CEREAL PRODUCTIVITY IN ETHIOPIA: AN ANALYSIS BASED ON *ERHS* DATA[1]

Getachew Ahmed Abegaz[2]

Abstract

This paper examines the recent phenomenon of cereal yield growth in Ethiopia and tries to see yield responses to modern inputs such as chemical fertilizer and improved seeds on major cereal crops. It bases its analysis on two rounds of the Ethiopian Rural Household Surveys (ERHS). Results show that cereal yield grew by 21 percent during the period 1999 and 2009, much lower than the national figure, which is 60 percent, for the same period. This growth was contributed by wheat, maize and barley, which grew by 62, 19 and 11 percent respectively. The study further indicates that the source of this yield growth can be partly explained by modestly increasing use of modern inputs. It shows more intensification of modern agricultural inputs than that which the CSA data shows during the period. Overall, regression results in the two periods show that yield response to fertilizer and improved seeds was found to be statistically significant. However, using panel data analysis, the study also found an indication of some yield growth, unrelated to inputs such as seeds and fertilizer. This cannot be explained by weather changes and needs further research to capture its source in a time series setting.

Keywords: Cereal yield, yield growth, yield response to modern agricultural inputs

JEL Classification: O4, O13, O33, Q18

---

[1] *ERHS* is short for Ethiopian Rural Household Surveys. See section 2 for a full description of these datasets.

[2] Ethiopian Economics Association/Ethiopian Economic Policy Research Institute (EEA/EEPRI)

1. Introduction

Official sources revealed that cereal yield recorded a significant growth in Ethiopia for the last 10 years. Across the main cereals, production increased from a level of 10.3 quintals per hectare in 1999 to 16.5 quintals per hectare in 2009; growth has been steady, with a temporary decrease due to drought in 2003 (see Table 1 and Annex 1). Production of all the major cereal crops seems to have been increasing in unison especially in recent years (see Annex 1). There is no significant difference in yield growth of these crops unlike the experiences of other countries, which showed high yield growth during the green revolution, and where crop specificity is an important feature (Gollin et. al., 2011).

Table 1: Yield (quintal per hectare) for Major Cereal Crops: 1999 and 2009

Cereals	1999	2009	Annual AverageGrowth Rates
Cereals	10.3	16.5	2.9
Teff	7.9	12.2	3.5
Barley	9.3	15.5	4.1
Wheat	11.3	17.5	3.9
Maize	18.5	22.2	1.8
Sorghum	12.7	17.4	3.2

Source: Data from Central Statistical Agency (CSA), Different Agricultural Sample Surveys

Cereal crop production has been lifted well above long-term levels. Although area expansion has been considerable, yield growth has accelerated more than area expansion particularly in the last few years (see Annex 2). However, the recent large yield increases do not seem to be explained by a sudden large increase in uses of modern inputs and improved farm management. Chemical fertilizer application is still low and only about 36.5 percent of total cereal acreage benefit from chemical fertilizer in 2009 (CSA, 2009; Gollin et. al., 2011). It also shows that little change was registered by these shares over the period 2005 and 2009. Similar stories can be told of other modern inputs: use of improved inputs did not expand in such an overwhelming rate, as the yield growth did.

The share (in percent) of area under improved farm management for three time points is provided in Annex 3. In addition to the low-level use of these improved farm management practices, little progress is shown for the last ten years. The maximum share of area cultivated with improved seeds is 20 percent for maize in 2009. On the other hand, the figure for wheat has even declined from 6 percent in 1998 to 4 percent in 2009. It is important to note that improved seeds are largely limited to only two crops, namely wheat and maize. More success is shown for wheat in terms of area where pesticide is applied, which is 41 percent in 2009 from 31 percent in 1998.

The foregoing analysis provides us with at least four salient features of the recent growth of cereal yield in Ethiopia. First, cereal yield recorded a significant growth in recent few years and this productivity is not crop specific. Second, although both output and area cultivated grew, the growth rate in output outpaced area expansion. Third, the use of modern inputs such as fertilizer, improved seeds and other modern farm management practices is low and has not been expanding for the same period under consideration. With these points in mind, and the fact that this increase in production cannot easily be justified given the massive increase in local food prices for an agriculture-dominated economy, where the rise and fall of food price is crucially dependent on agricultural production, it is imperative to see whether these changes are structural breaks or an unexpected shift.

Few studies were made on these recent phenomena. Minot (2008) cited in Dercon and Hill (2009) shows that this could be an overestimation of yields in CSA estimations in recent years. Some of the yield growth rates appear to outpace other East African countries and even Green Revolution India, especially taking into account the low growth in input intensity (Gollin et. al., 2011; Dercon and Hill, 2009).

A study on total factor productivity in agriculture by Fantu (2012) using a specification that used three primary inputs indicate that annual changes in total factor productivity averaged 4.5 percent during the five year period of 2004/05-2005/06 through 2008/9-2009/10. This figure declines to 1.4 percent for the same period with another specification accounting for factors that contribute to increased agriculture output. Pingali and Heisey (2009) documented studies on sources of total factor productivity in many developing countries. The majority of these studies found that modern farming technologies such as improved seeds (or high yielding varieties) and technology embodied in chemical fertilizer contribute to growth in total factor productivity. However, a study by Dercon and Hill (2009) shows that despite some rhetoric, to the contrary, the availability of appropriate and high return technologies on the ground is limited at present in the Ethiopian case.

There are many factors for this low-level use of high return technologies in Ethiopia and other sub-Saharan Africa. High transport costs, unfavorable climate, price risk, and illiteracy of household how to apply them, availability of fertilizer, price policies and credit availability, pricing environment and distribution costs, and infrastructural development are hurdles in the effective use of fertilizer markets (see for e.g. Daniel and Larson, 2010; Mwangi, 1996).

An impressive growth in cereal production with no evidence on intensification of agriculture such as increase in the use of modern inputs at the same time need a way of validating or invalidating the data on cereal production using other sources. This piece of work bases its analysis on the Ethiopian Rural Household Surveys (ERHS). The objective of the study is to examine recent cereal yield growth in Ethiopia and tries to see yield responses to modern inputs such as chemical fertilizer and improved seeds on major cereal crops.

Cereal crops constitute the largest share of farming household's production and consumption activities. Only five major cereals (barley, maize, sorghum, teff and wheat) account for about 70 percent of area cultivated and 65 percent of output produced (Alemayehu et. al, 2009). Moreover, according to Household Income, Consumption, and Expenditure Survey (HICE), these cereal crops account for 46 percent of household's total consumption. Therefore, a closer look at what is happening in cereal production has an important welfare and policy implication in Ethiopia.

The second section briefly describes the nature and source of data used. Furthermore, a production function is specified to see the impact of some important inputs on cereal yield. Discussion of results will be made in the third section. The fourth section concludes.

2. Data and Methods

This study uses data from the Ethiopian Rural Household Surveys (ERHS). ERHS is a longitudinal dataset and have been supervised by the Economics Department of Addis Ababa University, the Centre for the Study of African Economies (CSAE), University of Oxford and the International Food Policy Research Institute (IFPRI). Production data were collected at plot level allowing analysis by plot and/or by crop and making sample size larger for consistent estimation.

Both descriptive and analytical methods are employed in this paper. A determinant of crop productivity analysis was made to see the response of cereal yield to modern inputs such as chemical fertilizer and improved seeds. The yield response functions can easily be derived from Agricultural Household Models. In this model (more on this is a classic work by Singh, Squire and Strausss (1986)), production and consumption decision making is thought to be a recursive one. Production decisions are made first through maximizing profits; consumption decisions are then made through utility maximization subject to those profits. However, this happens if markets (both input and output markets) exist and function perfectly. In this ideal case, these decisions are separable, household preferences and endowments do not affect production, and the models can be solved recursively.

This raises a question of whether these output and input markets exist and function perfectly. Markets are either thin or absent in Ethiopia. Hence, it is impossible to separate the production and consumption decisions. As a result, the following production function is used.

$$\boldsymbol{Y = Y(V, K, P, E)}$$

Where Y- Level of output of crop produced by a household; V = Variable input level; K= Fixed input levels; P= Household preferences; E= Household endowments. The following empirical model is specified to link yields with major inputs such as fertilizer

$$\boldsymbol{y_{ic} = \beta_1 + \beta_2 f_{ic} + \beta_3 f_{ic}^2 + s_{ic} + \sum_{n=1}^{n=k} x_{ic}' \beta_n + \varepsilon_{ic}}$$

Where,

$\boldsymbol{y_{ic}}$= production of crop c per hectare of land (yield) for household i

$\boldsymbol{f_{ic}}$= chemical fertilizer application (in kg) for crop c for household i

$\boldsymbol{s_{ic}}$= amount of improved seeds used for crop c by household i

$\boldsymbol{x_{ic}'}$= other control variables including characteristics of the farm household (holder), characteristics of land, traditional fertilizer, dummy for villages, year etc.

Table 2: List of Some of the Questions/Variables in the Ethiopian Rural Household Survey (ERHS)

Variables (questions)	Variable Description/responses
Total output by crop/plot (in quintals)	Teff, Barley, Wheat, Maize & Sorghum
Plot Area	Plot size in hectares
How much DAP/Urea/DAP+Urea did you apply to this plot?	Amount in kg
Improved seed saved from last year/ bought/exchanged	Amount in kg
Did you apply manure/compost to this plot?	Yes=1, No=0
Extension coverage	Yes=1, No=0
Is the land Lem, Lem-teuf or teuf land?	Leum, Teuf, Leum-teuf
Slope of land	Medda, dagathama, geddel
How much of this plot is irrigated?	Percentage of irrigated land
Education of household head	Highest grade obtained
Age of the household head	Number of years
Household size of the household	Number of household members
Gender of household head	Male=1, Female=0

A list of variables and their description is offered in Table 2. This model is estimated by OLS. This is done for the total sample and for each of the cereal crops under study. A fixed effects model of the following type was specified to allow time-constant unobserved heterogeneity correlate with explanatory variables.

$$\boldsymbol{y_{itc} = \beta_1 + \beta_2 f_{itc} + \beta_3 f_{itc}^2 + s_{itc} + \sum_{n=1}^{n=k} x_{ic}' \beta_n + \alpha_{ic} + \varepsilon_{itc}},$$

where α_{ic} is individual specific unobserved heterogeneity. This model is estimated using the levels and logs of yield. Fixed effects estimator was used.

3. Discussion of Results

3.1. Cereal yield and use of modern agricultural inputs in Ethiopia

After taking into account some adjustments on the data[3], Table 3 presents estimates of yield for the five major cereal crops (teff, barley, wheat, maize and sorghum) in the two rounds of ERHS (1999 and 2009). The CSA estimation of the same variable is presented for comparison. We have to be careful interpreting these results. These are just two years and they do not tell the stories in full. However, it is still useful for mean yield comparison between national figures. Moreover, even though the national trend seems robust to weather events (i.e. not just explained by 'good weather' in 2009–see Gollin et. al., 2011), in a smaller sample this may affect the findings. Furthermore, these are not necessarily representative areas for actual growth – nevertheless, as the national data suggest broad and widespread growth, as a first approximation for a comparison, this is still worth exploring.

We present growth rates for all cereals and for specific crops, and compare growth rates between ERHS and CSA sources. The general observation is cereal yields in the CSA data are much higher than that in ERHS: overall, the difference is 7 percent in 1999 and 41 percent in 2009. This difference is highest for maize in 1999 and sorghum in 2009. Maize and wheat *growth* rates are nevertheless comparable: for maize, yields in CSA increased by 20 percent and in the ERHS by 19, wheat yields increased 55 percent in the CSA data and by 62 percent in the ERHS in this period. For teff, barley and sorghum, the gaps are huge. For example, stagnation of teff yields in the ERHS translates into a 54 percent growth

[3] Data on output and inputs per crop and plot were collected. However, due to a matching problem with the 2009 data, it has proven impossible to link all the input data to the output data. We had to restrict the matching to plots that did not intercrop nor had multiple crops on, and only those used in one season (either Belg or Meher). Arguably, that is not fundamentally problematic as long as in both years (1999 and 2009) the same restrictions on plots are applied, and our interest is not in studying and comparing the *levels* of yields in 1999 and 2009, but to study growth rates. As we have access to the fully matched data set for 1999, we offer some comparison between the 'adjusted' 1999 and the 'raw' 1999 data. Furthermore, yield data are based on self-reported output, and outliers appear to affect the results, especially in 2009 data – and the most unlikely values accounting for 1.3% of observations were dropped. It involves observations with reported harvests of more than 100 quintals per hectare. To decide the cut-off point for unlikely yields, we used as a benchmark yields from a series of high yielding countries. The assumption is that with all the available technology, natural fertility of the land and good weather prevailing in the country, a yield bigger than this benchmark yield must be an outlying case. Then, observations with bigger yield than these high yielding countries were dropped. This is done by crop. In the case of teff (local cereal crop), a benchmark of 3Xmedian better treats outliers and bigger than this benchmark were dropped. As a result, in total 37 observations (barley=6, wheat=4, maize=18, sorghum=2 and teff=7), much less than 2 percent of total sample, were dropped.

in the CSA data. The overall result is that the national yield data grew three times as fast as in the ERHS.

Table 3: Cereal Productivity (yield in quintal per hectare) based on ERHS & CSA by Crop: 1999 and 2009

Crops	ERHS			CSA			Difference (%)	
	1999	2009	Period Growth (%)	1999	2009	Period Growth (%)	1999	2009
	A	B	(B/A-1)*100	C	D	(D/C-1)*100	(C/A-1)*100	(D/B-1)*100
Cereals	9.6	11.7	21.2	10.3	16.5	60.2	6.9	41.3
Teff	8.3	8.2	-0.4	7.9	12.2	54.4	-4.3	48.3
Barley	8.7	9.7	11.1	9.3	15.5	66.7	6.7	60
Wheat	10.1	16.3	61.6	11.3	17.5	54.9	12.3	7.6
Maize	12.1	14.5	19.4	18.5	22.2	20	52.7	53.4
Sorghum	13.1	8.2	-37.4	12.7	17.4	37	-3.3	111.7

Source: ERHS with author's calculations, and CSA Agricultural Sample Survey

Do increased yields in the ERHS data for some cross match with input use data[4]? Looking into input use, there seems to be a modest increase in the use of modern inputs such as improved seeds and chemical fertilizers. Table 4 shows the share of area cultivated with modern inputs (improved seeds, fertilizer and irrigation). In the case of improved seeds, while the share is still low (i.e. only 14 percent of the cereal crop area cultivated with improved seeds), there is a big jump in the application of these seeds over the ten years period. Wheat and maize have the largest share of cultivated area with improved seeds in both periods with 25 and 33 percent in 2009 compared to 10 and 16 percent in 1999 respectively. A comparison can be made between the estimation of ERHS and CSA data using Table 4 and Annex 3. Overall, the estimation based on the ERHS data shows more intensification of improved seeds.[5]

[4]In addition, the input data were characterized by outliers. A few negative values were dropped. There were also other unlikely observations in terms of inputs per hectare. However, as the data in the table are reported in area under modern inputs, this does not affect the data here.

[5]One may worry about the data on improved seeds in surveys like this. With limited certification, farmers may believe certain seeds to be 'improved' even if they are not, or continue to recycle seeds e.g. for maize, even though they would gradually lose their extent of improvement. In self-reported data, this problem is not easily overcome. One example is the reported improved seeds for teff. In general, they are rarely used in Ethiopia – although this does not mean that farmers and research stations have used selection to get better yielding seeds in various places. The data has, however, substantial observations with improved seeds for teff as shown in the tables. Moreover, sorghum has no improved seeds applied, barley has only three observations with improved seeds, and this variable is left out of the regression model in the main text.

