*81.7*	*75.6*	*78.0*	*70.8*	*69.3*	*57.5*	*62.4*	*27.4*	*33.5*
Cereal production	51.2	47.7	43.3	39.1	38.2	33.8	30.9	30.4	16.7	16.9
Tubers production	20.6	21.3	23.4	22.3	18.6	18.6	13.3	13.7	4.3	3.9
Bean and oilseed produced	6.1	10.0	7.4	11.7	10.8	10.9	8.9	9.3	3.6	3.8
High-value crops produced	0.9	2.7	1.5	4.9	3.2	5.9	4.5	9.0	2.9	8.8
Other crop production	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.1	0.0	0.0
Vegetable and fruit sales	*3.1*	*1.9*	*3.0*	*1.9*	*4.5*	*3.9*	*4.6*	*4.4*	*5.9*	*6.0*
Livestock and fish income	*3.7*	*6.0*	*4.4*	*5.5*	*4.7*	*7.4*	*5.3*	*7.4*	*3.3*	*5.5*
Animal sales	2.2	5.0	2.6	4.5	2.8	5.7	2.6	5.6	1.5	3.2
Egg and milk production	1.4	1.0	1.7	1.1	1.8	1.7	2.6	1.8	1.7	2.3
Fish sales	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.1	0.0
Fish consumption	0.0	0.0	0.1	0.0	0.1	0.0	0.1	0.0	0.0	0.0
Farm work	**1.7**	**2.0**	**1.5**	**1.4**	**2.8**	**2.3**	**2.8**	**3.8**	**2.5**	**1.9**
On large-scale farm	*0.1*	*0.3*	*0.6*	*0.4*	*2.0*	*1.7*	*2.3*	*3.2*	*1.8*	*1.4*
On small-scale farm	*1.6*	*1.7*	*0.8*	*1.0*	*0.8*	*0.6*	*0.5*	*0.6*	*0.7*	*0.5*
Non-farm work	**1.2**	**2.2**	**2.0**	**3.0**	**3.1**	**4.1**	**10.3**	**6.6**	**26.8**	**22.3**
Own business income	**6.8**	**3.8**	**10.0**	**8.3**	**11.2**	**11.2**	**16.8**	**14.2**	**32.8**	**29.9**
Remittances	**4.8**	**2.4**	**3.6**	**1.8**	**3.0**	**1.7**	**2.7**	**1.3**	**1.2**	**0.8**
Sum (%)	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
Sum (in thousands of 2004 ZMK)	**66.0**	**61.6**	**149.1**	**152.4**	**251.9**	**261.6**	**431.4**	**456.9**	**1,341.8**	**1,713.2**

Note: Weights are used. Italic values are the total of the figures below, and bold values are the total of the italic values.

incomes per adult equivalent grew by 27.7%. This is mostly due to increases of farm income and own-business income.

But can average incomes be as low as they are reported? In 2004, total income per adult equivalent for the third quintile was 261,600 ZMK (129 PPP-$ in 2005 prices). Can a household survive at such income levels? Of this, 88,500 ZMK came from cereal production, including 155 kg maize worth 76,000 ZMK, while 48,500 ZMK came from tubers, including 152 kg of cassava worth 44,000 ZMK. Since the energy content of 1 kg of maize is 3,800 kcal, and the energy in 1 kg cassava is 1,600 kcal, this equals 2,280 kcal/adult equivalent per day. So although the maize and cassava produced per adult equivalent by a typical household only was worth 120,000 ZMK (59.4 purchasing power parity adjusted US$ in 2005 prices), maize and cassava by themselves provided almost as much energy as is needed by an adult equivalent (2464 kcal per day as per guidelines from the World Health Organization, 1985).

Bigsten and Tengstam (2010) study how consumption developed in rural Zambia during 1998–2004 and find that moderate rural consumption poverty fell from 83.5 to 77.5% (the moderate poverty line includes food, health, shelter and education). They decompose the change into one part due to per capita consumption growth and one part due to consumption inequality change. The growth is found to have contributed 6.5 percentage points (pp) of the fall, while inequality change reversed the fall by 0.2 pp. (There was also a 0.3 pp residual.) The data presented in Table 3 show that income inequality rose among rural smallholders during 2001–04. These two findings are not contradictory. They partly cover different years, and income inequality within the rural smallholder group can rise at the same time as consumption inequality in the full rural group is almost unchanged. Given that smallholders in general are relatively poor in rural Zambia, this group getting on average higher incomes should contribute to lower rural inequality.

To be able to identify livelihood strategies, we classify households according to their sources of income. To simplify, we do not take remittances into account (this is only 1–2% of total income). This leaves us with fifteen potential activity combinations, and those for which no income at all was registered. A household that derived farm income but had no other income has the activity combination farming-only, denoted F. A household that derived income only from farming and agricultural wage work has the activity combination FA, and so on. Tables 3 and 4 present the activity combinations for 2001 and 2004, respectively.

Table 3: Income by Activity Combination, 2001 (in 2004 ZMK per Adult Equivalent, in Thousands)

	Farm income	Farm work	Non-farm work	Own business	Total income	Frequency, %
F	253.2	0.0	0.0	0.0	253.2	50.9
FA	216.0	173.1	0.0	0.0	389.1	4.8
FN	210.0	0.0	624.9	0.0	834.9	9.2
FB	226.4	0.0	0.0	341.9	568.3	25.5
FAB	171.2	79.8	0.0	136.3	387.3	2.3
FNB	221.7	0.0	424.3	344.4	990.4	5.1
A	0.0	720.5	0.0	0.0	720.5	0.0
B	0.0	0.0	0.0	777.8	777.8	0.5
N	0.0	0.0	923.1	0.0	923.1	0.2
AB	0.0	19.7	0.0	206.9	226.6	0.0
AN	0.0	546.1	509.3	0.0	1,055.4	0.0
NB	0.0	0.0	169.5	116.7	286.1	0.1
FAN	155.0	109.4	185.6	0.0	450.0	0.4
FANB	216.8	100.7	245.4	121.7	684.6	0.4
None	0	0	0	0	0	0.5

Notes: F, farm income; A, agricultural wage work; N, non-agricultural work; B, own-business income. Activity frequency is based on population and not households. Weights are used.

The overall pattern changed little between the two years. Households engaged in non-agricultural wage work or own-business generally have higher incomes than others. Comparing Tables 3 and 4, we see that the full-time farmer share increased. Generally, farmers were less diversified in 2004 than in 2001.

Table 5 shows transitions from one type of combination in 2001 to another in 2004. We use the panel data set, i.e., the households for which we have observations for both years. The entries in the table show where those who started in a certain activity combination in 2001 ended up in 2004. For example, the first row contains the households that were in activity combination F in 2001. Of those, 67.2% still were in F in 2004, while 4.4% were in FA and 6.0% in FN. We see that 16.9% of the households that were in the activity combination F in 2001 had diversified into FB by 2004.

We note that 70.9% (Table 5, row 2, columns 1, 3, 4 and 6) of those earning income from a combination of their own farm and agricultural wage work in 2001 did not receive any agricultural wage income in 2004.

Table 4: Income by Activity Combination, 2004 (ZMK per Adult Equivalent, in Thousands)

	Farm income	Farm work	Non-farm work	Own business	Total income	Frequency, %
F	309.9	0.0	0.0	0.0	309.9	52.9
FA	201.0	160.2	0.0	0.0	361.2	5.3
FN	308.1	0.0	614.7	0.0	922.7	10.1
FB	339.3	0.0	0.0	445.9	785.2	22.0
FAB	221.4	94.3	0.0	177.5	493.2	2.4
FNB	298.4	0.0	423.8	389.3	1,111.5	4.8
A	0.0	119.5	0.0	0.0	119.5	0.2
B	0.0	0.0	0.0	905.8	905.8	0.4
N	0.0	0.0	615.6	0.0	615.6	0.2
AB	0.0	15.0	0.0	11.7	26.7	0.0
AN	—	—	—	—	—	0
NB	0.0	0.0	997.0	468.9	1,465.9	0.1
FAN	192.1	90.2	90.1	0.0	372.4	0.6
FANB	195.9	71.9	162.9	135.9	566.6	0.5
None	0	0	0	0	0	0.5

Notes: F, farm income; A, agricultural wage work; N, non-agricultural work; B, own-business income. Activity frequency is based on population and not households. Weights are used.