Another important input expected to have an impact on cereal production is fertilizer. We find that the area cultivated using fertilizer is 64 percent, for all cereals, in 2009, which has increased from its level of 55 percent in 1999. This is bigger than the estimations using CSA data, which is still close to 36.5 percent (see section 1). The crop with the highest application of fertilizer in both periods is wheat (90 percent in 2009 from 71 percent in 1999). Cereal crop area cultivated with irrigation is also a bit higher, with 4 percent of the total area irrigated, than the national average (CSA data). Overall, we note higher intensification in the ERHS sample than the national data, although only for wheat and maize, they kept up with national yield growth. In the next section, we explore this further and study how the yields respond to inputs in the ERHS data for 1999 and 2009.

Table 4: The Share of Areas (in percent) Cultivated with Modern Inputs: (1999-2009)[6]

Crops	Crop Area with Improved Seeds		Crop Area Applied with Fertilizer		Area Cultivated with Irrigation	
	1999	2009	1999	2009	1999	2009
Cereals	4.1	14.2	55.4	64.4	3.3	3.9
Teff	1.3	7.2	66.6	71.7	2.3	4.0
Barley	0.5	0.9	51.5	65.5	0.9	3.3
Wheat	10.0	25.0	70.7	89.7	0.5	3.2
Maize	15.8	32.7	39.4	49.6	14.5	4.9
Sorghum	0.0	0.0	9.4	5.2	4.2	4.1

Source: Estimations using ERHS data

3.2. Responses of yield to modern inputs: Fertilizer and improved seeds

Regression results are reported based on the 2009 and1999 cross section and on panel datasets separately. The data used are the 'adjusted' data as referred to in the descriptive analysis – correcting for outliers and ensuring that only plots that can be matched to inputs are included. Multi-collinearity test was made using variance inflating factor (VIF) and this data problem was not found. However, Breusch-Pagan/Cook-Weisberg test for heteroskedasticity was made and heteroskedasticity is present in both datasets as expected because we are dealing

[6] In this table, we only report the 'adjusted' data – i.e. the data that can be compared between 1999 and 2009 for growth rates. Note that the only significant difference appears to be that for maize, once we include the intercropping rates, the data for 1999 show higher crop areas with improved seeds (23%) and with fertiliser (49%). This is consistent with the yields on the maize plots in 1999 higher when the full sample is used and not just the plots that can be compared over time. As there is no obvious reason why the *growth* in input use on these plots outpaces the other plots, it is unlikely to affect the overall results comparing 1999 and 2009.

with cross sectional datasets. We corrected this problem using robust standard errors.

Table 5 below presented estimations based on the 2009 cross section dataset. The model estimated below includes some interaction terms and all other variables. A descriptive statistics of the important variables was offered in Annex 4. Overall, we can see that, the estimated coefficients for chemical fertilizer, traditional fertilizer and improved seeds were found to be significant at 5 percent. Yield seems to respond to fertilizer application for teff, barley and wheat: for example, for the latter, using about 100 kg per hectare (the usual recommended quantity) would add about 200 kg yield – or about 14 percent relative to the mean yield observed.[7]

[7]This includes the impact of the concavity (the squared term on fertiliser is significant) reducing yield at 100 kg per hectare by 27 kg.

Table 5: Estimates of the Parameters of Quadratic Production Functions in Levels of Yield (2009)

Dependent Variable=levels of yield	Total	Teff	Barley	Wheat	Maize	Sorghum
Dummy for improved seeds	546.77	358.70		35.80	624.77	
	[135.9046]***	[319.2028]		[160.0459]	[358.1169]*	
Fertilizer application in kg per hectare	2.32	2.86	1.90	2.36	3.18	5.85
	[0.5456]***	[1.0259]***	[1.0395]*	[0.8566]***	[2.7600]	[8.4786]
Square of chemical fertilizer used in kg	-0.0004	-0.0005	0.0019	-0.0027	-0.0068	-0.0009
	[0.0002]**	[0.0003]	[0.0027]	[0.0011]**	[0.0080]	[0.0189]
Interaction term: Fertilizer & improved seeds	-0.33	-1.74		1.95	0.84	
	[0.7612]	[1.6658]		[1.1787]*	[2.2073]	
Use of traditional fertilizer	354.61	87.66	147.57	111.84	317.32	875.45
	[72.0241]***	[116.9787]	[65.8935]**	[134.4445]	[146.4417]**	[541.3394]
Irrigation: 1 if irrigated, 0 otherwise	16.93	20.30	-109.46	-501.10	287.82	-36.81
	[120.4961]	[127.1920]	[79.8081]	[579.7204]	[232.6713]	[266.8143]
Interaction term: fertilizer and Irrigation	2.10	-5.25	1.09	6.13	0.66	
	[1.2652]*	[5.5965]	[1.0566]	[4.1300]	[1.8808]	
Fertility of Soil (Lem=1, best quality)	222.45	193.87	170.14	189.22	223.12	300.70
	[56.1273]***	[84.7161]**	[81.4826]**	[145.4284]	[182.3776]	[161.7882]*
Fertility of Soil (Lemteuf=1, medium quality)	37.66	191.17	6.02	-168.49	101.28	167.82
	[52.0428]	[92.1537]**	[62.0759]	[153.3343]	[192.6097]	[140.8739]
Age of holder	1.59	25.22	10.58	16.68	-13.55	22.24
	[8.1870]	[19.5199]	[13.4008]	[21.5762]	[16.4296]	[18.1412]
The square of age of holder	-0.0214	-0.2135	-0.1336	-0.1543	0.1283	-0.2203
	[0.0723]	[0.1832]	[0.1199]	[0.1865]	[0.1271]	[0.1677]
Household size of holder	-1.20	-0.07	4.85	-23.76	9.54	-26.57

Dependent Variable=levels of yield	Total	Teff	Barley	Wheat	Maize	Sorghum
	[9.8921]	[15.1795]	[13.8531]	[16.2527]	[32.4125]	[27.4108]
Gender of holder : 1 if Male, =0 if Female	73.96	68.84	109.05	-17.51	171.94	158.52
	[46.5607]	[65.9981]	[70.7740]	[108.2452]	[138.2774]	[242.4820]
Level of Education of holder						
Literate	33.046	148.571	-70.304	-9.779	134.426	-1.413
	[56.1197]	[75.5337]**	[80.4375]	[135.9657]	[173.6497]	[141.0628]
Primary (1-8)	28.31	109.97	-71.17	64.54	80.04	247.98
	[60.9615]	[75.2230]	[95.2352]	[114.9094]	[165.1064]	[171.8447]
Secondary (9-12)	151.91	715.98	262.79	158.11	-189.60	
	[109.4432]	[280.5677]**	[258.7840]	[150.3108]	[262.9586]	
Higher Education	-91.30	224.07		-87.67	1.22	
	[270.0945]	[295.4294]		[876.2344]	[549.0548]	
Constant	-289.42	-414.81	-33.85	-354.44	420.47	606.34
	[234.0535]	[528.8299]	[422.7629]	[585.5817]	[694.3032]	[770.2023]
Observations	1721	484	427	372	408	119
R-squared	0.39	0.38	0.62	0.52	0.25	0.36
F Stat	F(37,1683)=66.1	F(30, 453)=11.2	F(25, 401)=52.8	F(30, 341)=32.0	F(30,377)=23.8	F(17,101)=5.8
Prob F>0	0.00	0.00	0.00	0.00	0.00	0.00
Root MSE	813.8	608.5	564.7	714.9	1114.6	693.2

Source: ERHS with author's estimations
Notes: Robust standard errors in brackets
*Significant at 10 percent, ** significant at 5 percent, *** significant at 1 percent.
OSS Estimations were implemented with dummies for villages whose coefficients were not reported to save space

For maize, no direct fertilizer response is significant in the data; improved seeds matter strongly for this crop, adding 625 kg per hectare of output, ceteris paribus, or more than 40 percent relative to the mean. Interaction terms between fertilizer and improved seeds were significant for wheat – adding another 190 kg yield if both improved seeds and 100 kg per hectare fertilizer were used. A few other results emerge from the controls. First, that land quality matters significantly for teff and barley (with the lowest quality land as the base group), and traditional fertilizer (manure) adds substantially too. A strong effect from secondary education for teff cultivation also emerges, but this is based on very few households –virtually no one has this grade in the data.

Table 6: Predicted Mean Difference Tests (2009)

Inputs		Sample	Total	Teff	Barley	Wheat	Maize	Sorghum
Predicted mean yield	Using Both (Seeds=1; fertilizer =100kg/ha)	A	1698.6	1189.8		1718.2	2047.4	
	Fertilizer Only (Fertilizer=100kg/ha; Seeds=0)	B	1151.8	831.1	1006.7	1682.4	1422.6	1345.6
	Seeds Only (Fertilizer=0; Seeds=1)	C	1470.6	908.6		1509.3	1796.9	
	Nether (No Fertilizer; Seeds=0)	D	923.8	549.9	798.0	1473.5	1172.2	770.1
Difference	A-B	chi2 (1)	16.2	1.3		0.1	3.0	
		Prob.	0.00	0.26		0.82	0.08	
	A-C	chi2 (1)	18.5	8.0		7.4	1.5	
		Prob.	0.00	0.00		0.01	0.23	
	A-D	chi2 (1)	24.7	3.4	6.5	2.2	4.8	0.8
		Prob.	0.00	0.07	0.01	0.14	0.03	0.38
	B-C	chi2 (1)	5.6	0.1		0.8	0.8	
		Prob.	0.02	0.81		0.36	0.38	
	B-D	chi2 (1)	18.5	8.0		7.4	1.5	
		Prob.	0.00	0.00		0.01	0.23	
	C-D	chi2 (1)	16.2	1.3		0.1	3.0	
		Prob.	0.00	0.26		0.82	0.08	

Source: Estimation from ERHS (2009)

In Table 6, the above analysis was explored further by comparing mean yield predicted on the model using the levels of yield. It shows the predicted yield values when using both fertilizer (at recommended amount which is 100kg/ha) and seeds (i.e. seeds=1); fertilizer only (recommended, 100kg) with no seeds (seeds=0); seeds only (seed=1) with no fertilizer; and no fertilizer and no seeds.

All other characteristics are at the mean values in the sample. A chi-squared test is offered to test the differences using one of the inputs and both. Overall, the data suggests using both fertilizer and seeds adds significantly to yields compared to using only either or to using neither, while for teff and wheat, using both adds to yield compared only to using seeds alone.

Table 7 provides the estimated results based on the 1999 cross section dataset. Similar analysis that we did for 2009 data was made. The same model is specified as in above and the same exercise is done in the choice of models.

Yield response for improved seeds was significant. Overall, a mean difference of 592 kg per hectare was found between plots where high yielding variety seeds are applied and those cultivated with indigenous seeds. However, these results show that there was no direct relationship found between chemical fertilizer and yield for all crops except sorghum in 1999. The fact that yield doesn't respond to chemical fertilizer in 1999 while this input has an impact in 2009 can be explained partly by a host of factors, that affect the effective utilization of chemical fertilizer (e.g. see Daniel and Larson, 2010; Mwangi, 1996), that can improve over time the application and utilization of the input.

Table 7: Estimates of the parameters of quadratic production functions (1999)

Dependent Variable = levels of yield	Total	Teff	Barley	Wheat	Maize	Sorghum
Dummy for improved seeds	591.79	1148.57		514.12	716.36	
	[225.2956]***	[635.2339]*		[296.8951]*	[396.7304]*	
Fertilizer application in kg per hectare	-0.03	1.09	-0.45	-1.01	0.57	16.73
	[0.5116]	[1.1046]	[0.5869]	[1.1222]	[2.7280]	[5.0023]***
Square of chemical fertilizer used in kg	0.0028	0.0028	0.0029	0.0045	0.0035	-0.0428
	[0.0011]***	[0.0035]	[0.0013]**	[0.0024]*	[0.0037]	[0.0128]***
Interaction term: Fertilizer & improved seeds	-0.4668	-9.6388		-0.9758	0.0054	
	[1.1185]	[3.6333]***		[1.4356]	[2.1045]	
Use of traditional fertilizer	117.87	52.81	182.67	156.36	8.74	184.13
	[73.2864]	[71.0905]	[100.7599]*	[110.6515]	[242.2739]	[230.8071]
Irrigation: 1 if irrigated, 0 otherwise	-42.03	59.16	53.09	1293.51	71.29	-517.29
	[141.6628]	[131.8240]	[381.7819]	[1,248.8145]	[557.8339]	[271.3775]*
Interaction term: fertilizer and Irrigation	0.138	1.286	-0.759	-1.238	-1.194	
	[0.6556]	[1.4085]	[0.9632]	[2.1763]	[1.8578]	
Fertility of Soil (Lem=1, best quality)	153.51	119.39	58.73	111.09	910.37	-70.35
	[64.6390]**	[96.0072]	[77.6122]	[121.7844]	[376.3740]**	[368.6901]
Fertility of Soil (Lemteuf=1, medium quality)	-3.81	7.91	19.78	-57.66	821.58	-534.92
	[56.9943]	[81.0555]	[72.2738]	[112.8357]	[371.3108]**	[289.8176]*
Slope of land (flat=1)	311.07		242.93	344.95	-165.46	1407.18
	[109.8475]***		[109.1677]**	[256.9057]	[642.1618]	[482.9372]***
Slope of land (slopy)	265.13	11.19	189.54	407.00	713.91	918.53
	[106.2958]**	[69.8737]	[94.5266]**	[273.4644]	[519.5981]	[408.2295]**

Dependent Variable = levels of yield	Total	Teff	Barley	Wheat	Maize	Sorghum
Age of holder	0.86	12.08	-11.16	9.79	-8.21	4.91
	[8.2563]	[9.2227]	[12.1470]	[15.5737]	[33.3819]	[41.3445]
The square of age of holder	-0.0490	-0.1036	0.0541	-0.1140	-0.0903	-0.1850
	[0.0734]	[0.0823]	[0.1065]	[0.1387]	[0.3118]	[0.3721]
Gender of holder : 1 if Male, =0 if Female	87.63	33.70	9.55	-19.73	830.93	289.37
	[52.6170]*	[62.4834]	[63.5428]	[90.7456]	[269.7779]***	[300.3283]
Level of Education of holder						
Literate	47.10	69.97	23.86	101.65	-342.81	19.04
	[56.1226]	[74.1757]	[78.1125]	[120.3314]	[319.2460]	[222.0788]
Primary (1-8)	55.44	75.62	11.75	282.78	-292.17	-12.17
	[62.9413]	[76.2555]	[95.6923]	[137.4341]**	[221.0256]	[311.3857]
Higher Education	-72.27	195.71		-14.42	-535.25	
	[99.0007]	[94.6586]**		[224.3400]	[269.1230]**	
Constant	587.71	-74.91	1122.97	4.72	625.01	-789.55
	[254.0041]**	[380.9521]	[373.4130]***	[533.2163]	[903.3561]	[1,085.7112]
Observations	1480	435	467	257	194	127
R-squared	0.31	0.41	0.15	0.33	0.50	0.49
F Stat	F(30,1449) = 28.99	F(25,409) =21.80	F(20,446) = 10.12	F(23,233) = 6.08	F(26,167) = 39.29	F(16,110) = 7.27
Prob F>0	0.00	0.00	0.00	0.00	0.00	0.00
Root MSE	704.17	453.69	601.63	588.26	1049.00	938.53

Source: ERHS with author's estimations.

Notes: Robust standard errors in brackets. *Significant at 10 percent, ** significant at 5 percent, *** significant at 1 percent. OSS Estimations were implemented with dummies for villages whose coefficients were not reported to save space

Bringing the 1999 and 2009 datasets together, fixed effects model was estimated and the result is reported in Table 8. First, the sample size after joining the data was not as expected (in the sense that it is much lower than 2 times the usually used ERHS datasets). This was because only part of the households was considered for analysis in both periods. On the 2009 dataset, we started with 2055 plots before data cleaning. After data cleaning, we are remained with only 2,009 observations. These are only 820 households. On the 1999 dataset, we started with 3166 plots. After cleaning 2,166 plots remained. These are 848 households. When we merge them, only 507 households matched.

Table 8: Estimates of the Parameters of Quadratic Production Functions- Panel (1999 and 2009)

Dependent Variable: Levels/logs of yield	levels of yield	logs of yield
	Model I	**Model II**
Dummy for improved seeds	317.48	0.00
	[183.6746]*	[0.1722]*
Fertilizer application per hectare	3.73	0.00
	[0.8665]***	[0.0008]***
Square of chemical fertilizer used in kg	0.003	0.000
	[0.0010]***	[0.0000]
Interaction term between Fertilizer and Improved seeds	-2.861	-0.001
	[0.8675]***	[0.0008]*
Irrigation: 1 if Part or all of plot irrigated, 0	223.6314	0.5181
	[102.1586]**	[0.2031]**
Interaction term between fertilizer and	0.228	-0.001
	[1.5910]	[0.0014]
Use of traditional fertilizer	152.67	0.16
	[74.7458]**	[0.0811]*
Year Dummy (2009=1)	28.50	0.13
	[39.2283]	[0.0474]***
Constant	571.41	6.22
	[59.1967]***	[0.0540]***
Observations	1006	942
Number of uid	507	506
R-squared	0.24	0.15
F Stat	F(8,506) = 43.40	F(8,505) = 26.06
Prob > F	0.0000	0.0000
sigma_u	518.8	0.5653
sigma_e	614.1	0.6896
rho (fraction of variance due to u_i)	0.4164	0.4019
corr(u_i, Xb)	-0.0311	0.0108

Source: ERHS with author's estimations

Notes: Robust standard errors in brackets. *Significant at 10 percent, ** significant at 5 percent, *** significant at 1 percent.

OSS Estimations were implemented with dummies for villages whose coefficients were not reported to save space

Annex 6 provides some descriptive statistics of the panel data. Table 8 presents two panel regression models. The first uses levels of yield while the second uses logarithms of yield. The log-linear regression model was run to have a rough estimate of the percentage change in the year dummy, a proxy for factors that change overtime but not considered in the model.

We find that yield responses to chemical fertilizer and improved seeds are statistically significant. The coefficient for year dummy seems to be significant. This can be translated into total factor productivity (TFP), and suggests a growth in TFP of 13 percent in this period. Although not comparable, similar results were also found by Fantu (2012). Possible candidates for the source of TFP in developing countries include institutional change, agricultural terms of trade, weather, infrastructure, and access to markets (see for e.g. Pingali and Heisey, 2009), none of which were included in this model. A rudimentary analysis is provided in Table 9 showing weather is unlikely to explain this difference: the relevant rainfall period for the 1999 and 2009 data are 1998 and 2008.

Table 9: Mean Annual Rainfall

Rainfall/Year	1998	1999	2008	2009
Mean Annual Rainfall in mm	1326.9	1292.7	1176.6	981.1
Average mean annual rainfall in mm		1136.9		
Difference (%)	16.7	13.7	3.5	-13.7

Source: Own computations using CSA annual abstract for different years

We can see that 2008 was a good year relative to the long-term mean (3.5% better than usual) but 1998 was ever relatively better, so this is unlikely to be responsible for the TFP growth in 2009 relative to 1999.

4. Concluding Remarks

Although figures from official sources show that cereal yield growth has been high in the last few years, it remains to be explained in terms of the use of modern agricultural inputs, which are very low. The objective of this study is to examine cereal yield growth and the sources of this growth using another dataset. It was found that cereal yield grew between 1999 and 2009. However, this growth is much lower than that estimated using the CSA data. The levels of yield are also lower compared with the national figures, which is higher, by 7 percent in 1999 and 41 percent in 2009. The study also showed that there seems to be a modest increase in the use of modern inputs such as improved seeds and chemical fertilizer. Overall, as far as modern input use is concerned, the estimation based on the ERHS data is somewhat higher.

Results on cross section regression show that use of improved seeds significantly affects yield in both years. They lose their economic importance in the panel data analysis however. In the more recent cross section data and panel data analysis, use of fertilizer was found to be significant, but the contribution to yield is economically rather insignificant, in comparison to seeds. An important point to note is the coefficient of the year dummy using the panel data analysis; it can be interpreted as offering TFP-style growth in yields over the decade, and it showed about 13% growth, at least in the log-specification. Important sources include institutional change over the period, the weather, infrastructure, access to markets and services not included in the model. Some simple correlates suggest that weather is not responsible for this growth. Further research is needed to capture these variables in a time series setting.

Important caution must be taken in interpreting the above results however. First, since this study uses only data from two periods, it may not properly show the trend in cereal yield over the decade, but at best a snapshot. Second, it may not represent the national cereal yield status as the sample considers only 18 villages. However, this latter caveat might not be a serious one. These villages are located spread all over the country and the national trend is broad and widespread; hence, the results can be first approximation.

References

Alemayehu S., Sinafikeh A. and Gerawork G. 2009. "Trends and Determinants of Cereal Productivity: Econometric Analysis of Nationally Representative Data." IFPRI, ESSP-II. Addis Ababa, Ethiopia

Cameroon, A. C., and Trivedi, P. K. 2005. *Microeconometrics: Methods and Applications*. Cambridge University Press.

CSA. 1998. Report on Farm Management Practices (Private Peasant Holdings, Meher Season), Ethiopian Agricultural Sample Survey (1997/98 (1990 E.C.)), Statistical Bulletin 193, Central Statistical Authority (CSA), Addis Ababa, Ethiopia.

________. 1999. Report on Area and Production for Major Crops, Ethiopian Agricultural Sample Survey (1998/99 (1991 E.C.)), Statistical Bulletin 200, Central Statistical Authority (CSA), Addis Ababa, Ethiopia.

________. 2000. Report on Area and Production for Major Crops, Ethiopian Agricultural Sample Survey (1999/2000 (1992 E.C.)), Statistical Bulletin 227, Central Statistical Authority (CSA), Addis Ababa, Ethiopia.

________. 2001. Report on Area and Production for Major Crops, Ethiopian Agricultural Sample Survey (2000/2001) (1993 E.C.)), Statistical Bulletin 245, Central Statistical Authority (CSA), Addis Ababa, Ethiopia.

________. 2005. Report on Farm Management Practices for Meher Season Crops, Ethiopian Agricultural Sample Survey (2004/05 (1997 E.C.)) – Volume III, Statistical Bulletin 331, Central Statistical Agency (CSA), Addis Ababa, Ethiopia.

________. 2006. Report on Area and Production of Crops, Ethiopian Agricultural Sample Survey (2005/06 (1998 E.C.)) – Volume I, Statistical Bulletin 361, Central Statistical Agency (CSA), Addis Ababa, Ethiopia.

________. 2007. Report on Area and Production for Major Crops, Ethiopian Agricultural Sample Survey (2006/2007) (1999 E.C.)), Statistical Bulletin 388, Central Statistical Authority (CSA), Addis Ababa, Ethiopia.

________. 2008. Report on Area and Production Crops, Agricultural Sample Survey 2007/2008 (2000 E.C.): Volume I -, Statistical Bulletin 417, Central Statistical Agency (CSA), Addis Ababa, Ethiopia.

________. 2009. Agricultural Sample Survey 2008/2009 (2000 E.C.): Volume I - Report on Area and Production Crops, Statistical Bulletin 446, Central Statistical Agency (CSA), Addis Ababa, Ethiopia.

________. 2009. Report on Farm Management Practices for Meher Season Crops, Ethiopian Agricultural Sample Survey (2008/09 (2001 E.C.))–Volume III, Statistical Bulletin 446, Central Statistical Agency (CSA), Addis Ababa, Ethiopia.

Daniel Z. and Larson, D. F. 2010. "Incomplete Markets and Fertilizer Use: Evidence from Ethiopia." The World Bank Development Research Group Agriculture and Rural Development Team.

Dercon, S. and Hill V. 2009. Growth from Agriculture in Ethiopia: Indentifying Key Constraints. DfID, UK, 2009.

Ethiopian Economics Association. 2009. *Report on the Ethiopian Economy*. Volume VII. 2007/08, Ethiopian Economics Association. Addis Ababa, Ethiopia

Fantu N. Bachewe. 2012. "Growth in Total Factor Productivity in the Ethiopian Agriculture Sector: Growth Accounting and Econometric Assessments of Sources of Growth." ESSP-II Working Paper, 2012 IFPRI-ESSP-II, Addis Ababa, Ethiopia.

Food and Agriculture Organization. FAO. Faostat.org accessed at January 2011

Gollin, D., Lee F, R, Scher, I., Schlemme, C. 2011. "Crop Production in Ethiopia: Assessing the Evidence for a Green Revolution." Department of Economics, Williams College; IGC Growth Week.

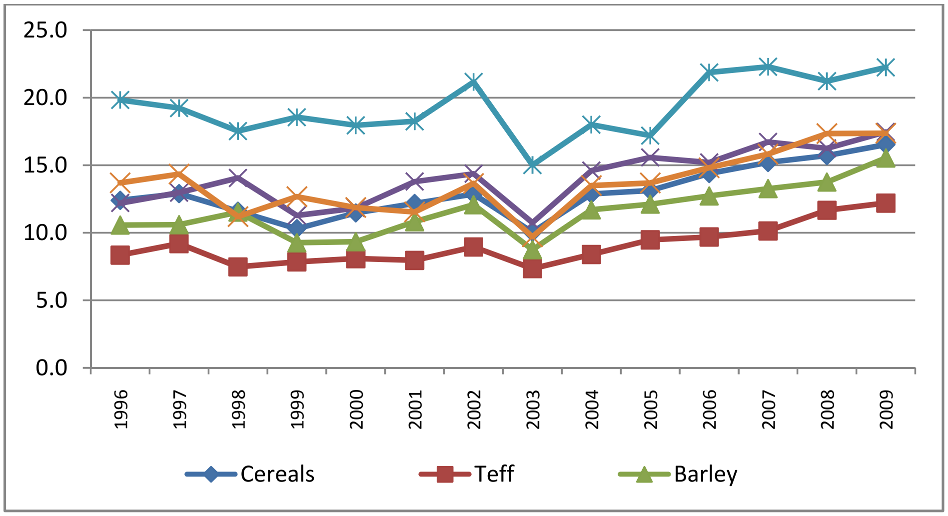
Mwangi, W. 1996. "Low Use of Fertilizers and Low Productivity in Sub-Saharan Africa." NRG Paper 96-05. Mexico, D.F.: CIMMYT.

Prabhu L. P. and Paul W. H. 1999. "Cereal Crop Productivity in Developing Countries: Past Trends and Future Prospects." CIMMYT, Working Paper 99-03.

Sing, I., Squire, L., Strauss, J. 1986. "A Survey of Agricultural Household Models: Recent Findings and Policy Implications." *The World Bank Economic Review*, Vol. 1 (Sept., 1986), pp. 149-179.

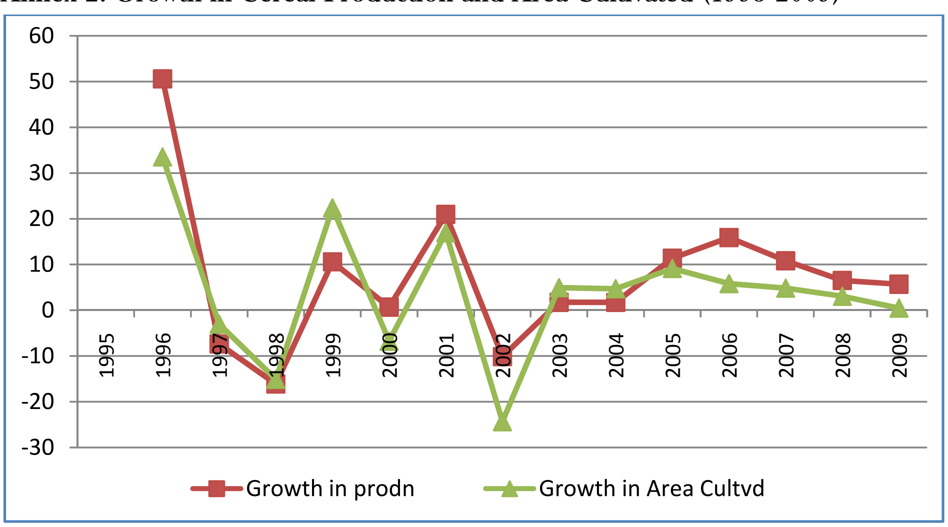
Annexes

Annex 1: Trends in Yields of Major Cereal Crops (1996-2009)



Source: Data from Central Statistical Agency (CSA), Different Agricultural Sample Surveys

Annex 2: Growth in Cereal Production and Area Cultivated (1995-2009)



Source: Data from Central Statistical Agency (CSA), Different Agricultural Sample Surveys

Annex 3: The Share of Area (in percent) under Improved Farm Management by Crop (1997/98-2008/09)

Improved Farm Management	Year	Cereals	Teff	Barley	Wheat	Maize	Sorghum
Share of Crop	1997/98	2.4	1.7	0.1	5.6	5.2	0.2
Area with	2004/05	4.0	0.7	0.5	3.8	15.9	0.5
Improved Seeds	2008/09	4.9	0.7	0.6	3.9	19.8	0.1
Share of Crop	1997/98	12.0	17.7	9.6	31.3	1.3	3.1
Area with	2004/05	16.7	24.5	11.7	38.0	2.1	2.4
Pesticide Application	2008/09	20.4	29.7	19.9	41.1	3.5	8.7
Share of Crop	1997/98	0.6	0.7	0.6	0.3	1.1	0.4
Area with	2004/05	0.8	0.4	0.7	0.3	2.1	0.8
Irrigation	2008/09	1.1	0.5	0.9	0.5	2.8	1.1
Share of Crop	2004/05	36.0	48.7	27.9	58.5	29.4	3.0
Area with Chemical Fertilizer	2008/09	36.5	51.9	27.8	60.1	31.0	3.0
Share of Crop	2004/05	18.2	17.2	16.7	29.5	24.9	6.5
Area with Extension Package	2008/09	15.6	16.2	8.8	20.5	23.0	2.9

Source: Authors' calculations based on the 1997/98, 2004/05 and 2008/09, Agricultural Sample Surveys, CSA.