Table 5: Percentage Moving from One Activity Combination to Another from 2001 to 2004

Activity 2001	Activity 2004							Sum	Frequency 2001
	F	FA	FN	FB	FAB	FNB	Rest		
F	**67.2**	4.4	6.0	16.9	1.9	2.3	1.3	100.0	50.9
FA	48.3	**23.4**	6.7	12.8	3.9	3.1	1.8	100.0	4.8
FN	24.6	4.5	**41.4**	9.2	0.9	14.9	4.5	100.0	9.2
FB	45.1	3.5	5.4	**36.5**	3.2	4.1	2.2	100.0	25.5
FAB	49.4	6.7	5.3	20.9	**9.3**	3.4	5.0	100.0	2.3
FNB	24.3	4.5	24.4	22.9	3.8	**16.0**	4.1	100.0	5.1
Rest	43.6	0.7	10.7	20.5	2.9	9.9	**11.6**	100.0	2.2

Notes: Rest includes all other activities, as well as having no activity at all. Weights are used. F, farm income; A, agricultural wage work; N, non-agricultural work; B, own-business income.

Thus, working on others' farms is not generally a permanent feature of smallholder income generation in Zambia. Most of the households that were full-time farmers in 2001 had the same activity combination in 2004.

Looking at the values on the diagonal in Table 5, we see that mobility is high. For example, only 41.4% of the households that started in FN had the same activity combination in 2004. We see that 56.3% of those that started in FN remained in FN or had diversified further into FNB by 2004, and that 40.4% of those that were in FNB in 2001 were still in FNB or FN in 2004. Non-agricultural wage work is apparently more of a long-term choice than agricultural wage work is. We further note that of those that started in FB, 43.8% stayed in FB or diversified further to FAB or FNB, while 45.1% fell back to F.

These descriptive tables show that there is extensive income diversification among Zambian smallholders, and that increasing diversification generally is associated with higher incomes. However, to be able to say something more substantive about associations we need econometric analysis.

6. Explaining income diversification of smallholders

In the econometric analysis, we look at several related aspects of smallholder income diversification and incomes. Before we do the analysis, let us present the variables used in the regressions.

6.1 Explanatory variables

From the theoretical review, we concluded that important determinants of household diversification and income are endowments, market access and access to finance. First, we have data reflecting the assets of households. We include a variable measuring the highest education in years among the labour force in the household. We include a measure of land per labourer; i.e., hectares of land (cultivated and fallow) per household member aged 23–59. We choose age 23 as the lower cut-off point to exclude the vast majority of students from the labour force.[8] The data set allows us to also study the effect of advice on cropping and information on agricultural commodity prices. To pick up possible effects of indivisibility, we include a variable for the size of the household labour force. We also include the age of the household head (Age) and its square (Age sq) to pick up potential life-cycle effects, plus a dummy for female headship.

We include market access in three ways. Income level in district captures local demand. Wage level in district captures the level of development of

[8] Using 16 years as a lower cut-off point instead of 23 years gives very similar results.

the labour market. Distance to vehicular transport reflects how easy it is to reach outside markets.

It is difficult to find good measures of access to finance. We include two variables as proxies. The first variable to capture access to finance is a measure of the households' wealth. The second is loans in district, which measures the share of households in the district that have obtained an agricultural loan not restricted to agriculture. This proxies for credit available for diversification out of agriculture.

Province dummies are included to capture endowments, access to markets or access to finance influences that vary between the provinces but are not captured by the other variables. For example, access to credit may, to some extent, be picked up by the provincial dummies, reflecting different levels of economic integration, including development of the financial system. We include provincial dummies for eight provinces: Lusaka, Central, Copper Belt, Southern Luapula, Northern, North Western and Western; Eastern is the default. For a detailed presentation of the variables, see Appendix A.

In the econometric analysis undertaken to explain income diversification of smallholders, we use the panel data set, i.e., the observations that existed in both years. There were 6,922 households in the 2001 survey, of which 1,580 were not in the 2004 survey. Out of the latter, 707 households had moved out of the standard enumeration area, 390 had dissolved and 362 had not been possible to contact (Republic of Zambia, 2004). Thus, the attrition rate was 22.8%. Households that left the survey had on average 95% of the income level of the total 2001 sample; 26.7% were female-headed compared to 21.7% for the whole sample; and they were on average 1.8 years younger, yet with 0.14 more years of schooling. Overall, they were not very different from the households that remained in the sample. However, since only the 'survivors' stay in the data set, and no new household are included, the household members become older on average.

6.2 Selection of activity combinations

The first key issue is to understand the determinants of choice of activity combination. We specify a multinomial logit model predicting the likelihood of each of the seven choices a household can make.[9] A household

[9] For comparison, we ran a standard logit and a standard probit for one activity combination at the time (F, FA, FN, etc.). They both gave estimates of the marginal effects that were very similar to those from the multinomial logit.

can choose one of the six activity combinations, F, FA, FN, FB, FAB and FNB, on which we focus. Or it can choose Rest, which includes all other activity combinations. The multinomial logit model (Green, 2003) for activity combination choice is:

$$\text{Prob(Act comb}_{it} = j) = \frac{e^{\beta_{0j}+\beta_{Xj}X_{it}+\beta_{Zj}Z_{it}}}{1+\sum_{k=1}^{6} e^{\beta_{0k}+\beta_{Xk}X_{it}+\beta_{Zk}Z_{it}}}, \quad (1)$$
$$j = 0, 1 \ldots, 6, \ \beta_0 = 0,$$

where X_{it} is a vector of variables capturing endowments (Land per labourer, Education, Cropping advice in district, Agri-commodity info in district and Labour force), access to markets (Distance to vehicular transport, Log income in district and Wage level in district) and access to finance (Loans in district). Z_{it} is a vector of control variables (age, age squared, gender of household head, dummies for province, a dummy for year 2004).

It is a concern that there might be reversed causality for Land per labourer if, e.g., households buy more land if they are full-time farmers. We therefore tested an IV approach using Average land per labourer within the district as an instrument. It is strong and valid. There are no statistical package that supports IV multinomial logit, so we estimated IV probits for one activity combination at the time. That this is a reasonable solution is supported by the fact that probits and logits without instruments gave results that were very similar to those of the multinomial logit without instruments. Since the Wald test indicates that there is only endogeneity in selection into two activity combinations, FB and FAB, and the instrumented parameter estimates differ only marginally from the non-instrumented, we do not use any instrumental approach.

Size of labour force and wage level in district are used in the selection regressions, but not in the income regressions. Size of labour force arguably matters for selection of activity combination, since individuals cannot be in two different places at the same time. However, there are less clear reasons for this variable to influence income level *per se*. Wage level (for non-skilled labour) in district is included to capture the level of development of the (non-skilled) labour market, and this matters for choice of activity combination. It cannot be included in the analysis of income, since wage level has a very strong effect of income levels, in itself. On the other hand, the wealth index is used in the income regressions, but not in the selection regression. Due to data limitations, the wealth index includes mostly agricultural tools (plough, harrow, water pump and

bicycle). It cannot be included in the selection regressions since it is strongly endogenous; a household that focus on faming uses agricultural tools. The wealth index is used as a crude proxy for access to finance in the income regressions, since this problem is much smaller in the income regressions.