Annex 4: Descriptive Statistics of Major Variables (2009)

Variables	Total Sample		Teff		Barley		Wheat		Maize		Sorghum	
	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.
Yield	1167.7	1038.9	822.6	747.7	968.8	885.4	1626.1	980.4	1447.5	1326.3	822.0	792.4
Improved Seeds in kg per hectare	12.7	52.7	3.4	16.6	1.9	24.1	45.8	101.5	6.3	18.7	0.0	0.0
Chemical Fertilizer in kg per hectare	79.8	150.3	97.7	215.4	80.7	102.2	103.5	103.7	56.3	145.9	7.7	50.7
Use of Traditional Fertilizer (Man/Comp=1)	0.2	0.4	0.0	0.2	0.2	0.4	0.1	0.3	0.4	0.5	0.1	0.3
Use of Irrigation (Yes=1)	0.0	0.2	0.0	0.1	0.0	0.2	0.0	0.2	0.0	0.2	0.0	0.2
Fertility of Land (Lem=1)	0.5	0.5	0.6	0.5	0.3	0.4	0.6	0.5	0.7	0.5	0.5	0.5
Fertility of Land (Lemteuf=1)	0.3	0.5	0.2	0.4	0.4	0.5	0.3	0.5	0.2	0.4	0.3	0.5
Fertility of Land (Teuf=1)	0.1	0.4	0.1	0.3	0.3	0.5	0.1	0.2	0.1	0.3	0.2	0.4
Slop of land (Medda=1)	0.8	0.4	0.8	0.4	0.8	0.4	0.8	0.4	0.8	0.4	0.8	0.4
Slop of land (Dagathama=1)	0.2	0.4	0.1	0.3	0.2	0.4	0.2	0.4	0.2	0.4	0.2	0.4
Slop of land (Geddel=1)	0.0	0.1	0.0	0.1	0.0	0.2	0.0	0.0	0.0	0.1	0.0	0.1
Age of holder (years)	52.4	14.0	52.2	13.1	55.5	14.1	51.5	13.5	50.8	14.7	49.7	13.7
Level of Education of holder												
Illiterate	0.4	0.5	0.4	0.5	0.4	0.5	0.4	0.5	0.4	0.5	0.4	0.5
Literate (church -mosque)	0.3	0.4	0.3	0.5	0.4	0.5	0.2	0.4	0.2	0.4	0.2	0.4
Primary (1-8)	0.3	0.5	0.3	0.5	0.2	0.4	0.3	0.5	0.3	0.5	0.3	0.5
Secondary (9-12)	0.0	0.2	0.0	0.2	0.0	0.1	0.1	0.3	0.0	0.2	0.0	0.1
Higher Education	0.0	0.1	0.0	0.1	0.0	0.0	0.0	0.1	0.0	0.1	0.0	0.1
Household size of holder	6.4	2.5	6.7	2.3	5.9	2.2	6.8	2.7	6.4	2.6	6.4	2.8
Gender of holder	0.7	0.4	0.8	0.4	0.6	0.5	0.7	0.4	0.8	0.4	0.8	0.4

Source: Estimated from ERHS (2009)

Annex 5: Descriptive Statistics of Major Variables (1999)

Variable	Total Sample		Teff		Barley		Wheat		Maize		Sorghum	
	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.
Yield	963.5	860.0	825.7	579.5	871.8	618.5	1006.3	697.8	1211.9	1373.3	1314.0	1399.5
Improved Seeds in kg per hectare	5.6	47.0	3.7	55.6	2.2	29.4	17.0	72.6	7.0	25.6	0.0	0.0
Chemical Fertilizer in kg per hectare	69.5	100.8	82.7	78.0	58.0	95.4	99.2	117.0	52.8	109.2	32.1	116.7
Traditional Fertilizer (Man/Comp=1)	0.2	0.4	0.0	0.2	0.2	0.4	0.1	0.3	0.4	0.5	0.4	0.5
Use of Irrigation (Yes=1)	0.0	0.2	0.0	0.1	0.0	0.1	0.0	0.1	0.2	0.4	0.0	0.2
Extension Coverage (Yes=1)	1.9	0.2	2.0	0.2	2.0	0.2	1.9	0.3	1.8	0.4	2.0	0.2
Fertility of Land (Lem=1)	0.5	0.5	0.6	0.5	0.4	0.5	0.5	0.5	0.7	0.5	0.4	0.5
Fertility of Land (Lemteuf=1)	0.3	0.5	0.3	0.4	0.3	0.5	0.3	0.5	0.3	0.4	0.4	0.5
Fertility of Land (Teuf=1)	0.2	0.4	0.1	0.3	0.3	0.5	0.1	0.3	0.1	0.2	0.2	0.4
Slop of land (Medda=1)	0.8	0.4	0.9	0.2	0.8	0.4	0.8	0.4	0.8	0.4	0.7	0.5
Slop of land (Dagathama=1)	0.1	0.3	0.1	0.2	0.2	0.4	0.1	0.4	0.2	0.4	0.3	0.4
Slop of land (Geddel=1)	0.0	0.1	0.0	0.1	0.0	0.2	0.0	0.1	0.0	0.1	0.0	0.1
Age of holder (years)	49.4	14.9	47.5	14.4	51.8	15.1	50.0	15.6	47.6	13.9	47.8	14.2
Gender of holder	1.2	0.4	1.1	0.3	1.2	0.4	1.2	0.4	1.2	0.4	1.1	0.3
Level of Education												
Illiterate	0.4	0.5	0.3	0.5	0.5	0.5	0.4	0.5	0.3	0.4	0.5	0.5
Literate (church-mosque)	0.2	0.4	0.1	0.3	0.2	0.4	0.1	0.3	0.1	0.3	0.3	0.4
Primary (1-8)	0.1	0.3	0.2	0.4	0.1	0.3	0.2	0.4	0.2	0.4	0.1	0.3
Secondary (9-12)	0.0	0.1	0.0	0.1	0.0	0.0	0.0	0.1	0.0	0.2	0.0	0.1
Higher Education	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.1	0.0	0.0

Source: Estimated from ERHS (1999)

Annex 6: Descriptive Statistics of Major Variables (1999 & 2009)-Panel

Variable		Mean	Std. Dev.	Min	Max	Observations
Yield	overall	1031.4	914.8	0.0	8333.3	N = 1655
	between		911.4	0.0	8333.3	n = 1156
	within		382.8	-1669.8	3732.6	T-bar = 1.431
Dummy for improved seeds	overall	0.12	0.33	0.00	1.00	N = 1668
	between		0.32	0.00	1.00	n = 1161
	within		0.14	-0.38	0.62	T-bar = 1.4366
Improved seeds used in kg per hectare	overall	7.6	40.7	0.0	1000.0	N = 1665
	between		42.2	0.0	1000.0	n = 1160
	within		16.6	-292.4	307.6	T-bar = 1.4353
Fertilizer application per hectare	overall	67.0	148.4	0.0	3400.0	N = 1665
	between		166.3	0.0	3400.0	n = 1160
	within		35.8	-308.0	442.0	T-bar = 1.4353
Square of chemical fertilizer used in kg	overall	26506.2	360610.3	0.0	11600000.0	N = 1665
	between		430843.3	0.0	11600000.0	n = 1160
	within		20212.2	-366018.0	419030.4	T-bar = 1.4353
Interaction term: Fertilizer & Improved seeds (dummy)	overall	13.3	54.7	0.0	1250.0	N = 1665
	between		55.8	0.0	1250.0	n = 1160
	within		23.2	-386.7	413.3	T-bar = 1.4353
Interaction term: Fertilizer & Improved seeds (continuous var)	overall	1702.2	32958.1	0.0	1250000.0	N = 1665
	between		37516.0	0.0	1250000.0	n = 1160
	within		8395.7	-238298	241702.2	T-bar = 1.4353

Variable		Mean	Std. Dev.	Min	Max	Observations
Irrigation: 1 if Part or all of plot irrigated, 0 otherwise	overall	0.06	0.24	0.00	1.00	N = 1668
	between		0.22	0.00	1.00	n = 1161
	within		0.12	-0.44	0.56	T-bar = 1.4366
Use of traditional fertilizer	overall	0.32	0.47	0.00	1.00	N = 1668
	between		0.42	0.00	1.00	n = 1161
	within		0.24	-0.18	0.82	T-bar = 1.4366
Interaction term between fertilizer and Irrigation	overall	3.20	22.63	0.00	400.00	N = 1665
	between		18.99	0.00	214.86	n = 1160
	within		13.26	-196.80	203.20	T-bar = 1.4353
Period dummy2009	overall	0.50	0.50	0.00	1.00	N = 2322
	between		0.00	0.50	0.50	n = 1161
	within		0.50	0.00	1.00	T = 2

Source: Estimated from ERHS (1999 & 2009)

ILLNESS AND CHOICE OF TREATMENT IN URBAN AND RURAL ETHIOPIA

Shiferaw Gurmu[1] and Solomon Tesfay Tesfu[2]

Abstract

Using large data set from a nationally representative sample of households and discrete choice models, we examine the effect of access to roads, transport and liquidity on seeking treatment for illness and health care provider choice in urban and rural Ethiopia. The results indicate that access to roads and public transport are important determinants of the decision to seek treatment for illness by the rural residents. We also find evidence that distance to all weather road, access to public transport and access to liquidity have a strong effect on the utilization of private healthcare facilities. The significance of distance to all weather roads in healthcare utilization is especially appealing in terms of policy design because it implies that construction of multi-purpose road networks can compensate for the absence of healthcare facilities in the proximity.

Keywords: Illness; healthcare choice; accessibility; liquidity constraint; discrete choice models; sample selection

JEL Classification: I1; C35

[1] Corresponding author: Department of Economics, P.O. Box 3992, Andrew Young School of Policy Studies, Georgia State University, Atlanta, GA 30302, USA; Phone: (404) 413-0161; Fax: (404) 413-0145; E-mail: sgurmu@gsu.edu

[2] Richard J. Bolte, Sr. School of Business, Mount St. Mary's University, 16300 Old Emmitsburg Road, Emmitsburg, MD 21727. Phone: (703) 298-7481; E-mail: tesfu@msmary.edu

1. Introduction

People in developed countries often visit healthcare facilities not only when they are sick but also for regular checkups. In developing countries on the other hand, many people do not seek treatment in the formal medical facilities even when they are sick let alone for routine check-ups. While access and affordability could be important factors influencing healthcare decisions in developing countries, awareness and socio-cultural beliefs could as well be important. Access is often limited because of poor infrastructure and limited distribution of healthcare facilities. Private healthcare facilities are often limited to urban areas that may not be easily accessible to the rural population because of poor road and transport facilities. Government healthcare facilities that provide low-cost healthcare services are often available in the rural areas but their distribution is usually uneven and their quality is low (Chaya, 2007). Traditional healthcare services are also widely practiced in a number of developing countries, particularly in the rural areas. For example, WHO (2008) estimates that up to 80% of the people in some Asian and African countries rely on traditional medicine for primary healthcare. When traditional and modern healthcare services coexist side by side, therefore, efforts to expand modern healthcare may not always produce the desired result. Understanding the factors that influence people's choices between alternative sources of healthcare service will thus be useful inputs into the process of developing effective healthcare services in developing countries.

Some recent empirical studies examined the influence of both the demand and supply side factors on healthcare choices in developing countries (e.g., Amaghionyeodiwe, 2008; Sepehri, 2008; Habtom and Ruys, 2007). The demand side factors include income, patient and household characteristics while the supply side factors include various measures of physical access and quality of healthcare providers. The most commonly analyzed measure of physical access to healthcare is distance to healthcare facilities. While physical access could also be influenced by the availability of transport facilities, the direct effect of such facilities has attracted little attention in empirical analysis of healthcare choice in developing countries. Easier access to road networks and public transport allows more convenient and quicker travel to towns or cities where more or better treatment options are often available and hence may influence healthcare decisions.

Using large survey data from a nationally representative sample of households, this paper examines the effect of access to road networks, means of transportation and liquidity[3] status on seeking treatment for illness and health care provider choice in rural and urban Ethiopia. We also analyze the effects of distance to health centers and the standard household and individual characteristics like income, asset ownership and education on choice of treatment for illness. We estimate selection probit models for seeking treatment for illness and multinomial logit models for healthcare provider choice, allowing for household level cluster-robust standard errors.

This paper contributes to the literature on the determinants of healthcare usage in developing countries by examining the direct effects of access to transport networks and liquidity that have not been investigated in existing literature in developing counties. Although healthcare services in the public facilities in Ethiopia as in many other developing countries are mostly provided free of charge or at nominal costs, getting there and financing food and accommodation for self and company require some money and hence household's access to liquidity at the time a member gets sick is important. However, to the best of our knowledge, there are no studies that examine the direct effect of liquidity constraints on healthcare utilization in developing countries.

Our empirical results indicate that access to roads and public transport are important determinants of the decision to seek treatment for illness by the rural residents. We also find evidence that distance to all weather road, access to public transport and access to liquidity have a strong effect on the utilization of private healthcare facilities. The significance of distance to all weather roads in healthcare utilization has important policy implication because it suggests that it may be possible to improve healthcare utilization not only by expanding healthcare facilities in the rural areas but also by building more roads. This may help the resource-constrained policy makers to reprioritize the budget so as to focus more on improving the quality of healthcare facilities and expansion of road networks (since roads are needed for non-health aspects of well-being as well) by scaling back the construction of new healthcare facilities.

The next section briefly describes the healthcare system in Ethiopia. That is followed by a brief review of the literature on healthcare utilization and provider choice focusing on the research in developing countries. Section 4 briefly describes the theoretical context and the empirical methodology followed in this paper. In section 5 we describe the data and present summary statistics for the

[3] Household's access to liquidity is defined in terms of the household's ability to obtain 100 Birr within a week if needed. Birr is the name of the Ethiopian currency and 1USD currently buys about 18 Birr.

variables used in the empirical models. Estimation results are presented in section 6 while section 7 concludes.

2. Health Care System of Ethiopia

With annual per capita expenditure on healthcare in PPP of about $23.00 (WHO, 2012), Ethiopia is one of the countries that spend the least on healthcare services in Africa. The average healthcare expenditure in comparable countries like Kenya and Ghana is almost five times higher. The meager resources allocated for the healthcare services are also reflected in the poor state of the healthcare infrastructure in the country. According to the Ministry of Health of Ethiopia (2010), the healthcare infrastructure in Ethiopia[4] consisted of 195 hospitals, 13,850 health centers, and health posts[5], 2,853 clinics and 1,322 pharmacies and drug shops in 2009[6]. The total number of beds in the hospitals was 15,111 for a hospital bed-population ratio of less than 2 beds per 10,000 which is among the lowest in the world.

The public sector dominates the provision of health care services in the country with nearly 75% of the hospitals being run by the ministry of health or other government agencies like the ministry of defense (MOH, 2010). About 20% of the hospitals are privately owned while the remaining 5% are run by NGOs. Health centers, health stations and health posts are almost exclusively owned by the government while the clinics are owned either by the private sector (90%) or NGOs (10%). The public health care facilities provide healthcare services at nominal prices or free of charge depending on the economic status of the patient. Most of the NGO facilities also provide subsidized health care services while the private facilities charge the market prices.

In terms of location, almost all the hospitals in the country are in the big cities and towns, 21% being in the capital city (MOH, 2010). Health centers are available both in small and large towns as well as some rural areas but their distribution and quality is uneven. Health posts are almost exclusively rural health outlets but mostly lack the necessary human resources and supplies to provide

[4] Based on the projections from the latest population census conducted in May 2007, Ethiopia's population would be about 84 million in July, 2012 (CSA, 2011).
[5] The health centers have more services and human resources than health posts. Healthcare facilities which used to be reported as "health stations" have been phased out or upgraded to health centers or downgraded to health posts as of 2009 (MOH, 2010).
[6] This is the latest year for which data on the number of both private and public healthcare facilities and their personnel are available.

anything more meaningful than referring the visitors (patients) to the nearest health center. The private clinics are mostly located in the urban areas where they can find people who can afford the higher prices that they often charge.

One of the reasons for the very low per capita healthcare utilization rate in Ethiopia (see MOH, 2008, Collier et al, 2003) could be these supply side constraints as reflected in the limited availability and uneven distribution of the healthcare facilities. While 85% of the country's population lives in rural areas most of the healthcare facilities are located in urban areas. The road networks and transport facilities that connect the rural areas to the towns and cities are underdeveloped and often deteriorate during the rainy season (Chaya, 2007). The distribution of these transport facilities is also uneven leading to people living in or near the towns having a comparative advantage over those who live in the remote areas.

Perhaps believing that distance is an important factor affecting the healthcare utilization rate in the country, the focus of the government's health care policy over the recent years has been to expand the physical access to health facilities (Collier et.al, 2003) largely by building large number of health posts as close to the people as possible. For example, the number of health posts has expanded from 1,432 in 2003 to 12,488 in 2009 while the number of health centers increased from 451 to 1,362 over the same period (see MOH 2008 and 2010). The government has also been constructing some rural road networks but the overwhelming emphasis on expanding the physical access could potentially be at the expense of quality as some evidence appears to indicate (see for example USAID, 2008). Now the question is can't the government scale back the construction of the new healthcare facilities and invest more in the quality of the existing facilities and more road networks that serve not only the healthcare needs of the public but also other aspects of their livelihood like access to markets? Answering this question requires analyzing the role of road networks in healthcare utilization of the public, and that is one of the key purposes of this study.

3. Literature

Several studies have examined the factors that influence healthcare choices both in developed and developing countries. The most common determinants of healthcare utilization can broadly be classified into the demand-side factors and supply-side factors. According to Ensor and Cooper (2004) the demand-side factors include income, education and awareness, attitudes and norms, opportunity costs of seeking healthcare, availability of substitutes and their prices, type and severity of illness, and various socio-demographic characteristics of the individual, the household and the community. On the other hand the supply-side factors include various measures of availability and quality such as distance to healthcare facilities, travel time, waiting time, number and qualification of the staff, availability of equipment and medical supplies. Other factors like price of healthcare service and rationing reflect the interaction between demand and supply sides. The role of each of these has been empirically examined in one form or another.

Demand side factors like access to and type of health insurance, health status, income and education are the dominant determinants of healthcare utilization in the United States where healthcare is largely financed through private or group health insurance schemes (e.g., see Sharma et.al, 2003; Manning et al, 1987; Leclere, 1994). Geographic and spatial factors are also important in the rural areas of the US (for example see Arcury et al, 2005; Nemet and Bailey, 2000). In other developed countries, where healthcare is largely financed through public health care schemes like the UK or Germany, healthcare utilization is also influenced by the supply side factors like access to and nature of transport facilities as well as distance to the healthcare facilities (See for example Field and Briggs, 2001; Pohlmeier and Ulrich, 1995). According to Ensor and Cooper (2004), demand side barriers to healthcare utilization in developed countries are particularly important for vulnerable groups like the poor and the elderly.

Explaining healthcare utilization behavior appears to be more complex in developing countries than in the developed countries. The much more pervasive use of self-care and traditional medicine, the still evolving process of modern medicine, the more complicated cultural attitudes and socio-economic conditions make it harder to accurately model healthcare utilization behavior in these countries (Kroeger, 1983). Yet, a large number of studies, with varying levels of methodological rigor and area of emphasis, have examined the determinants of healthcare utilization in developing countries. While some of these studies analyzed healthcare utilization as a binary outcome variable capturing as to whether healthcare service was sought or not (e.g. Mwabu et al 1993; Develay et al, 1996; Appleton 1998; Collier et al 2003, Pillai et al, 2003; Sepehri et al 2008), others analyzed the frequency of visits to a doctor or medical facilities (e.g. Trivedi 2002; Chang and Trivedi, 2003). Even a larger number of studies have examined the factors that influence the choice of healthcare provider in developing countries (e.g. Amaghionyeodiwe 2008; Habtom and Ruys 2007;

Thuan et al 2008; Leonard 2007; Borah 2006; Akin et al 1998; e.g. Bolduc et al 1996; Tembon 1996).

Distance to healthcare facilities and income are the most commonly analyzed determinants of healthcare utilization in developing countries. In an African context, the effect of distance on healthcare usage was analyzed, for example, by Appleton (1998) and Buor (2003) for Kenya and Ghana, respectively. Both studies found out that distance has a strong negative effect on healthcare utilization. Another study by Collier et al (2003) found a similar negative association between distance and usage in a sample of rural districts in Ethiopia. Collier et al as well as Buor also found positive association between healthcare usage and household income. Some of these studies also show that education may positively influences healthcare utilization in developing countries (Collier et al, 2003; Appleton, 1998). The effects of user fees and transport cost on utilization of healthcare were analyzed by Develay et al (1996) but found theoretically implausible positive effects for both variables as did Buor (2003) for waiting time and transport cost in Ghana. According to Collier et al as well as Mwabu et al (1993) quality of healthcare services is another important factor that positively influences health care usage.

A number of other studies examined the effects of income, distance, education, user fees, waiting time, travel time, and quality of service on the choice of healthcare provider in Africa. For example, Habtom and Ruys (2007) found household income to have significant positive effect on the utilization of private medical facilities for profit and no significant effect on the utilization of public or non-profit catholic healthcare facilities in Eritrea. Their study also found education to have positive effect on utilization of both private and public facilities while user fees and transport costs had negative effect. Similar findings were obtained for the effect of user fees and education by Asfaw (2003) in a rural Ethiopian village. Another study by Amaghionyeodiwe (2008) showed negative relationship between distance and user fees and utilization of both public and private clinics and hospitals in Nigeria. Quality was found to have a positive influence on utilization of both private and public facilities by Amaghionyeodiwe (2008) as well as Akin et al (1995) for Nigeria. Similar findings were reported for the effect of quality and distance on provider choice by Bolduc (1996) for Benin, Leonard (2007) for Tanzania and Tembon (1996) for Cameron.

While almost all of the studies of healthcare utilization in developing countries essentially analyzed healthcare usage by the people with some kind of self-reported illnesses, only a few of them recognized and attempted to address the selectivity problem that could arise due to the self-reported nature of the illness and treatment data. Selectivity may arise because those people who are more likely to report illness may actually end up being those who care and know more about their health status and hence are more likely to seek treatment. Therefore, parameters estimates based on the data from the sample with self-reported illnesses may be biased. In the context of a binary healthcare utilization model, the selectivity problem was addressed by Appleton (1998) and Collier et al (2003). Appleton finds selectivity to be unimportant in a binary choice probit model for seeking treatment in Kenya while Collier et al find it to be significant in a data from rural Ethiopia.

This paper attempts to expand the literature in this area by estimating selectivity-corrected binary choice model for healthcare usage and multinomial logit model for provider choice. The focus of analysis is on the effect of access to transport networks and access to liquidity on healthcare usage that has not been directly examined in the existing literature in developing countries. Access to transport facilities is represented by distance to all weather roads and the available means of traveling to the healthcare center. Access to liquidity is captured by the ability to raise 100 Birr within a week if needed. In the absence of formal credit markets for personal finances in Ethiopia, people will have to rely on their friends or social networks to raise the money that may be needed for emergency and understanding the role of liquidity constraints in healthcare utilization may indirectly explain the degree to which the informal networks are able to cover the role of modern credit facilities.

4. Theoretical Context and Methods

The theoretical framework for healthcare utilization has been formulated by a number of studies before as summarized for example in Behrman and Deolalakiar (1988). Considering a health care system consisting of j alternative providers, an individual faces $j+1$ alternatives including self-care (not seeking outside treatment) as $j+1^{th}$ option. Hence, the utility individual i expects to derive from choosing treatment alternative j may be stated as

$$u_{ij} = u_{ij}(c_{ij}, h_{ij}), \tag{1}$$

where h_{ij} is the health status of individual i after receiving treatment from alternative j and c_{ij} is a composite good representing the value of all goods and services consumed other than healthcare. The health status for individual i, h_{ij}, is assumed to depend on initial health status h_0, a vector of individual characteristics (x_i) like sex, age and education as well as the vector of provider characteristics (z_{ij}) like accessibility, affordability and quality of services it provides as defined for example by Mwabu et al (1993),

$$h_{ij} = h_{ij}(h_0, x_i, z_{ij}). \tag{2}$$

The budget constraint may also be stated as

$$y_i = c_{ij} + p_{ij}, \tag{3}$$

where p_{ij} represents both the direct costs of transport and treatment as well as opportunity costs of time spent in seeking treatment at provider j.

Maximizing (1) subject to (2) and (3) and substituting the resulting reduced form demand functions for h_{ij} and c_{ij} back into the conditional utility function leads to the standard conditional indirect utility function that may be stated as

$$v_{ij} = v_{ij}(h_0, x_i, z_{ij}, y_i, p_{ij}). \tag{4}$$

Equation (4) represents the maximum utility individual i can derive by seeking treatment at provider j, controlling for initial health status h_0, a vector of individual characteristics x_i, provider characteristics z_{ij}, income y_i, and direct and opportunity costs of treatment p_{ij}. In practice, however, some of the factors that influence utility may not be observable and restating the utility function by adding these random components (ε_{ij}) to the indirect utility provides the basis for the empirical model:

$$U_{ij} = v_{ij}(.) + \varepsilon_{ij}\ . \tag{5}$$

Denoting healthcare provider alternatives available for individual i by d_i, alternative j will be chosen over alternative k if j provides at least as much utility as k, i.e. the probability that j is chosen over k is given as,

$$\begin{aligned} P(d_i = j) &= P[U_{ij} \geq U_{ik}],\ \text{for all}\ j \neq k. \\ &= P[v_{ij}(.) + \varepsilon_{ij} \geq v_{ik}(.) + \varepsilon_{ik}],\ \text{for all}\ j \neq k. \\ &= P[\varepsilon_{ik} - \varepsilon_{ij} \leq v_{ij} - v_{ik}],\ \text{for all}\ j \neq k. \end{aligned} \tag{6}$$

Assuming that the errors are distributed as type I extreme value and v_{ij} is linear in its arguments, the probabilities in (6) can be modeled as multinomial logit.

For empirical analysis in this paper, the alternative sources of healthcare service are classified into five categories consisting of self-treatment, government facilities, private facilities (including NGO facilities), pharmacy, and other facilities that include traditional healers. Self-treatment (no formal treatment) is taken as the benchmark option for which the vector of coefficients is normalized to zero. Among the arguments in v_{ij} our analysis focuses on two of the access variables: distance to all weather roads and availability of public transport, and one of the household characteristics: an indicator for liquidity constraint or access to finance. We expect distance to all weather roads to have negative influence on the utilization of all healthcare facilities but for the rural sample it is expected to have larger negative effect on the utilization of private healthcare facilities that are mostly located in larger cities and hence require road networks to attract rural residents. Similarly, availability of public transport is expected to have positive influence on the utilization of the healthcare facilities its effect being larger for the usage of private facilities by the rural residents. The indicator for access to liquidity that takes a value of 1 if the household can obtain at least 100 Birr within a week if it wants and 0 otherwise is expected to have positive influence on utilization of all the facilities but intuitively its effect is anticipated to be larger on the usage of the private healthcare facilities that mostly charge the market price as opposed to the government facilities that provide the services at nominal prices or free of charge.

Our healthcare provider choice models also include some of the most commonly analyzed determinants of healthcare usage such as distance to health centers, individual and household characteristics such as age, gender, marital status, education, employment status, occupation, income, and assets owned. We estimate our models both with and without region fixed effects to see if our results are influenced by some unobserved differences in the access and quality of services across the administrative regions of the country. We also correct the standard errors for the household level clustering since the household characteristics included in the econometric models are similar for the members of the same household.

The binary choice probit model of healthcare utilization we estimate in the first part of our econometric analysis can be thought of as a special case where all the four sources of formal health care are lumped together against the self-care option. The dependent variable in this case is a dichotomous variable that takes a value of 1 if the person chose to seek outside treatment in any of public or private facilities, pharmacy or other facilities. The regressors in the probit equations are similar to those that are included in the multinomial healthcare provider choice models. To account for selectivity bias that could arise as a result of common unobserved influences on the tendency to report an illness and seek treatment, we jointly estimate the treatment equation along with a binary outcome equation for illness specified as probit.[7] The illness equation includes some additional regressors like sources of drinking water, type of toilet facilities, sources of cooking fuel and lighting that are thought to influence the health status of household members. The data and variables used for estimation of the empirical models are described in the next section.

5. Data

Analysis in this paper is based on data from the Ethiopian Welfare Monitoring Survey (WMS) conducted in the year 2000 by the central statistical agency of Ethiopia (henceforth WMS2000). WMS2000 was a nationally representative survey of 17,285 households in the rural areas and 8,643 households in urban areas proportionately distributed across the 11 regions of the country on the basis of their population size. The rural sample includes about 83,390 individuals with complete information while the urban sample includes 38,931 individuals with complete data. The data contain reasonably detailed information on the characteristics of the household and its members as well the community and the environment they live in. The data also contain information on incidence of illness and type of treatments sought, household income and key assets owned, education and basic demographics for each member of the household, employment status and occupation for each member older than 9, type of dwelling and type of access to water, energy and toilet facilities, distance to health center, distance to dry and all weather roads and type of transport to health centers as well as an indicator for access to finance. The names and description of the variables of interest are presented in Table 1.

[7] See, for example, Wooldridge (2010), Chapters 15-16 and 19, for further details on multinomial logit model and binary probit regression with selection.

Table 1: Names and descriptions of the variables used in the models

Variable	Description
Dependent variables (Last 2 Months)	
Illness in last 2 months	= 1 if person had health problem in last 2 months
Seek treatment	= 1 if person sought treatment in last 2 months (if ill)
Health care choice(for persons who reported health problems over the last two months)[a]	
Self-treatment (benchmark)	= 1 if person did not seek outside treatment
Government facilities	= 1 if person used health care services provided by government
Private facilities	= 1 if person used health care service provided by private/mission/NGO facilities
Pharmacy	= 1 if person was treated at Pharmacy
Other facilities	= 1 if person used other sources of treatment
Explanatory variables	
Distance to health center[a]	Distance to health center in kilometers
Distance to all weather road[a]	Distance to all weather road (AWR) in kilometers
Age	Age in years
Male	= 1 if person is male
Household size[a]	Household size
Married	= 1 if person is currently married
Primary education	= 1 if education of person is at least primary complete
Access to liquidity[a]	= 1 if household can get 100 Birr within a week
House ownership[a]	= 1 if household owned a house a year ago
Low income (Benchmark)[a]	= 1 if monthly income is from 0 to 80 Birr
Lower middle income[a]	= 1 if monthly income is from 81 to 180 Birr
Middle income[a]	= 1 if monthly income is from 181 to 395 Birr
High income[a]	= 1 if monthly income is from 396 to 9100 Birr
Number of cattle[a]	Number of cattle owned
Number of equines[a]	Number of Equines owned
Number of sheep[a]	Number of sheep/goats owned
Occupation (differential groups for urban vs. rural)	
Service and sales	= 1 if occupation in service and sales
Agriculture	= 1 if occupation is in agriculture
Craft and trade	= 1 if occupation in craft and trade
Plant and machine	= 1 if occupation in plant and machine
Laborer	= 1 if the person is laborer
Other occupation – rural	= 1 if occupation is other (for rural resident)
Other occupation – urban	= 1 if occupation is other(urban resident)
Young	= 1 if person is younger than 10
Unemployed	= 1 if person is unemployed

Variable	Description
Means of transport to health center	
On foot[a]	= 1 if transportation to health center is on foot
Public transport[a]	= 1 if transport to health center is public
Animal and other transport[a]	= 1 if transport to health center is on animals or other
Water, toilet, lighting and cooking	
Unprotected water[a]	= 1 if household uses unprotected water
Flush toilet[a]	= 1 if household uses flush toilet
Pit-latrine toilet[a]	= 1 if household uses pit-latrine toilet
Other (control)[a]	= 1 for other toilets
Number of rooms[a]	Number of rooms in the house
Kerosene[a]	= 1 if lighting source is kerosene
Electricity[a]	= 1 if lighting source is electricity
Other lighting[a]	= 1 for other sources of lighting
Natural fuel[a]	= 1 if household uses natural fuel for cooking
Other fuel[a]	= 1 if household uses other sources of fuel for cooking
Living standard[a]	= 1 if self-assessed general living standard decreased over last year
Fixed effects	
Killil fixed effects[a]	Dummy variables for 11 major regions

Source: Welfare Monitoring Survey (WMS)- Ethiopia (2000).