The marginal effects reported in Table 6 give strong support for the hypothesis that endowments and access to markets matter for diversification. Endowments have in general the strongest effect. Having little land makes it more likely to be in FN or FA. One standard deviation less of land/labourer makes it 2.9 and 2.6 pp, respectively, more likely that a household is in FN and FA, respectively (see appendix for summary statistics). This can be assumed to be distress diversification, since A is mostly non-skilled work, while N contains both skilled and non-skilled work. Education clearly opens up for non-agriculture wage work. One standard deviation more of education makes it 6.5 pp more likely to be in FN. Being in a district where a large share of the households have received advice on cropping makes it more likely for households to be full-time farmers rather than diversified into business. One standard deviation more of cropping advice in the district makes it 4.1 pp less likely that a household is diversified into FB, and 7.0 pp more likely that it is a full-time farmer. One standard deviation more of agri-commodity price information in the district makes it 2.2 pp more likely that a household is diversified into FB.[10] Households with a large labour force seem to be more diversified than smaller households, but the effects are not strong.

Access to markets measured as being closer to transport facilities increases the likelihood of non-agriculture wage work. Being in a district with high-income levels primarily makes it more likely for households to diversify into activity combinations including business. The effect in pp/standard deviation is 1.7 and 0.5 for FB and FAB, respectively. Finally, we look at access to finance. The results are not clear-cut, but credit available for diversification out of agriculture seems to facilitate diversification into activity combinations including own business. Due to data limitations, the measure of access to financial markets is a rough proxy, as discussed in Section 6.1

[10] Agricultural trading, retailer/ship-owner and marketer/hawker/vendor makes up more than 40% of the persons involved in business, and all these categories should benefit from the information.

Table 6: Marginal Effects for Selection into Activity Combinations, Multinomial Logit

	F	FA	FN	FB	FAB	FNB
Land per labourer	0.021**	−0.017**	−0.019**	0.018**	−0.0011	0.0046**
Education	−0.022**	−0.0036**	0.018**	0.0024*	−0.00098*	0.0069**
Cropping advice in district[a]	0.57**	−0.017	−0.083*	−0.33**	−0.028	−0.083**
Agri commodity info in district[a]	−0.27**	0.090**	0.041	0.15**	0.014	0.0024
Labour force	−0.0062	−0.00025	−0.0091**	0.0061	0.0021	0.0061**
Distance to vehicular transport	0.0018**	0.000071	−0.0011**	0.00024	−0.00021	−0.00067**
Log income in district[a]	−0.14**	0.014	0.040**	0.051**	0.016*	0.017
Wage level in district[b]	−0.00043	0.00080*	−0.00045	−0.00017	0.00067**	−0.00079*
Non ag loans in district[a]	−0.24	−0.42**	−0.079	0.30	0.12	0.23*
Age	−0.00085	0.00044	0.0046**	−0.0063**	0.0014	0.0011
Age sq	0.000052*	−5.7e − 06	−0.000047**	0.000034	−0.000018*	−0.000017
Female head	0.047**	−0.0049	−0.046**	0.0082	0.011*	−0.016**
Lusaka	−0.32**	0.065*	−0.022	0.11**	0.057*	0.052
Copperbelt	0.0071	0.026	−0.039**	0.014	0.0080	−0.0079
Central	−0.037	0.024	−0.036**	0.065**	0.0040	−0.019*
Southern	−0.025	0.044**	−0.035**	0.018	0.016	−0.023**
Luapula	−0.13**	−0.018*	−0.055**	0.11**	0.052*	0.0096
Northern	0.012	−0.037**	−0.044**	0.060**	−0.0013	−0.0057
Nwestern	0.075**	−0.020*	−0.025*	0.0022	−0.021**	−0.017
Western	−0.26**	0.034*	−0.025*	0.17**	0.045*	0.033
t4	0.029**	0.0056	0.0056	−0.036**	0.0025	−0.0052
Observations	9,481	9,481	9,481	9,481	9,481	9,481

Notes: F, farm income; A, agricultural wage work; N, non-agricultural work; B, own-business income. Robust standard errors are used. The reported effects are the average of the marginal effects for the sample. Weights are used. Eastern is the default province.

[a]The average in the district, excluding the observation itself.

[b]The average in the district, including the observation itself.

*Significant at 5%.

**Significant at 1%.

We note that having a female household head makes it considerably less likely for a household to diversify into non-agricultural wage work and somewhat more likely to diversify into business.[11] This result may be due to the fact that females are less geographically mobile (because of traditional household or family duties) than males, while non-farm wage work often requires long-distance travel.

The endowments, markets and finance variables are able to explain quite a lot of the selection into activities. The province dummies do add considerably less to the goodness-of-fit of the model than the endowments, markets and finance variables do. When using all variables, the prediction accuracy rate is 48.8% and the pseudo-R^2 is 0.0932. When running the regression without province dummies, the prediction accuracy rate is slightly higher, 50.6%, and the pseudo-R^2 is a bit lower, 0.0695. Running the regression without the endowments, markets and finance variables, the prediction accuracy rate is 48.5%, while the pseudo-R^2 is much lower, 0.0333.[12]

Province dummies are used as control variables, and the explanatory power of our model does not rely on the province dummies to any considerable extent. Still one can, as a robustness check, ask if the estimated parameter values are reasonable. For example, households in Lusaka tend to be in FB much more, and in F much less, than what our other variables predict. In Lusaka, a much larger share of the population is involved in two of the business categories: marketer/hawker/vendor (8 pp more than average) and 'other' (12 pp more than average). This reflects the fact that Lusaka has a more developed business climate that is hard to capture with our control variables.

6.3 Determination of level of income

Next we look at the determinants of the level of real income. We estimate six equations describing the determinants of log income per labourer.

[11] The age variable and the dummy for female head and province are mostly control variables. We would like to control for any effect they might have. We see that age does not seem to matter much. Having a very young or very old household head make it less likely to be in FN.

[12] Prediction accuracy rates do not change much for different sets of explanatory variables, since the model, irrespective of which explanatory variables that are included, predicts the activity combination F for almost every observation. This is expected since the frequency for F is more than double the frequency of the second most common activity combination. Studying pseudo-R^2 is therefore more informative in cases like these.

For the first straightforward OLS, we specify

$$\text{Log}\,\frac{Y_{it}}{L_{it}} = \beta_0 + \beta_A A_{it} + \beta_X X_{it} + \beta_Z Z_{it} + \varepsilon_{it}, \tag{2}$$

where Y_{it} is the income and L_{it} is the size of labour force. A_{it} is a vector of dummy variables indicating if household i at time t are doing that particular activity combination (FA, FN, FB, FAB, FNB and Rest, with F as the default). X_{it} is a vector of variables capturing endowments (Land per labourer, Education, Cropping advice and Agri-commodity info), access to markets (Distance to vehicular transport and Log income in district) and access to finance (Loans in district and Wealth). Z_{it} is a vector of control variables (age, age squared, gender of household head, dummies for province, a dummy for year 2004 and a dummy for having a new household head since the period before).