[a] Indicates household level variables.

As stated in the previous section, the illness equation and treatment equations are jointly estimated to control for selectivity problem. WMS2000 gathered detailed information on illness and treatment for a period of 2 months preceding the date of interview. The incidence of illness by region, age group, income and gender is presented in Table 2. About 29% of the individuals in the rural sample and about 22% in the urban sample reported at least one incidence of illness over the period. Prevalence of illness was the highest among the elderly (60 or older) about 47.5% of them reporting to have been sick at least once over the last two months. Incidence of illness was also high among the under-five children 35% of which were reported to have been sick over the two months' period. The incidence of illness also appears to increase with the falling income, 30% of the individuals belonging to the lowest income households reporting some kind of illness whereas only 22% of those who belong to the top income class having suffered some illness over the previous two months. In addition, women appear to be slightly more vulnerable to disease than men.

Table 2: Health problems by age, gender and income group

Rural vs. urban	
Rural	29.0
Urban	21.6
By age group	
Age 0 to 4	35.38
Age 5 to 16	17.08
Age 16 to 59	28.24
Age 60 and over	47.52
By income class	
Low Income	29.75
Low Middle Income	28.66
Middle Income	25.43
High Income	21.78
By gender	
Female	27.89
Male	25.38
Total	26.68

Source: Computed from WMS-Ethiopia (2000).

There is a marked difference in utilization of health care between urban and rural areas as shown in Table 3. Among those who reported illness in the rural Ethiopia, almost 60% did not seek any kind of formal treatment whereas only 30% of those who reported an illness in urban areas did not seek outside treatment. This appears to be quite high even by African standards. For example, only 7 to 8% of those who reported illness in a Nigerian sample failed to seek outside treatment (Amaghionyeodiwe, 2008). Among those who sought treatment nearly half visited government facilities both in urban and rural areas but utilization of private facilities is proportionately higher in urban areas. About 17% of those who sought treatment from the rural areas visited private facilities while 26% of the healthcare users from the urban areas visited private providers. This is not surprising since the private facilities are mostly available in urban areas and the urban residents are at a comparative advantage in terms of location and perhaps ability to afford the higher prices that the private facilities often charge. On the other hand, reliance on pharmacy and other sources of healthcare is proportionately higher among the rural users than urban users (32% versus 20%). The ‘other’ category mostly includes the traditional healers that are more prevalent in the rural areas than the urban areas. However, on the basis of these data (some of which could be reporting error) the overall reliance on traditional healers doesn't seem to be as high as generally presumed in the context of developing countries as cited for example in WHO (2008).

Table 3: Health care facility choice of individuals reporting health problems

Facility choice	Rural	Urban
Self-Care	59.93	30.41
Government	20.42	37.48
Private	6.75	18.34
Pharmacy	6.02	7.50
Other	6.87	6.26
Total (%)	100	100
Total (Obs.)	24205	8428

Source: Computed from WMS-Ethiopia (2000).

The summary statistics for the rest of the variables used in both the healthcare utilization and illness equations are presented in Table 4. As stated before our key variables of interest are access to roads and means of transport to healthcare facilities. The WMS2000 data contain information on the distance of the household unit in the sample from the nearest dry weather road and nearest all-weather road both measured in kilometers (kms). However, we think that all-weather roads are better measures of access to transport than dry weather roads since the later often become dysfunctional during the rainy season which could run for several months in some parts of the country. Therefore, we include distance to all-weather roads as a regressor in healthcare utilization models. In the rural areas represented by the sample of households in the WMS2000 data, the average distance from all weather roads is 12.5kms with considerable variation as reflected in the large standard deviation. Not surprisingly, the mean distance from all weather roads in urban areas is much smaller (less than half a kilometer) but there is a large variation in the urban areas as well. The average distance to the nearest health center is about 9kms in the rural areas and 1.5kms in the urban areas again with considerable variation.

Table 4: Summary statistics for the variables used in the econometric models

	Rural (Obs.=83390)		Urban (Obs.=38931)	
Variable	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**
Illness in last 2 months	0.290	0.454	0.216	0.412
Seek treatment[a]	0.401	0.490	0.696	0.460
Distance to health center	8.977	9.899	1.465	3.349
Distance to all wealth. road	12.537	16.699	0.391	3.295
Public transport	0.035	0.184	0.077	0.267
Animal and other transport	0.138	0.345	0.166	0.372
Transport on foot	0.827	0.380	0.756	0.429
Access to liquidity	0.707	0.455	0.686	0.464
Age	21.082	18.375	23.164	17.480
Age squared	782.097	1219.64	842.082	1209.93
Male	0.497	0.500	0.450	0.498
Household size	5.924	2.283	5.930	2.696
Married	0.325	0.468	0.258	0.437
Married*age	12.238	19.417	9.992	18.226
Primary education	0.032	0.177	0.360	0.480
House ownership	0.938	0.242	0.542	0.498
Low income (Benchmark)	0.251	0.434	0.152	0.359
Lower middle income	0.204	0.403	0.288	0.453
Middle income	0.128	0.334	0.451	0.498
High income	0.097	0.295	0.043	0.203
Number of cattle	4.400	9.173	1.265	16.877
Number of equines	1.047	13.837	0.333	10.535
Number of sheep	3.818	12.352	1.002	10.845
Service and sales			0.087	0.281
Agriculture	0.189	0.391		
Craft and trade			0.064	0.245
Plant and machine			0.012	0.110
Laborer	0.206	0.404	0.077	0.267
Young	0.343	0.394	0.230	0.421
Unemployed	0.227	0.419	0.459	0.498
Other occupation - rural	0.035	0.184		
Other occupation - urban			0.071	0.257
Unprotected water	0.798	0.402	0.065	0.247
Flush toilet	0.007	0.085	0.086	0.280
Pit-latrine toilet	0.108	0.311	0.699	0.459
Other (control)	0.884	0.320	0.215	0.411
Number of rooms	1.595	0.939	2.558	2.060
Kerosene	0.703	0.457	0.187	0.390
Electricity	0.015	0.120	0.789	0.408
Other lighting	0.282	0.450	0.024	0.154
Natural fuel	0.961	0.193	0.591	0.492
Other fuel	0.039	0.193	0.409	0.492
Living standard	0.413	0.492	0.318	0.466

Source: Computed from WMS-Ethiopia (2000).

[a] There are 24205 and 8428 observations for the rural and urban samples, respectively.

The other indicator of access to transport facilities whose effect we examine in our empirical models for healthcare utilization is the means of transport the household members usually use to travel to the healthcare centers. Apparently, the most common means of transport to the nearest health center is on foot almost 83% of the rural and 76% of the urban respondents stating it as the available means of transport. About 4% of the rural and 8% of the urban households stated that they use public transport to travel to the nearest health center. Other means of transport including animals and perhaps some private cars account for 14% of the rural and 17% of the urban means of traveling to a health center.

Another variable we focus on in our econometric analysis is an indicator for access to liquidity at the household level capturing the response to a question asking whether the household can raise at least 100 Birr within a week if needed. The proportion of households that positively responded to this question is roughly equal in rural and urban areas (71% for rural vs. 69% for urban) which is somewhat surprising given the general observation that financial services in developing countries are much more developed in urban areas than the rural areas. This also appears to be inconsistent with the information on income that shows much smaller proportion of urban respondents than rural respondents falling in the lowest income category (15% for urban vs. 25% for rural). However, it could mean that the rural residents are socially closer and ready to help each other than the urban residents. The results for the effects of these and a large set of control variables on healthcare usage and provider choice are presented in the next section.

6. Estimation Results

6.1. Treatment for illness

The first set of results we present in this section are for a probit model representing utilization of outside healthcare service by those people who reported an incidence of illness over the past two months at the time of the survey. We present results from four specifications. The base specification (Spec 1) controls for key access measures, demographic and socio-economic factors. Specification 2 adds livestock variables, while Spec 3 additionally includes occupational effects. To account for the possible influence of unobserved regional characteristics on our estimates, we have estimated the models both with and without 'Killil'[8] fixed effects. Thus, Spec 4 includes the full set of controls, including Killil fixed effects. We have also corrected the standard errors for within household correlations since the household level characteristics are the same for the members of the same household. To avoid the potential selection bias that could arise because of common unobserved influences on the probability of reporting an illness and the probability of seeking treatment we use the Heckman-type maximum likelihood approach to jointly estimate the binary outcome illness equations and healthcare utilization equations.

Estimated average marginal effects for covariates in the treatment equations for the rural and urban samples are reported in Tables 5 and 6, respectively. We provide parameter estimates for selected treatment equations in Table A1 in the Appendix A.[9] Results from Wald test for independent equations (ρ = rho = 0) show evidence of selectivity for the rural results and urban results with Killil fixed effects (see bottom of Table A1). In addition, models with Killil fixed effects dominate those without Killil effects for rural and urban samples in terms of both Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). Therefore, our discussion of the results mostly focuses on the estimates with Killil fixed effects.

Consistent with many other studies on healthcare utilization in developing countries reviewed previously, physical access to healthcare facilities seems to be an important determinant of seeking treatment for illness in rural but not in urban Ethiopia. Distance to the nearest health center has highly significant negative effect on the decision to seek treatment after falling ill in rural Ethiopia and the absolute magnitude of the coefficient actually increases, albeit slightly, when we control for Killil fixed effects. We observe the same pattern in the estimated coefficient of distance to all weather roads for the rural sample. While the coefficient of access to all weather roads is smaller in absolute magnitude it is negative and highly significant both with and without killil fixed effects. Availability of public transport to the nearest health center also has a significant positive effect on the decision to seek treatment. On the other hand, none of these measures of physical access to healthcare are significant in the urban healthcare utilization equations. This is not surprising given that physical access is less of an issue in urban areas compared to the demand side issues like affordability.

[8] ' Killil' refers to an administrative region in Ethiopia. There are 11 administrative regions including two city administrations (Addis Ababa and Dire Dawa). Killil 1 or Tigray is the excluded category.
[9] Results for the illness equation (akin to first stage estimates) are available from the authors.

Table 5: Average marginal effects of covariates on probability of seeking treatment for illness Probit model with selection (Rural, obs.= 83390)

Variable	Spec 1	Spec 2	Spec 3	Spec 4
Distance to health center/10	-0.0101***	-0.0095***	-0.0093***	-0.0114***
Distance to all weather road/10	-0.0050***	-0.0049***	-0.0049***	-0.0058***
Public transport	0.0245***	0.0243***	0.0220***	0.0204**
Animal plus other	-0.0277***	-0.0276***	-0.0300***	-0.0285***
Liquidity constraint	-0.0066*	-0.0059	-0.0054	0.0017
Age in years/10	-0.0057***	-0.0056***	0.0002	0.0001
Male	0.0060**	0.0061**	0.0032	0.0045
Household size/10	-0.0377***	-0.0360***	-0.0333***	-0.0380***
Married	0.0496***	0.0493***	0.0410***	0.0413***
Primary education	0.0372***	0.0372***	0.0327***	0.0298***
Lower-middle income	0.0203***	0.0205***	0.0215***	0.0197***
Middle income	0.0284***	0.0293***	0.0300***	0.0309***
High income	0.0415***	0.0426***	0.0423***	0.0450***
Living standard	0.0044	0.0046	0.0047	0.0080**
Age-squared and Age*Married	Yes	Yes	Yes	Yes
Livestock variables		Yes	Yes	Yes
occupation dummies			Yes	Yes
Killil effects				Yes
- Log likelihood	64344	64336	64194	63310
Degrees of freedom	40	46	54	74
AIC	128768	128765	128495	126769
BIC	129142	129194	12899	127460

*, **, and *** indicates that the average marginal effect is statistically significantly different from zero at the 10%, 5%, and 1% , respectively.

The statistics at the bottom of the table pertain to coefficient estimates from the seek treatment and illness equations.

The selection equation additionally includes the water, toilet, lighting and cooking variables identified in Table 1.

Table 6: Average marginal effects of covariates on probability of seeking treatment for illness Probit model with selection (Urban, obs.=38931)

Variable	Spec 1	Spec 2	Spec 3	Spec 4
Distance to health center/10	-0.017	-0.0169	-0.0178	-0.0083
Distance to all weather road/10	-0.0014	-0.0015	-0.002	0.004
Public transport	-0.0067	-0.0065	-0.0063	0.0111
Animal plus other	-0.0296	-0.0295	-0.032	-0.0123*
Liquidity constraint	0.0545***	0.0545***	0.0559**	0.0075
Age in years/10	-0.0143**	-0.0144**	-0.0187***	0.0220***
Male	0.0281***	0.0280***	0.0263***	-0.0098*
Household size/10	-0.0075	-0.0078	-0.0074	-0.0775***
Married	0.0328	0.0328	0.037	0.0493***
Primary education	0.0200***	0.0201***	0.0177**	-0.0008
Lower-middle income	0.0248	0.0248	0.0255	-0.0061
Middle income	0.0463*	0.0463*	0.0487*	0.0079
High income	0.0565*	0.0563*	0.0589*	0.0228
Living standard	-0.0222**	-0.0222**	-0.0225**	0.0046
Own house	0.0062	0.006	0.0071	-0.0042
Age-squared and Age*Married	Yes	Yes	Yes	Yes
Livestock variables		Yes	Yes	Yes
occupation dummies			Yes	Yes
Killil effects				Yes
- Log likelihood	24503	24501	24401	23752
Degrees of freedom	42	48	60	80
AIC	49090	49098	48922	47664
BIC	49450	49510	49436	48349

*, **, and *** indicates that the average marginal effect is statistically significantly different from zero at the 10%, 5%, and 1%, respectively.

The statistics at the bottom of the table pertain to coefficient estimates from the seek treatment and illness equations.

The selection equation additionally includes the water, toilet, lighting and cooking variables identified in Table 1.

One variable related to access to transport that appears to have statistically significant but somewhat counterintuitive sign both in rural and urban equations is the use of animal or other forms of transport. According to these results, those who have to use animals or other non-public means of transport are less likely to seek treatment for illness both in rural and urban areas compared to the benchmark category that have to walk to the healthcare centers. This result could be capturing the possibility that for these people the healthcare centers are too inaccessible that they had to resort to animals and other forms of transport to get there. If they don't have sufficient access to these forms of transport, however, they may not be able to get to the healthcare centers even if they are sick and hence we may observe the negative relationship.

Overall, however, access to roads and public transport appears to significantly influence the utilization of healthcare in rural Ethiopia. While the average marginal effect of reducing the distance to all weather roads on the probability of seeking treatment is small in absolute magnitude (see Table 5), it is not substantially different from the average marginal effect of reducing the distance to the healthcare center. According to our estimates for the rural sample, reducing the distance to the nearest healthcare center by 1 kilometer increases the probability of seeking treatment for illness by 0.11% on average whereas reducing the distance to all weather roads by 1 kilometer increases the probability of seeking treatment by 0.06%. These estimated effects for the rural sample imply that, given the respective sample standard deviations of about 10 and 17 kilometers for distance to nearest health center and all weather roads, reducing either the distance to the nearest healthcare center by one standard deviation or distance to all weather roads by one standard deviation increases the probability of seeking treatment by about one percent, all else equal. Also, those who report they can use public transport to travel to the nearest healthcare center are at least 2% more likely to seek treatment for illness than those who have to walk.

These results indicate that there is some degree of substitutability between reducing the distance to healthcare facilities by building more or making the existing facilities more accessible by building road networks and making public transport accessible. Therefore, when policy makers operate with tight budget as in Ethiopia, they could reprioritize the budget so as to focus on building smaller number of healthcare facilities and save resources for building multi-purpose road networks and improve the quality of the existing healthcare facilities.

The other potentially critical issue for utilization of healthcare in developing countries could be access to finance at the time of need. While assets owned and reported monthly income could capture the overall resource profile of the household, they do not necessarily show that the household will have the necessary cash available to finance travel, accommodation and sometimes treatment costs of the emergency illnesses. We have estimated the effects of the indicators for both income and access to liquidity along with ownership of livestock assets and a house in the urban areas. While the effects of assets owned are generally insignificant in both rural and urban healthcare usage equations, the probability of seeking treatment for illness appears to be increasing with rising income profile of the household in the rural areas. The results for urban residents are somewhat mixed, the statistical significance of the effect of income profile essentially vanishing when we control for Killil fixed effects except in the case of high income whose effect remains at least modestly significant even after we control for Killil fixed effects (see Table 6). Education that in some cases could be a better indicator of income than the reported income itself (because of the underreporting of income in household surveys) also has strong positive effect on the decision to seek treatment for illness, particularly for the rural sample.

On the other hand, the effects of our indicator for access to liquidity are mixed but it is interesting to note that access to liquidity appears to be somewhat more important for utilization of healthcare in urban areas than the rural areas. While the coefficient of access to quick cash is not statistically significant in the rural healthcare utilization equations it is significant and positive in the urban equations where we do not control for Killil fixed effects. This again may have to do with the possibility that the closer personal relationships among the rural residents make it easier for them to raise quick cash at the time of emergency than the urban areas. It could also be because the urban residents rely more on the private healthcare that could be afforded only when one has access to liquidity at the time of sickness. In fact, the results in the next sub-section as well as the descriptive evidence presented in Table 3 appear to support this later possibility. The information presented in Table 3 shows that almost three times more people rely on private healthcare in urban areas than the rural and the results presented in the next subsection show that the average marginal effect of access to liquidity on the usage of private healthcare is about five times larger in urban areas.

6.2 Healthcare provider choice

To analyze the effect of access to roads, means of transport and access to liquidity on the choice of healthcare provider, we estimate multinomial logit models including a similar set of explanatory variables we used in the binary choice probit model for the decision to seek treatment for illness. The available treatment options are classified into five categories consisting of self-care, government facilities, private facilities (including NGO facilities), pharmacy, and other facilities including traditional healers. Self-care (no outside treatment) is taken as the benchmark option. Since the magnitudes of the coefficient estimates from multinomial choice model do not tell much by themselves, our discussion here focuses on the average marginal effects of our key variables of interest on provider choice.

The average marginal effects of the selected covariates on the choice among the five alternative treatment options are presented in Tables 7 and 8 for rural and urban samples, respectively.[10] Corroborating our findings from our binary choice models for seeking treatment for illness, the average marginal effects from multinomial logit models show that longer distance to health center and all weather roads both lead to higher probability that people in the rural areas will limit themselves to self-care option. Similarly, access to public transport significantly reduces the probability that sick people in the rural areas will opt for self-treatment. Distance to health center has a positive but weakly significant effect on the probability of self-care in urban areas as well. As opposed to our findings from the binary choice model, the average marginal effects from our multinomial models show that access to liquidity has strong negative effect on the choice of self-care both for urban and rural samples, the effect being more than three times stronger for the urban sample. Education and income also have significant negative effect on the probability of choosing self-care the effect in each case being slightly larger for the rural sample. Now, the question is, how do these measures of physical access and financial constraints influence the probability of choosing among alternative forms of outside treatments?

[10] The multinomial logit coefficient estimates with and without Killil fixed effects for both urban and rural samples are available from the authors. In each case, the model with Killil fixed effects dominates the version without killil effects in terms of both AIC and BIC..

Table 7: Average marginal effects of covariates on healthcare provider Choice Multinomial logit model (Rural, obs.=24205)

Choice/Variable	Spec 1	Spec 2	Spec 3	Spec 4
Self-care				
Distance to health center/10	0.0412***	0.0382***	0.0378***	0.0425***
Distance to all weather road/10	0.0189***	0.0184***	0.0182***	0.0185***
Public transport	-0.0924***	-0.0891***	-0.0881***	-0.0762***
Animal plus other	0.0933***	0.0909***	0.0904***	0.0840***
Liquidity constraint	-0.0192**	-0.0215**	-0.0212**	-0.0275***
Age in years/10	0.0201***	0.0196***	0.0337***	0.0320***
Male	-0.0496***	-0.0497***	-0.0310***	-0.0338***
Household size/10	-0.0262	-0.0332	-0.0417**	-0.0160
Married	-0.0743***	-0.0726***	-0.0663***	-0.0680***
Primary education	-0.1748***	-0.1727***	-0.1623***	-0.1318***
Lower-middle income	-0.0505***	-0.0515***	-0.0514***	-0.0502***
Middle income	-0.0863***	-0.0894***	-0.0892***	-0.0952***
High income	-0.1411***	-0.1456***	-0.1450***	-0.1436***
Living standard	0.0085	0.0069	0.0067	-0.0022
Public Facilities				
Distance to health center/10	-0.0417***	-0.0437***	-0.0436***	-0.0459***
Distance to all weather road/10	0.0006	0.0005	0.0005	-0.0015
Public transport	-0.0241	-0.0232	-0.024	-0.014
Animal plus other	-0.0890***	-0.0899***	-0.0896***	-0.0796***
Liquidity constraint	-0.0013	-0.0037	-0.0039	0.0077
Age in years/10	-0.0112***	-0.0115***	-0.0149***	-0.0148***
Male	0.0250***	0.0251***	0.0144**	0.0144**
Household size/10	-0.0068	-0.0131	-0.0096	-0.0155
Married	0.0401***	0.0409***	0.0364***	0.0390***
Primary education	0.1057***	0.1078***	0.1070***	0.0957***
Lower-middle income	0.0325***	0.0322***	0.0323***	0.0304***
Middle income	0.0463***	0.0440***	0.0444***	0.0467***
High income	0.0531***	0.0502***	0.0508***	0.0525***
Living standard	-0.0041	-0.0046	-0.0044	0.0017
Private Facilities (Including NGOs)				
Distance to health center/10	-0.0118***	-0.0103***	-0.0099***	-0.0107***
Distance to all weather road/10	-0.0071***	-0.0070***	-0.0068***	-0.0055**
Public transport	0.0808***	0.0798***	0.0789***	0.0615***
Animal plus other	-0.0138**	-0.0132**	-0.0133**	-0.0151**
Liquidity constraint	0.0124**	0.0141***	0.0142***	0.0117**
Age in years/10	-0.003	-0.0027	-0.0060**	-0.0051**
Male	0.0082**	0.0082**	0.0069*	0.0080**
Household size/10	0.0230**	0.0281***	0.0294***	0.0188*
Married	0.0126*	0.0120*	0.0115*	0.0108*
Primary education	0.0481***	0.0464***	0.0388***	0.0274***
Lower-middle income	0.0097*	0.0098*	0.0095*	0.0101**
Middle income	0.0254***	0.0272***	0.0265***	0.0288***
High income	0.0485***	0.0504***	0.0496***	0.0482***
Living standard	0.0034	0.0039	0.004	0.0066

Choice/Variable	Spec 1	Spec 2	Spec 3	Spec 4
Pharmacy				
Distance to health center/10	0.0058**	0.0073***	0.0071**	0.0052*
Distance to all weather road/10	-0.0085***	-0.0082***	-0.0082***	-0.0090***
Public transport	0.0243*	0.0231*	0.0236*	0.0244*
Animal plus other	0.0044	0.005	0.0051	0.005
Liquidity constraint	0.0017	0.0027	0.0026	0.0043
Age in years/10	-0.0030*	-0.0027*	-0.0037	-0.0037
Male	0.0104***	0.0104***	0.0081**	0.0085**
Household size/10	0.0163	0.0183*	0.0198*	0.0166*
Married	0.0135*	0.0127*	0.0125*	0.0130*
Primary education	0.0074	0.0066	0.0075	0.0036
Lower-middle income	0.0157***	0.0160***	0.0161***	0.0152***
Middle income	0.0138**	0.0152**	0.0155**	0.0152**
High income	0.0209***	0.0216***	0.0216***	0.0195**
Living standard	-0.0092**	-0.0086*	-0.0087**	-0.0087**
Other Sources (Including Traditional Care)				
Distance to health center/10	0.0064***	0.0086***	0.0086***	0.0089***
Distance to all weather road/10	-0.0039**	-0.0037**	-0.0036**	-0.0025
Public transport	0.0113	0.0095	0.0096	0.0043
Animal plus other	0.0051	0.0073	0.0074	0.0057
Liquidity constraint	0.0064	0.0084*	0.0083*	0.0038
Age in years/10	-0.0030*	-0.0027	-0.0091***	-0.0083***
Male	0.0060*	0.0061*	0.0017	0.003
Household size/10	-0.0063	-0.0001	0.0021	-0.0039
Married	0.008	0.007	0.0059	0.0052
Primary education	0.0137	0.012	0.009	0.0051
Lower-middle income	-0.0073	-0.0064	-0.0065	-0.0054
Middle income	0.0007	0.003	0.0029	0.0045
High income	0.0188**	0.0234**	0.0230**	0.0234**
Living standard	0.0014	0.0024	0.0024	0.0025
Other controls used in all regressions				
Age-squared and Age*Married	Yes	Yes	Yes	Yes
Livestock variables		Yes	Yes	Yes
occupation dummies			Yes	Yes
Killil effects				Yes
- Log likelihood	27370	27323	27278	26565
Degrees of freedom	72	84	100	140
AIC	54884	54814	54756	53411
BIC	55467	55494	55566	54544

*, **, and *** indicates that the average marginal effect is statistically significantly different from zero at the 10%, 5%, and 1%, respectively.

The statistics at the bottom of the table pertain to coefficient estimates from multinomial logit regressions.

Table 8: Average marginal effects of covariates on healthcare provider choice Multinomial logit model (Urban, obs.=8428)

Choice/Variable	Spec 1	Spec 2	Spec 3	Spec 4
Self-care				
Distance to health center/10	0.0927*	0.0922*	0.0905*	0.0717*
Distance to all weather road/10	-0.0048	-0.0047	-0.0028	0.0043
Public transport	0.0134	0.0132	0.0119	-0.0411*
Animal plus other	0.0672***	0.0670***	0.0669***	0.0360**
Liquidity constraint	-0.0879***	-0.0880***	-0.0844***	-0.0762***
Age in years/10	0.0244***	0.0245***	0.0185***	0.0098*
Male	-0.0450***	-0.0450***	-0.0377***	-0.0328***
Household size/10	0.0694**	0.0703**	0.0618**	0.0559**
Married	-0.1003***	-0.1004***	-0.0940***	-0.0775***
Primary education	-0.012	-0.0122	-0.0156	-0.0377***
Lower-middle income	-0.0319	-0.0318	-0.031	-0.0367
Middle income	-0.0785***	-0.0787***	-0.0779***	-0.0768***
High income	-0.1062***	-0.1060***	-0.1031***	-0.0992***
Living standard	0.0316**	0.0316**	0.0302**	0.0298**
Own house	-0.0238*	-0.0235*	-0.0243*	0.0142
Public Facilities				
Distance to health center/10	-0.0593	-0.059	-0.0591	-0.031
Distance to all weather road/10	0.0102	0.01	0.0087	0.0122
Public transport	-0.0237	-0.0236	-0.0233	0.0285
Animal plus other	-0.1056***	-0.1055***	-0.1039***	-0.0633***
Liquidity constraint	0.0198	0.0198	0.0171	0.0115
Age in years/10	-0.0224***	-0.0225***	-0.0136**	-0.0042
Male	0.0206*	0.0205*	0.0190*	0.0129
Household size/10	-0.0422	-0.0433	-0.04	-0.0126
Married	0.0676***	0.0677***	0.0660***	0.0453**
Primary education	-0.0155	-0.0151	-0.015	0.013
Lower-middle income	-0.0008	-0.001	-0.0017	0.0051
Middle income	0.0149	0.015	0.0131	0.0063
High income	0.0114	0.0111	0.0076	-0.008
Living standard	-0.016	-0.016	-0.0151	-0.0124
Own house	0.0178	0.0173	0.0164	-0.0256*
Private Facilities (Including NGOs)				
Distance to health center/10	-0.0787**	-0.0783*	-0.0774*	-0.0780**
Distance to all weather road/10	0.0013	0.0012	0.0015	-0.0012
Public transport	0.0419*	0.0423*	0.0431*	0.0346
Animal plus other	0.0323**	0.0325**	0.0319**	0.0224
Liquidity constraint	0.0613***	0.0611***	0.0609***	0.0584***
Age in years/10	0.0059	0.0059	0.005	0.002
Male	0.0026	0.0025	0.0022	0.0038
Household size/10	-0.0178	-0.0179	-0.0179	-0.0301
Married	-0.0119	-0.0119	-0.0125	-0.0034
Primary education	0.0302***	0.0303***	0.0318***	0.0239**
Lower-middle income	0.0212	0.0211	0.0211	0.0225
Middle income	0.0482**	0.0480**	0.0487**	0.0509**
High income	0.0929***	0.0928***	0.0931***	0.0986***
Living standard	-0.0306***	-0.0304***	-0.0305***	-0.0294***
Own house	-0.0107	-0.0109	-0.0105	-0.0025

Choice/Variable	Spec 1	Spec 2	Spec 3	Spec 4
Pharmacy				
Distance to health center/10	0.0068	0.0068	0.0078	0.0018
Distance to all weather road/10	-0.0072	-0.0072	-0.0079	-0.0142
Public transport	-0.0314***	-0.0315***	-0.0315***	-0.0158
Animal plus other	-0.0411***	-0.0411***	-0.0412***	-0.0293***
Liquidity constraint	-0.0066	-0.0064	-0.0067	-0.0065
Age in years/10	-0.0059**	-0.0059**	-0.0081**	-0.0025
Male	0.0075	0.0076	0.0049	0.0039
Household size/10	0.0182	0.0183	0.022	0.0209
Married	0.0254*	0.0255*	0.0231*	0.0087
Primary education	-0.0128*	-0.0129*	-0.012	-0.006
Lower-middle income	0.0164	0.0167	0.0167	0.016
Middle income	0.0057	0.006	0.0062	0.0113
High income	0.0036	0.0036	0.0045	0.0109
Living standard	0.0150*	0.0149*	0.0151*	0.0126
Own house	0.0227***	0.0229***	0.0234***	0.0118
Other sources (Including Traditional Care)				
Distance to health center/10	0.0384***	0.0383***	0.0381***	0.0354***
Distance to all weather road/10	0.0006	0.0007	0.0005	-0.0011
Public transport	-0.0002	-0.0004	-0.0002	-0.0062
Animal plus other	0.0473***	0.0471***	0.0464***	0.0341***
Liquidity constraint	0.0134*	0.0135*	0.0131*	0.0128*
Age in years/10	-0.0019	-0.0019	-0.0018	-0.005
Male	0.0142**	0.0143**	0.0116**	0.0122**
Household size/10	-0.0275*	-0.0272	-0.0259	-0.0341**
Married	0.0192**	0.0192**	0.0174*	0.0268**
Primary education	0.01	0.01	0.0108	0.0068
Lower-middle income	-0.0049	-0.005	-0.005	-0.0069
Middle income	0.0097	0.0098	0.0099	0.0082
High income	-0.0017	-0.0016	-0.0021	-0.0024
Living standard	0	-0.0001	0.0003	-0.0006
Own house	-0.006	-0.0057	-0.0049	0.0022
Other controls used in all regressions				
Age-squared and Age*Married	Yes	Yes	Yes	Yes
Livestock variables		Yes	Yes	Yes
occupation dummies			Yes	Yes
Killil effects				Yes
- Log likelihood	11501	11498	11475	11039
Degrees of freedom	76	88	112	152
AIC	23154	23173	23174	22382
BIC	23689	23791	23962	23452

*, **, and *** indicates that the average marginal effect is statistically significantly different from zero at the 10%, 5%, and 1%, respectively.