We need to normalise for household size, and the natural way to do this if we want to explain income levels is to do it per labourer. Hausman tests suggest that a panel estimation is preferred to a pooled estimation, and that a fixed effects panel estimation is preferred to a random effects panel estimation. A time dummy is added. As a comparison, we also run a pooled OLS and an RE panel. Since the effect of the time invariant explanatory variables cannot be estimated in a FE specification, we use the RE panel estimation for this purpose. Although the RE estimates are somewhat biased according to the Hausman test, this is the only way to get at least some idea of what the effects of the time invariant explanatory variables are.[13]

A possible concern is that there are endogeneity in the estimations. One risk is that the activity combinations might be correlated with unobserved ability. We would like to use instruments for activity combinations. The obvious IV candidates are the selection variables in the estimations in Table 7. If they are uncorrelated with unobserved ability, they are valid instruments. Unfortunately, they are too few to be instruments for activity combinations.[14] We have not been able to find any other instruments. But

[13] We have tested the effect of a dependency ratio (people aged 0–14 and 65 and older divided by the labour force aged 15–64). It turns out that the effect on output per labourer is positive. This reflects the fact that the dependents after all do contribute something to output. When we ran the regression on income per adult equivalent instead, we found a strongly negative effect. This shows that dependents indeed add something to output, but much less than proportionately.

[14] One alternative could be to use the predicted probabilities from the multinomial logit model as instruments. They are not too few. Since the predicted probabilities are not

Table 7: Pooled OLS Regressions (Columns 1–2) and Panel Regressions (Columns 3–6) for Log Level of Income, 2001 and 2004

	Pool1	Pool2	RE1	RE2	FE1	FE2
FA	0.32**		0.31**		0.34**	
FN	0.94**		0.85**		0.59**	
FB	0.64**		0.62**		0.65**	
FAB	0.57**		0.58**		0.62**	
FNB	1.11**		1.06**		0.96**	
Land per labourer	0.21**	0.21**	0.19**	0.19**	0.15**	0.16**
Education	0.024**	0.049**	0.022**	0.044**	−0.0098	−0.0023
Cropping advice	0.11**	0.14**	0.11**	0.13**		
Agri commodity info	0.18**	0.21**	0.19**	0.22**		
Distance to vehicular transport	0.000031	−0.0011	5.9e − 06	−0.0012		
Log income in district[a]	0.46**	0.55**	0.50**	0.57**	0.62**	0.70**
Non ag loans in district[a]	−0.93	−0.81	−0.69	−0.58		
Wealth	0.22**	0.20**	0.23**	0.21**		
Age	−0.0097*	−0.0070	−0.0083	−0.0062	0.023	0.034
Age sq	0.000056	3.6e – 06	0.000048	−2.6e – 07	−0.00026	−0.00039*
Female head	−0.15**	−0.20**	−0.16**	−0.22**	−0.098	−0.18

Lusaka	−0.0072	0.23*	−0.14	0.079		
Copperbelt	0.041	0.074	0.025	0.088		
Central	−0.13**	−0.080	−0.15**	−0.100*		
Southern	−0.25**	−0.20**	−0.26**	−0.20**		
Luapula	−0.13**	0.019	−0.15**	0.0021		
Northern	0.021	0.081*	0.015	0.083*		
N. Western	0.0063	0.073	−0.024	0.037		
Western	−0.19**	0.048	−0.21**	0.028		
t4	0.089**	0.067**	0.039*	0.025	0.084**	0.066**
Newhead	−0.26**	−0.27**	−0.27**	−0.27**	−0.28**	−0.26**
Constant	6.80**	5.74**	6.42**	5.47**	4.17**	3.29**
Observations	9,450	9,450	9,450	9,450	9,450	9,450
R^2	0.339	0.236			0.190	0.081
Number of households			4,956	4,956	4,956	4,956

Notes: Weights are used, except in RE where Stata cannot do it. Robust standard errors are used. F is default activity combination. A dummy for remaining activity combinations was included but the result is not reported here. Eastern is the default province.

[a]The average in the district, excluding the observation itself.

*Significant at 5%.

**Significant at 1%.

as long as unobserved ability does not change over time, this problem is solved by the FE estimations. And as mentioned before, a Hausman test indicates that the FE estimations should be used.

Households that changed gender of the household head seem to have had a negative development of their incomes. But since households often change to female headship because they have had a death in the family, it might be this that drives the results. To be able to separate these two effects, we want to control for getting a new head. We therefore include a dummy for having changed household head since last period.[15]

Our main interest is the activity variables. Does it pay to diversify? The results are reported in Table 7. In regressions 2, 4 and 6, A_{it} is not included. In regressions 3 and 4, a random effects specification is estimated. In regressions 5 and 6, a fixed effects specification is estimated. The OLS1, RE1 and FE1 regressions give very similar results. We see that FA is associated with about 40% higher income than F.[16] Further we see that diversifying into FB or FAB gives about 80% higher income. What seems to give the highest income is diversification into FN or FNB. FN gives about 145% above farming only (80% in the FE), and the most diversified households. FNB has 200% (160% in the FE) higher income than specialised farmers.

The only notable difference between the three estimation methods is that the FN (and to a smaller extent the FNB) parameter has a considerably lower estimates in the FE estimation than in the other two. This indicates that the households that do non-agriculture wage work earn more—not only because they do this type of work but also because they are different (more different than our control variables capture). Among Zambian smallholders, education seems to be correlated with non-agriculture wage work.

linear combinations of the selection variables, it is mathematically possible to do it, but we have severe multi-collinearity. This is shown in the under-identification test, Kleibergen–Paap, which has a *P*-value of 0.26. So this is not doable.

[15] Unfortunately, we do not have any information on whether a household of 2001 had changed head since the period before. We put it to zero for 2001 to be able to do the regressions and take advantage of the information we have on 2004. Compared to the estimates without adding new head, the only notable difference is that the parameter of the female dummy is a bit less negative, as expected.

[16] Since the left-hand variable is the logarithm of income and the estimate of the FA dummy is, e.g., 0.31 in the RE1 estimation, $e^{0.31} - 1 = 0.36$ gives the effect on income level in the RE1 estimation.

The FE1 estimation, and the RE1 estimation for the time invariant variables, suggests that endowments and access to markets, as well as access to finance, have a strong positive direct effect on income per labourer. All variables except distance to vehicular transport and loans in district have statistically and economically significant effects. It can also be observed that education does not seem to have any positive effect in the FE estimation. This probably is partly because the effect was overestimated in the RE due to unobserved ability being positively correlated with education. But the result in the FE should be handled with care, since the reported values indicates that there are considerable measurement errors.[17] Even though there is no obvious reason why these errors should be systematic, they might have an substantial influence on the FE estimates, since the true variance in the difference of education, from education changing over time, is very small. The true variance in education levels, used in the pooled regression and in the RE regression, is much higher. This means that the variation of the education variable coming from the measurement error is not dominant in OLS and RE.

Looking at the RE2 estimation, in which the activity combinations are excluded as explanatory variables, we get estimates of the total effect, that is, both the direct effect on income and the indirect effect via activity combination choice.

Even though the magnitudes of the coefficients of the province dummies are not negligible, the effect of endowments, access to markets and access to finance are more important. Looking at the pooled regression including activity combinations, we see the following: when all variables are used, the R^2 is 0.3388. When provinces variables are dropped, the R^2 is 0.3320. When keeping only provinces, activity combinations and time dummy, the R^2 is 0.1786. In other words, the marginal contribution of the province dummies to R^2 is 0.0068, while the marginal contribution of the variables for endowments, access to markets and access to finance to R^2 is 0.1602. Both groups of variables, the province dummies, on the one hand, and the rest, on the other hand, pass an F-test to be included.

6.4 Income and changes of activity combinations

In this section, we split the sample into subsamples, based on activity combination in 2001. We estimate six equations describing the determinants of

[17] The education variable changes very much from year 2001 to 2004 for quite a few of the households. This is probably due to some of the respondents giving imprecise answers. It could also be a mistake from the enumerator. It is, in other words, a measurement error.

log income per labourer of 2004, one for each activity combination the households were doing in 2001. The specification is:

$$\text{Log}\frac{Y_{i2004}}{L_{i2004}} = \beta_0 + \beta_A A_{i2004} + \beta_X X_{i2004} + \beta_Z Z_{i2004} + \varepsilon_{i2004}. \qquad (3)$$

The specification is similar to equation (2), but it differs in the following three respects: it is about log income per labourer in 2004. Lagged log income per labourer, $\text{Log}(Y_{i2001}/L_{i2001})$, is included in the control variables Z_{it}. In A_{i2004}, the vector of dummy variables indicating which activity combination was in 2004, the default activity combination differs among the six regressions. The activity combination in 2001 is used as default. In other words, now A_{i2004} captures change of activity combination, and the default is that the household does not change.

Within each subsample, we look at how shifting from one activity combination to another influences income. By controlling for initial income, we should at least significantly reduce reverse causality. The six regressions, one for each initial activity combination, are shown in Table 8. The first regression includes households that had the activity combination F in 2001, and the second includes those that had the activity combination FA in 2001, etc. The dummy FN04 captures households that had changed to FN by 2004, etc. Having the same activity combination both years is the default. In both regressions, we control for initial income, i.e., log income per labourer in 2001. We use the same controls as in the previous section.

Let us first consider the households that started in F. This is the intuitively most appealing regression. The effect of changing to any other activity combination was large and statistically significant. Changing to FA gave an approximately 25%[18] higher income, and changing to FN, FB or FAB resulted in a 70–80% increase. Yet, changing to FNB seems to be the most lucrative alternative, since it gave a 100% higher income compared to staying in F.

Next, we look at those that started in FB. For these households, it was statistically significantly negative to revert back to F, since it lowered income levels by 53%. It was also negative to change to FA, since this change was associated with a decrease of income by 41%. The other dummy estimates are not statistically significant.

Looking at the other four regressions, one should keep in mind that these subsamples are quite small, and these estimations therefore have

[18] $e^{0.22} - 1 = 0.25$.

Table 8: Log Income per Labourer 2004 (Separate Regressions by Activity Combination 2001)

Variables	2001					
	F	FA	FN	FB	FAB	FNB
F		−0.92**	−0.92**	−0.76**	−1.32**	−1.65**
FA	0.22**		−0.50	−0.52**	−1.02*	−0.84*
FN	0.57**	−0.052		−0.21	0.40	−0.47*
FB	0.59**	0.12	−0.011		−0.25	−0.99**
FAB	0.50**	−0.091	−0.56	−0.077		−0.32
FNB	0.70**	−0.37	0.090	0.16	0.66	
Land per labourer	0.27**	0.33**	0.32**	0.26**	0.53**	0.45**
Education	−0.00095	−0.066**	0.052**	−0.0030	0.0068	0.022
Cropping advice	0.093	0.42	0.19	0.31**	−0.17	0.28
Agri commodity info	0.24**	−0.029	0.15	0.16	0.44	0.11
Distance to vehicular transport	0.0015	0.0027	0.0021	−0.0022	−0.0049	0.00064
Log income in district[a]	0.40**	0.63	0.24	0.17	0.95*	0.31
Non ag loans in district[a]	1.03	0.0064	1.09	−0.20	−15.6*	−2.83
Wealth	0.15**	0.19*	0.00012	0.20**	0.041	0.093
Age	−0.0061	0.040	0.050	−0.0065	0.016	−0.054
Age sq	0.000017	−0.00049	−0.00065*	−0.000014	−0.00033	0.00048
Female head	−0.16**	−0.048	−0.43	−0.069	−0.062	−0.22
Lusaka	−0.086	−0.35	0.39	−0.35	−1.15*	0.14
Copperbelt	−0.032	−0.012	−0.27	0.099	−0.91*	−0.12
Central	−0.23**	−0.26	−0.25	−0.35**	−1.57**	−0.11
Southern	−0.36**	−0.091	−0.22	−0.48**	0.48	−0.48
Luapula	−0.10	−0.17	−0.29	−0.090	−1.07**	−0.020
Northern	−0.040	0.37	−0.038	−0.021	−0.71	0.020

(*continued on next page*)

Table 8: *Continued*

Variables	2001					
	F	FA	FN	FB	FAB	FNB
N. Western	0.078	0.38	−0.15	−0.14	0	0.31
Western	−0.27**	0.24	−0.26	−0.25*	−0.87*	−0.068
Lag Log income per labourer	0.25**	0.28**	0.44**	0.21**	0.35**	0.39**
Newhead	−0.15	−0.73*	−0.051	−0.27*	0.30	−0.084
Constant	4.44**	0.81	3.17	8.70**	−2.64	5.82
Observations	2,338	206	367	1,230	106	210
R^2	0.374	0.479	0.613	0.372	0.626	0.625

Notes: Weights and robust standard errors are used. The default activity combination in respective regression is the activity combination used for 2001. A dummy for remaining activity combinations was included but the result is not reported here. Eastern is the default province. The independent variables are for 2004.

[a]The average in the district, excluding the observation itself.

*Significant at 5%.

**Significant at 1%.

less power. Nevertheless, the overall picture is that it generally pays to switch from a less diversified to a more diversified activity state. And furthermore there is a negative impact on income of becoming less diversified. Starting in any activity combination other than F, changing to F led to a lower income in 2004.

In sum, the magnitudes of the estimates are in line with what were found in Table 7.[19] We did the estimations in Table 8 to a large extent to try to handle the possible endogeneity problems in Table 7. Luckily, the results in Table 7 do not seem to have been driven by endogeneity.

Even though the magnitudes of the coefficients of the province dummies are not negligible, the effect of endowments, access to markets and access to finance are more important. As in the regressions in Table 7, the marginal contribution of the province dummies to R^2 is much less than the marginal contribution of the variables for endowments, access to markets and access to finance to R^2.

6.5 Income growth within an activity combination

So far we have analysed the impact of activity combinations and changes in activity combinations on income per labourer. What remains to investigate are the deeper determinants of income growth within activity combinations. We thus estimate annual income growth per labourer for 2001–04 by activity combination in 2001. Once again, the sample is split into six subsamples, depending on which activity combination was in 2001. A test for sample selection bias based on a Heckman selection model and the selection variables that were used in Table 6 shows no evidence of sample selection bias.[20] We estimate six equations describing the determinants of annual growth of income per labourer for the period 2001–04.

[19] The estimated coefficient for loans in district for the subsample starting from FAB is huge—something fishy is going on here. Since there is only 106 households in this subsample, a few outliners can be driving this.

[20] Since there is no clear-cut test for selection bias based on a multinomial logit model, we use a Heckman selection model instead. This means taking one activity combination at a time. But these approaches should give similar results, and as mentioned earlier, we have compared our multinomial estimates to probit estimates, and they are very similar. We find that Mills lambda is statistically insignificant in all Heckman regressions. We conclude that there is no evidence of sample selection bias. There are not any clear theoretical reasons to expect sample selection bias, either.

The specification is:

$$\frac{\text{Log}(Y_{i2004}/L_{i2004}) - \text{Log}(Y_{i2001}/L_{i2001})}{3} = \beta_0 + \beta_X X_{i2001} + \beta_Z Z_{i2001} + \varepsilon_i, \quad (4)$$

where Y_{it} is the income and L_{it} is the size of labour force. X_{it} is a vector of variables capturing endowments (Land per labourer, Education, Cropping advice and Agri-commodity info), access to markets (Distance to vehicular transport and Log income in district) and access to finance (Loans in district and Wealth). Z_{it} is a vector of control variables [age, age squared, gender of household head, dummies for province and log income per labourer, $\text{Log}(Y_{i2001}/L_{i2001})\text{Log}(Y_{i2001}/L_{i2001})$].

The results of the OLS regression (Table 9) show that endowments and access to finance have a direct effect on income growth, and it shows that these effects differ across activity combinations. Endowments in the form of education are positive for income growth if the household has FN. This is expected, since non-agricultural wage work, to a large extent, is skilled. Cropping advice is as expected positive for those in F as well as for those in FB. Also information on commodity prices is positive if the household starts in F or FB. The positive effect for those in F may indicate that the information facilitated diversification from F to FB from 2001 to 2004.

For access to markets, few statistically significant effects are found. Access to finance represented by the wealth index has a positive effect on income growth if a household starts in F, FB or FNB. Since this index, due to data limitations, includes mostly agricultural tools (plough, harrow, water pump and bicycle), it is a rough measure.

Finally, part of the income growth is left to be explained by provinces, and some of the province dummies are quite large. Endowments, access to markets and access to finance make a better job at explaining income per labourer *levels* in earlier sections, than *growth* in this section. It is reasonable that growth over a 3-year period, to a larger extent, is driven by temporary shocks and fluctuations, such as weather. These shocks are often concentrated to some provinces. But, as in the regressions in Tables 7 and 8, the marginal contribution of the province dummies to R^2 is much less than the marginal contribution of the variables for endowments, access to markets and access to finance to R^2.

Table 9: Income per Labourer Growth (Annual) 2001–04 (Separate Regressions by Activity Combination 2001)

	F	FA	FN	FB	FAB	FNB
Land per labourer	0.011	−0.036	−0.014	0.022**	−0.015	0.014
Education	0.0043	−0.022*	0.027**	0.0017	0.0092	0.015
Cropping advice	0.062**	0.15	0.10	0.11**	0.20	0.052
Agri commodity info	0.11**	0.0063	0.053	0.081*	0.27*	0.081
Distance to vehicular transport	0.00021	−0.0012	−0.0013	−0.0013*	−0.0054	−0.00048
Log income in district[a]	0.011	0.13	−0.065	−0.023	0.35*	0.084
Non ag loans in district[a]	0.34	2.28	−0.18	−0.44	−3.50	−0.37
Wealth	0.069**	0.054	0.036	0.070**	0.0030	0.057*
Age	−0.0028	0.015	0.0041	−0.0059	0.0014	−0.0081
Age sq	5.5e – 06	−0.00021	−0.00010	0.000020	−0.000057	0.000058
Female head	−0.043*	−0.098	−0.13	−0.018	0.10	−0.23*
Lusaka	0.066	0.0030	0.27	−0.017	−0.36	0.19
Copperbelt	0.0060	−0.082	−0.046	0.095	−0.23	0.041
Central	−0.063*	−0.023	−0.0029	−0.075	−0.38	−0.023
Southern	−0.11**	−0.038	−0.043	−0.15**	0.058	−0.22
Luapula	−0.037	−0.071	−0.091	−0.030	−0.27	−0.12
Northern	0.014	0.12	0.033	0.025	0.067	0.019
N. Western	−0.00075	−0.0082	−0.11	−0.058	0	0.0027
Western	−0.12**	0.0095	−0.11	−0.079	−0.00049	−0.015
Lag Log income per labourer	−0.24**	−0.20**	−0.14**	−0.24**	−0.11	−0.16**
Constant	3.11**	0.65	2.47**	3.61**	−3.20	1.08
Observations	2,396	212	378	1,279	111	216
R^2	0.339	0.367	0.243	0.318	0.295	0.273

Notes: Robust standard errors and weights are used. Eastern is the default province. The independent variables are for 2001.
[a]The average in the district, excluding the observation itself.
*Significant at 5%.
**Significant at 1%.

7. Policy conclusions

The analyses in the previous sections show that smallholders in Zambia are dependent on a range of off-farm income sources. Our analysis shows that greater diversification is associated with higher income per labourer. Paving the way for diversification is thus a key ingredient in a package of poverty-reduction policies. The diversification route to higher income for rural households requires a well-functioning economic environment and policies that make it possible for new income-generating activities to emerge. Of course, poverty may also be reduced by households leaving agriculture altogether and migrating to town. In fact, this is expected to be the long-term pattern, but at the present stage in the development of Zambia this type of migration is only realistic for a minority of the rural poor.

We identified three main sets of determinants of diversification that can be affected by policy interventions, namely factors related to endowments, markets and transaction costs and finance. The most basic endowment of a smallholder household is land. Our random effects estimations show that land has a strong impact on income. There is a high variation within districts in terms of land ownership. Re-allocation of land towards the poorer households could play a significant role in a policy aiming at poverty reduction. Human capital is another key endowment of smallholder households. Farmers with higher education can improve productivity on the farm, but education also opens up opportunities for non-agricultural wage work. The effect of advice on cropping and information on agricultural commodity prices was found to be positive for household income, and it does affect diversification behaviour.

The second set of determinants of diversification relates to market access. Local demand opens up for diversification, both into wage work and own business. Access to transportation primarily benefits diversification into non-agricultural wage work. This suggests that measures aimed to reduce costs of transactions and interactions for households support rural growth.

Third, we found weak evidence that the lack of finance is a constraint on income diversification and growth in rural Zambia. Being in a district with credit available for diversification out of agriculture seems to facilitate diversification into own business. We found that wealthier households had higher income growth during the period, which further highlights the importance of cash constraints. They hinder diversification into both business and new crops. This suggests that improved access to credit matters should be part of the rural development strategy.

We have shown that rural income growth is associated with diversification of income sources rather than specialisation in early stages of development. Policy-makers should keep in mind that rural household incomes are not derived from agriculture alone, and a strong focus in development policy should be placed on measures to facilitate smallholder income diversification. These are policies aimed at building up the smallholder households' assets, to develop the economic environment so that smallholders get better market access and to develop credit facilities that are accessible for smallholders.

Funding

We gratefully thank the Swedish International Development Cooperation Agency, Department for Research Cooperation, for financial support.

Acknowledgements

We thank all those who offered comments. The authors also would like to thank the Food Security Research Project in Zambia, and in particular Michael Weber, for access to data and helpful comments. We further are thankful to Finn Tarp, Jan Pettersson, Måns Söderbom and two anonymous referees for comments and suggestions. All remaining errors are our own.

References

Barrett, C.B. (1998) 'Immiserized Growth in Liberalized Agriculture', *World Development*, 26 (5): 743–53.

Barrett, C.B., M. Bezuneh and A. Aboud (2001a) 'Income Diversification, Poverty Traps and Policy Shocks in Côte d'Ivoire and Kenya', *Food Policy*, 26 (4): 367–84.

Barrett, C.B., M. Bezuneh, D.C. Clay and T. Reardon (2005) 'Heterogeneous Constraints, Incentives and Income Diversification Strategies in Rural Africa', *Quarterly Journal of International Agriculture*, 44 (1): 37–60.

Barrett, C.B. and T. Reardon (2000) Asset, Activity, and Income Diversification Among African Agriculturalists: Some Practical Issues, Working Paper 2000–19. Department of Applied Economics and Management, Ithaca, NY: Cornell University.

Barrett, C.B., T. Reardon and P. Webb (2001b) 'Nonfarm Income Diversification and Household Livelihood Strategies in Rural Africa: Concepts, Dynamics, and Policy Implications', *Food Policy*, 26 (4): 315–31.

Bigsten, A., B. Kebede and T. Mekkonen (2003) 'Growth and Poverty Reduction in Ethiopia: Evidence from Household Panel Surveys', *World Development*, 31 (1): 87–106.

Bigsten, A. and S. Tengstam (2010) 'Renewed Growth and Poverty Reduction in Zambia', *Zambia Social Science Journal*, 1 (1): 1–28.

Block, S. and P. Webb (2001) 'The Dynamics of Livelihood Diversification in Post-famine Ethiopia', *Food Policy*, 26 (4): 333–50.

Canagarajah, S., C. Newman and R. Bhattamishra (2001) 'Non-farm Income, Gender, and Inequality: Evidence from Rural Ghana and Uganda', *Food Policy*, 26 (4): 405–20.

Carter, M.R. and C.B. Barrett (2006) 'The Economics of Poverty Traps and Persistent Poverty: An Asset-based Approach', *Journal of Development Studies*, 42 (2): 178–99.

Carter, M.R. and J. May (1999) 'Poverty, Livelihood and Class in Rural South Africa', *World Development*, 27 (1): 1–20.

Central Statistical Office, Republic of Zambia (2005) Living Conditions Monitoring Survey Report 2004. Lusaka: Central Statistical Office.

Dercon, S. (1998) 'Wealth, Risk and Activity Choice: Cattle in Western Tanzania', *Journal of Development Economics*, 55 (1): 1–42.

Dercon, S. and P. Krishnan (1996) 'Income Portfolios in Rural Ethiopia and Tanzania: Choices and Constraints', *Journal of Development Studies*, 32 (6): 850–75.

Doss, C., J. McPeak and C. Barrett (2008) 'Interpersonal, Intertemporal, and Spatial Variation in Risk Perceptions: Evidence from East Africa', *World Development*, 36 (8): 1453–68.

Ellis, F. (1998) 'Survey Article: Household Strategies and Rural Livelihood Diversification', *Journal of Development Studies*, 35 (1): 1–38.

Ellis, F. (2000) 'The Determinants of Rural Livelihood Diversification in Developing Countries', *Journal of Agricultural Economies*, 51 (2): 289–302.

Ellis, F. and G. Bahiigwa (2003) 'Livelihoods and Poverty Reduction in Uganda', *World Development*, 31 (6): 997–1013.

Ellis, F. and H.A. Freeman (2004) 'Rural Livelihoods and Poverty Reduction Strategies in Four African Countries', *Journal of Development Studies*, 40 (4): 1–30.

Evans, H.E. and P. Ngau (1991) 'Rural–Urban Relations, Household Income Diversification and Agricultural Productivity', *Development and Change*, 22 (3): 519–45.

Green, W.H. (2003) *Econometric Analysis*, 5th edn. Upper Saddle River, NJ: Prentice Hall, Pearson Education.

IMF (2007) International Financial Statistics. May. Washington, D.C.: International Monetary Fund.

Jalan, J. and M. Ravallion (2002) 'Geographic Poverty Traps? A Micro Model of Consumption Growth in Rural China', *Journal of Applied Econometrics*, 17 (4): 329–46.

Jayne, T.S., J. Govereh, P. Chilonda, N. Mason, A. Chapoto and H. Haantuba (2007) Trends in Agricultural and Rural Development Indicators in Zambia, FSRP Working Paper No. 24. Lusaka: Food Security Research Project.

Lanjouw, P., J. Quizon and R. Sparrow (2001) 'Non-agricultural Earnings in Peri-urban Areas of Tanzania: Evidence from Household Survey Data', *Food Policy*, 26 (4): 385–403.

Lay, J. and T.O. Mahmoud (2008) 'Few Opportunities, Much Desperation: The Dichotomy of Non-agricultural Activities and Inequality in Western Kenya', *World Development*, 36 (12): 2713–32.

Lay, J., U. Narloch and T.O. Mahmoud (2009) 'Shocks, Structural Change, and the Patterns of Income Diversification in Burkina Faso', *African Development Review*, 21 (1): 36–58.

Lewis, W.A. (1954) 'Economic Development with Unlimited Supplies of Labor', *The Manchester School of Economic and Social Studies*, 22: 139–91.

Reardon, T. (1997) 'Using Evidence on Household Income Diversification to Inform Study of the Rural Non-farm Labor Market in Africa', *World Development*, 25 (5): 735–48.

Reardon, T., C. Delgado and P. Matlon (1992) 'Determinants and Effects of Income Diversification amongst Farm Households in Burkina Faso', *Journal of Development Studies*, 28 (1): 264–96.

Reardon, T., J.E. Taylor, K. Stamoulis, P. Lanjouw and A. Balisacan (2000) 'Effects on Nonfarm Employment and Rural Income Inequality in Developing Countries: An Investment Perspective', *Journal of Agricultural Economics*, 51 (2): 266–88.

Republic of Zambia (2001) *Supplemental Survey to the 1999/00 Post-harvest Survey. Interviewers' Instruction Manual*. Lusaka: Central Statistical Office.

Republic of Zambia (2004) *Rural Income and Livelihoods Survey 2004. Second Supplemental Survey to the 1999/00 Post-harvest Survey. Interviewers' Instruction Manual*. Lusaka: Central Statistical Office.

Singh, I., L. Squire and J. Strauss (eds) (1986) *Agricultural Household Models*. Baltimore: John Hopkins.

Thurlow, J. and P. Wobst (2004) The Road to Pro-poor Growth in Zambia: Past Lessons and Future Challenges, Mimeo. Washington, D.C.: IFPRI.

Woldenhanna, T. and A. Oskam (2001) 'Income Diversification and Entry Barriers: Evidence from the Tigray Region of Northern Ethiopia', *Food Policy*, 26 (4): 351–65.

World Bank (2007) World Development Indicators. New York: World Bank.

World Health Organization (1985) Energy and Protein Requirements. Geneva: WHO Technical Report Series 724.

Zimmerman, F.J. and M.R. Carter (2003) 'Asset Smoothing, Consumption Smoothing and the Reproduction of Inequality under Risk and Subsistence Constraints', *Journal of Development Economics*, 71 (2): 233–60.

Zulu, B., T.S. Jayne and M. Beaver (2007) Smallholder Household Maize Production and Marketing Behaviour in Zambia and Its Implications for Policy, FSRP Working Paper No. 22. Lusaka: Food Security Research Project.

Appendix A. The Variables

Activity combinations

F	Farm income
A	Agricultural wage income (or 'Farm work')
N	Non-agricultural wage income (or 'Non-farm work')
B	Own-business income
FA	Farm income and agricultural wage income
FN	Farm income and non-agricultural wage income
etc...	
FAB	Farm income, agricultural wage income and own-business income
etc...	

Other variables

Female head	The household has a female household head
Age	The age of the household head
Agesq	Age squared
Labour force	The number of household members aged 23–59
Total income	The total income for the household in June/July 2004 ZMK. The discount factor 1.7619 was used (IMF 2007), based on CPI for April/May 2001 and June/July 2004
Income per labourer	Total income/Labour force
Income per labourer growth	Annual percentage growth of Income per labourer
Log income in district	Average Log income per labourer in district, excluding the household itself.

Land (ha)	Total area land, cultivated and fallow. Virgin land was not asked for in 2004 so it cannot be included. Garden land plots smaller than 0.0625 ha (0.25 lima) is not considered cultivation in line with Central Statistics Office.
Land per labourer	Land/Labour force
Education (years)	The highest education of any household member 23–59 years old (i.e., of the Labour force)
Distance to vehicular transport (km)	'How many kilometres is your homestead from the point where household members can get vehicular transport?' Question only included for 2004 (time invariant).
Loans in district (%)	The percentage of the households in a district that have obtained an agricultural loan from a source that are not specialised to agricultural activities, excluding the household itself. Average of 2001 and 2004. The questions included for 2001: 'Did you receive any of the agricultural loans applied for from the formal sector for the 1999/2000 agricultural season?' and 'Did you obtain agricultural loans in cash from the informal sector (i.e., friends, neighbours, money lenders) for the 1999/2000 agricultural season?' The question included for 2004: 'Did the household borrow money or obtain a loan (cash or in-kind) from an individual or company to support agricultural production during the 2002/2003 season?' If the source of the loan was a cotton, tobacco or other crop company, or the Food Reserve Agency or farmers' association, the source was considered targeting only agriculture, and it was not included in our variable. More suggestions of sources were given than for 2001 (time invariant).

Cropping advice (Y/N)	The households have received cropping advice. Average of 2001 and 2004. It is an average of two dummies, so it takes the values 0, 0.5 or 1. The question included for 2001: 'Did this household receive advice during the 1999/2000 agricultural season on how to run farm operations as a business, or any advice on improved/recommended cropping practices?' The question included for 2004: 'Have you ever received any advice to plant 20 kg/ha of hybrid maize seed, to apply 4 × 4 bags/ha of basal and top dressing fertiliser, on practising minimum tillage (planting basins or ripping), to leave crop residues in the field, or on nitrogen-fixing crop rotation?' (time invariant).
Cropping advice in district (%)	The percentage of the households in a district that have received cropping advice (Cropping advice = 1 for that particular year), excluding the household itself. Average of 2001 and 2004 (time invariant).
Agri-commodity info (Y/N)	The households have received information on agricultural commodity prices. Average of 2001 and 2004. It is an average of two dummies, so it takes the values 0, 0.5 or 1. The question 2001 was about the 1999/2000 agricultural season. The question 2004 was if the households have ever received such information, and suggestions of sources of information were given 2004 (time invariant).
Agri-commodity info in district (%)	The percentage of the households in a district that have received information on agricultural commodity prices (Agri-commodity info = 1 for that particular year), excluding the household itself. Average of 2001 and 2004 (time invariant).

Wage level (in thousands of ZMK)	'If you were to hire someone to weed a 1 lima field, how much would you have to pay in ZMK?' Question only included for 2004 (time invariant).
Wage level in district (in thousands of ZMK)	Average wage in district, including the wage demand reported by the household itself (time invariant).
Wealth (index)	Wealth is an index of whether a household owns any of plough, harrow, water pump and bicycle.[21] These dummies were normalised by subtracting the average and dividing by the standard deviation. Using principal component analysis, an appropriate linear combination (index) of the four variables was found. The first principal component (eigenvalues) was found. Finally, the index was constructed as the sum of the four variables, weighted by the principal components. The average of the index for 2001 and that for 2004 are used in order to avoid reversed causality (time invariant).

[21] These four variables were used since they were the best indicators of wealth included in the survey both years.

Appendix B. Summary Statistics for the Panel Data Set

Variables	2001						2004					
	Weighted		Unweighted				Weighted		Unweighted			
	N	Mean	Mean	Standard deviation	Minimum	Maximum	*N*	Mean	Mean	Standard deviation	Minimum	Maximum
Female head	5,242	0.215	0.203	0.402	0	1	5,243	0.234	0.221	0.415	0	1
Age	5,211	46.06	46.62	15.18	13	97	5,211	48.63	49.18	15.01	16	95
Age sq	5,211	2,353	2,404	1,529	169	9,409	5,211	2,591	2,643	1,582	256	9,025
Land per labourer	4,931	1.213	1.330	1.648	0.0150	35	4,911	0.851	0.927	1.355	0	40
Labour force	5,358	1.885	1.946	1.162	0	15	5,358	2.361	2.450	1.471	0	16
Log income in district	5,358	13.30	13.31	0.327	12.08	14.31	5,358	13.22	13.22	0.329	11.99	14.21
Agri commodity info	5,358	0.295	0.302	0.320	0	1	5,358	0.295	0.302	0.320	0	1
Agri commodity info in district	5,358	0.294	0.292	0.149	0.0285	0.719	5,358	0.294	0.292	0.149	0.0285	0.719
Cropping advice	5,358	0.333	0.342	0.345	0	1	5,358	0.333	0.342	0.345	0	1
Cropping advice in district	5,358	0.278	0.278	0.123	0.0462	0.572	5,358	0.278	0.278	0.123	0.0462	0.572
Non ag loans in district	5,358	0.0262	0.0270	0.0230	0	0.164	5,358	0.0262	0.0270	0.0230	0	0.164
Wage level in district	5,358	29.42	29.46	7.145	12.97	56.28	5,358	29.42	29.46	7.145	12.97	56.28
Wealth	5,358	−0.152	−0.0369	1.134	−0.842	7.180	5,358	−0.152	−0.0369	1.134	−0.842	7.180

Distance to vehicular transport	5,331	8.693	8.582	15.87	0	145	5,331	8.693	8.582	15.87	0	145
Education	4,963	6.527	6.672	3.595	0	19	5,073	6.903	7.070	3.660	0	19
Log income per labour	4,915	13.31	13.41	1.109	7.186	17.84	5,056	13.22	13.28	1.174	8.112	18.76
Income per labourer growth							4,823	−0.030	−0.044	0.401	−2.289	1.942
Newhead	5,358						5,358	0.0629	0.0622	0.241	0	1
F	5,327	0.531	0.530	0.499	0	1	5,353	0.561	0.554	0.497	0	1
FA	5,327	0.0446	0.0449	0.207	0	1	5,353	0.0495	0.0495	0.217	0	1
FB	5,327	0.261	0.263	0.440	0	1	5,353	0.212	0.217	0.412	0	1
FN	5,327	0.0781	0.0760	0.265	0	1	5,353	0.0887	0.0899	0.286	0	1
FAB	5,327	0.0232	0.0240	0.153	0	1	5,353	0.0224	0.0226	0.149	0	1
FNB	5,327	0.0429	0.0430	0.203	0	1	5,353	0.0401	0.0404	0.197	0	1
Other activity combination	5,327	0.0183	0.0191	0.137	0	1	5,353	0.0254	0.0265	0.161	0	1
Central	5,358	0.108	0.107	0.309	0	1	5,358	0.108	0.107	0.309	0	1
Copperbelt	5,358	0.0566	0.0582	0.234	0	1	5,358	0.0566	0.0582	0.234	0	1
Eastern	5,358	0.222	0.210	0.407	0	1	5,358	0.222	0.210	0.407	0	1
Luapula	5,358	0.118	0.115	0.319	0	1	5,358	0.118	0.115	0.319	0	1
Lusaka	5,358	0.0253	0.0300	0.171	0	1	5,358	0.0253	0.0300	0.171	0	1
Northern	5,358	0.176	0.188	0.391	0	1	5,358	0.176	0.188	0.391	0	1
N. Western	5,358	0.0723	0.0601	0.238	0	1	5,358	0.0723	0.0601	0.238	0	1
Southern	5,358	0.113	0.123	0.328	0	1	5,358	0.113	0.123	0.328	0	1
Western	5,358	0.109	0.109	0.312	0	1	5,358	0.109	0.109	0.312	0	1

Appendix C. Descriptive Tables

Table A1: The Provinces

Province	Capital	Average income per household and month (in thousands)	Population	Density (people/ km^2)	Line of rail	Distance to market (km)[a]
Lusaka	Lusaka	734	1,391,329	63.5	×	4.2
Copperbelt	Ndola	665	1,581,221	50.5	×	3.9
Central	Kabwe	443	1,012,257	10.7	×	17.6
Southern	Livingstone	474	1,212,124	14.2	×	16.4
Eastern	Chipata	490	1,306,173	18.9		20.0
Luapula	Mansa	318	775,353	15.3		18.6
Northern	Kasama	378	1,258,696	8.5		25.0
Northwestern	Solwezi	427	583,350	4.6		19.7
Western (Barotseland)	Mongu	356	765,088	6.1		23.0
Zambia	Lusaka	502	9,885,591	13.1		14.8

Note: Income from Central Statistical Office (2005). Population (for 2000) and population density from Administrative Divisions of Countries by Gwillim Law. Distance to markets from Thurlow and Wobst (2004).
[a]Distance to market = average distance from household to food and input markets.