The statistics at the bottom of the table pertain to coefficient estimates from multinomial logit regressions.

Consistent with the findings in many other empirical studies of provider choice in developing countries, our results show that distance to health center has statistically strong negative effect on the utilization of both private and public healthcare facilities by the rural residents, its effect being much larger in the case of public facilities. Distance to health center also appears to have a modestly significant negative effect on the utilization of private facilities by the urban residents. It also appears that distance to health center encourages people to resort to other sources of treatment including traditional care. The marginal effect of distance to health center on other sources of healthcare is consistently positive and strongly significant both for the rural and urban samples. It is not surprising that people may resort to traditional and other forms of informal care when the formal sources are inaccessible.

On the other hand, according to these results neither the distance to all weather roads nor access to public transport has significant effect on the utilization of public health care facilities. However, both have statistically strong effect on the utilization of private healthcare facilities by the rural residents. For example, the results indicate that availability of public transport to the rural residents may increase the probability of seeking treatment in a private facility by 6% to 8%. These findings are consistent with the fact that private facilities are almost exclusively located in urban areas and the rural residents can only make effective use of these facilities if they have accessible roads and means of transport.

Access to liquidity has strong positive effect on the probability of choice of private facilities for both the rural and urban samples the magnitude of the effect being much larger for the urban sample (1.2% in the rural vs. 5.8% in the urban areas). On the other hand, the probability of utilization of public facilities is not significantly affected by household's ability to raise quick cash when needed. This is consistent with the fact that public healthcare is generally provided free of charge or at low cost while only money can buy the private health care services. However, this is not fully supported by our findings about the effects of income and education both of which could be considered as measures of the overall financial capability of the household although they may not be sources of quick liquidity.

According to our results education has significant positive effect on the probability of choosing public facilities only among the rural residents while it significantly affects the probability of choosing private facilities both by the rural and urban residents. In fact, for the rural sample the effect of education on the probability of using public facilities is more than three times larger than its effect on the probability of using private facilities. This may indicate that the effect of education on healthcare utilization comes largely through its effect on health awareness rather than its effect on financial capability. The effect of income follows similar pattern to that of education. It has significant positive effect on the probability of utilization of private facilities by both urban and rural residents whereas it has significant effect on the usage of public facilities only for the rural sample.

In general, for public facilities, physical presence rather than transport facilities or access to finance seem to be more important whereas for private facilities both means of transport and access to liquidity seem to be important. On the basis of these results, therefore, expanding transport facilities (along perhaps with financial services) will facilitate increasing exposure of both the rural and urban residents to private healthcare options. While private healthcare is generally expensive and could be unaffordable for most people in Ethiopia, making it more physically accessible to those who can afford it could help the government release some resources for improvement in the quality of the existing public healthcare facilities and expansion of transport facilities.

7. Conclusion

This paper examines the effect of access to roads, transport and liquidity on seeking treatment for illness and health care provider choice in urban and rural Ethiopia. We estimate selection probit models for seeking treatment for illness and multinomial logit models for healthcare provider choice controlling for a large set of individual and household characteristics. Our analysis is based on a large data set from a nationally representative sample of households.

Our results from the selection probit models indicate that access to roads and public transport are important determinants of the decision to seek treatment for illness by the rural residents but not for the urban residents. According to our estimates for the rural sample, reducing the distance to the nearest all weather road by one kilometer increases the probability of seeking treatment for illness by 0.06%. Also, those who report they can use public transport to travel to the nearest healthcare center are at least 2% more likely to seek treatment for illness than those who have to walk. We do not find strong evidence that liquidity constraint is a major factor influencing the decision to seek treatment, particularly in rural areas. This makes sense in a country where healthcare is largely provided by the state free of charge or at nominal charges. On the other hand, results from multinomial logit models for healthcare provider choice indicate that distance to all weather road, access to public transport and access to liquidity have stronger effect on the utilization of private facilities than public facilities. These results are consistent with the fact that private healthcare facilities are almost exclusively urban-based and are mostly utilized by those who have access to transport and liquidity.

The results imply that it may be possible to improve healthcare utilization not only by focusing on expanding healthcare facilities in the rural areas but by expanding the road networks as well. The significance of distance to all weather roads is appealing in terms of policy design because it implies that construction of road networks can compensate for the absence of healthcare facilities in the proximity. This may help the policy makers to reprioritize the budget so as to focus more on improving the quality of the healthcare facilities and expansion of road networks (since roads are needed for non-health aspects of well-being as well) by scaling back the construction of new healthcare facilities. This is particularly relevant in Ethiopian context where the government has been focusing on building large number of public healthcare centers and health posts (MOH, 2010) but the quality of the facilities is quite low both in terms of human resources, equipment and medical supplies. Expansion of road networks will also enhance the utilization of private healthcare by the rural residents encouraging the later to take a more significant role in the provision of healthcare services in the country.

References

Akin, J.S., Guilkey D.K, Hutchinson, P.L. and McIntosh, M.T. 1998. "Price Elasticities of Demand for Curative Health Care with Control for Sample Selectivity on Endogenous Illness: An analysis for Sri Lanka," *Health Economics* 7: 509-531.

Amaghionyeodiwe A.L. 2008. "Determinants of the choice of health care provider in Nigeria," *Health Care Management Science* 11: 215-227.

Appleton, Simon. 1998. "The Impact of Public Services on Health Care and Illness: A Treatment Effects Model with Sample Selectivity", *Journal of African Economies* 7(1): 1-33

Arcury, Thomas A., John S. Preisser, Wilbert M. Gesler, and James M. Powers. 2005. "Access to Transportation and Health Care Utilization in a Rural Region," *Journal of Rural Health*, 21(1):31-38

Asfaw, A. 2003. "How Poverty Affects Health Status and the Health Care Demand Behavior of Households? The Case of Rural Ethiopia," Working paper, Centre for Development Research, University of Bonn.

Behrman, J. and Deolalikar, A. 1988. Health and nutrition (chapter 14). In: Chenery, H. and Srinivasan, T.N. (eds.) *Handbook of Development Economics* (volume I), Elsevier Science Publishers B. V., pp.631–711.

Bolduc D, Lacroix G, Muller C. 1996. The choice of medical providers in rural Benin: a comparison of discrete choice models. *Journal of Health Econ* 15:477–498.

Buor, Daniel. 2003. "Analyzing the primacy of distance in the utilization of health services in the Ahafo-Ano South district, Ghana," *International Journal of Health Planning and Management* 18: 293–311.

Central Statistical Agency of Ethiopia (CSA). 2000. Welfare Monitoring Survey (WMS2000).

________. 2011. National Statistical Abstract.

Chaya, Nada. 2007. "Poor Access to Health Services: Ways Ethiopia is Overcoming it." Publications of Population Action International, Vol.2, Issue 2.

http://www.populationaction.org/Publications/Working_Papers/Poor_Access_to_Health_Services_in_Ethiopia/Summary.shtml

Chang, Fwu-Ranq and Pravin K. Trivedi. 2003. "Economics of self-medication: theory and evidence", *Health Economics* 12: 721–739.

Collier, Paul, Stefan Dercon and John Mackinnon. 2002. "Density versus Quality in Health Care Provision: Using Household Data to Make Budgetary Choices in Ethiopia," *The World Bank Economic Review* 16(3):425-448.

Develay, A., Sauerborn, R. and Diesfeld, H.J. 1996. "Utilization of Health Care in an African Urban Area: Results form a Household Survey in Ouagadougou, Burkina Faso" *Social Science and Medicine* 43(11):1611-1619.

Ensor T, Cooper S. 2004. "Overcoming barriers to health services access: influencing the demand side," *Health Policy and Planning*, 19: 69–79.

Field, K.S.; Briggs, D.J. 2001. "Socio-Economic and Locational Determinants of Accessibility and Utilization of Primary Health-Care", *Health and Social Care in the Community 9:294*–308.

Habtom, G. K and Pieter Ruys. 2007. "The choice of a health care provider in Eritrea," *Health Policy*, 80(1):202-217.

Kroeger A. 1983. "Anthropological and socio-medical health care research in developing countries," *Social Science and Medicine* 17(3): 147-161.

Leclere, Felicia B., Leif Jensen, Ann E. Biddlecom. 1994. "Health Care Utilization, Family Context, and Adaptation Among Immigrants to the United States," *Journal of Health and Social Behavior,* 35(4):370-384.

Leonard, Kenneth L. 2007. "Learning in Health Care: Evidence of Learning about Clinician Quality in Tanzania," *Economic Development and Cultural Change* 55:531–555.

Manning, W.G., Newhouse, J.P., Duan, N., Keeler, E.B., Leibowitz, A. and Marquis, M. S. 1987. "Health Insurance and the Demand for Medical Care: Evidence from a Randomized Experiment," *American Economic Review*, 77(3): 251-277.

MOH, Federal Democratic Republic of Ethiopia. 2008. Health and Health Related Indicators, available at http://cnhde.ei.columbia.edu/indicators/

_______. 2010. Health and Health Related Indicators, available at http://www.ethiomedic.com/index.php/national-guidlines/file/8-healthandhealthrelatedindicators2001.html?start=5

Mwabu G., Ainsworth M., Nyamete A. 1993. "Quality of medical care and choice of medical treatment in Kenya: an empirical analysis," *Journal of Human Resources* 28(4):838–862.

Nemet, G.F. and Bailey, A.J. 2000. "Distance and Health Care Utilization among the Rural Elderly," *Social Science and Medicine* 50:1197–208

Pohlmeier, W. and Ulrich, V. 1995. "An Econometric Model of the Two-Part Decision making Process in the Demand of Health Care," *The Journal of Human Resources*, 30(2): 339-361.

Pillai, Rajamohanan K., Sankey V. Williams, Henry A. Glickc, Daniel Polsky, Jesse A. Berlin and Robert A. Lowe. 2003. "Factors affecting decisions to seek treatment for sick children in Kerala, India," *Social Science & Medicine* 57: 783–790.

Sepehri, Ardeshir, Saeed Moshiri, Wayne Simpson and Sisira Sarma. 2008. "Taking account of context: how important are household characteristics in explaining adult health-seeking behavior? The case of Vietnam," *Health Policy and Planning*, 23:397-407.

Sharma, Rajiv, Mitchell Haas, and Miron Stano. 2003. "Patient Attitudes, Insurance, and Other Determinants of Self-Referral to Medical and Chiropractic Physicians," *American Journal of Public Health* 93(12):2111-2117.

Tembon, A.C. 1996. "Health care provider choice: the North-west province of Cameroon," *International Journal of Health planning Management* 11:53–57.

Thuan, Nguyen Thi Bich, Curt Lofgren, Lars Lindholm and Nguyen Thi Kim Chuc. 2008. "Choice of healthcare provider following reform in Vietnam", *BMC Health Services Research* 8:162.

Trivedi P. 2002. "Pattern of health care utilization in Vietnam: analysis of 1997-98 Vietnam Living Standards Survey data," Working Paper No. 2775, Washington, D.C.: The World Bank.

USAID. 2008. Powering Health: Options for improving energy services at health facilities in Ethiopia, downloaded at http://pdf.usaid.gov/pdf_docs/PNADM769.pdf.

WHO. 2008. "Traditional Medicine", Fact Sheet No. 134. Available at http://www.who.int/mediacentre/factsheets/fs134/en/

______. 2012. World Health Statistics, Geneva, Switzerland.

Wooldridge, J.M. 2010. *Econometric Analysis of Cross Sectional and Panel Data*, Second Edition. Cambridge: The MIT Press.

Table A1: Parameter estimates from probit model with binary selection for seeking treatment

Variables	Rural (83390 obs.)		Urban (38931 obs.)	
	Spec3	Spec4	Spec3	Spec4
Distance to health center/10	-0.0477***	-0.0584***	-0.0932	-0.0342
	(0.011)	(0.011)	(0.065)	(0.039)
Distance to all weather road/10	-0.0248***	-0.0299***	-0.0105	0.0165
	(0.006)	(0.006)	(0.039)	(0.032)
Public transport	0.1124***	0.1046**	-0.0337	0.0457
	(0.039)	(0.043)	(0.076)	(0.041)
Animal plus other	-0.1531***	-0.1466***	-0.1588***	-0.0507***
	(0.025)	(0.026)	(0.056)	(0.027)
Liquidity constraint	-0.0276	0.0087	0.2776***	0.0309
	(0.018)	(0.019)	(0.043)	(0.037)
Age in years/10	-0.0482**	-0.0507**	-0.1072***	0.0823***
	(0.019)	(0.020)	(0.041)	(0.030)
Age-squared/100	0.0105***	0.0109***	0.0041	0.0026
	(0.002)	(0.002)	(0.005)	(0.003)
Male	0.0166	0.0229	0.1395***	-0.0405
	(0.014)	(0.015)	(0.030)	(0.025)
Household size/10	-0.1700***	-0.1954***	-0.0391	-0.3200***
	(0.040)	(0.040)	(0.120)	(0.052)
Married	0.2093***	0.2122***	0.3358**	0.2415***
	(0.038)	(0.040)	(0.138)	(0.065)
Married*Age/10	0.0038	0.0048	-0.0466**	-0.0152
	(0.009)	(0.010)	(0.023)	(0.014)
Primary education	0.1670***	0.1531***	0.0946**	-0.0034
	(0.041)	(0.044)	(0.043)	(0.029)
Lower-middle income	0.1099***	0.1012***	0.1124***	-0.0250
	(0.022)	(0.022)	(0.067)	(0.050)
Middle income	0.1533***	0.1587***	0.2277***	0.0325
	(0.024)	(0.026)	(0.068)	(0.051)
High income	0.2160***	0.2311***	0.2834***	0.0943***
	(0.031)	(0.034)	(0.080)	(0.056)
Living standard	0.0242	0.0413**	-0.1154***	0.0190
	(0.017)	(0.017)	(0.040)	(0.027)
Own house			0.0372	-0.0174
			(0.044)	(0.025)
Livestock variables	Yes	Yes	Yes	Yes
Occupation dummies	Yes	Yes	Yes	Yes
Killil effects		Yes		Yes
Rho	0.940	.923	-0.537	.943
P-value for testing rho = 0	<0.0001	<0.0001	<0.164	<0.0001
-Log likelihood	64194	63311	24401	23752
Degrees of freedom	54	74	60	80
AIC	83390	83390	38931	38931
BIC	64194	63311	24401	23752

Cluster-robust standard errors are shown within parenthesis.
*, **, and *** indicates that the average marginal effect is statistically significantly different from zero at the 10%, 5%, and 1%, respectively.
The binary illness status selection equation (estimates not shown) additionally controls for the water, toilet, lighting and cooking variables identified in Table 1.

SOCIOECONOMIC FACTORS AFFECTING CHILDHOOD MORTALITY IN ETHIOPIA: AN INSTRUMENTAL VARIABLE APPROACH

Fitsum Zewdu Mulugeta[1]

Abstract

The main causes of death in most early childhood mortality are diseases which are preventable and curable. This is the reason why childhood mortality is treated as a development issue rather than a simple health problem. Ethiopia is among the places where the rate of such deaths is high, which is an indication of the poor quality of life that its people have. It is important to study the important factors of childhood mortality and design intervention in order to improve the situation. This study attempts to identify the important factors of childhood deaths by using the Ethiopian Demographic and Health Survey conducted in 2011. We have fitted an instrumental variable probit *model to identify the structural relation between childhood mortality and maternal, child specific and household related variables. Maternal education, maternal age at first birth, total number of children ever born, access to facilities like toilet, safe water, radio and electricity turned out to be inversely related to childhood mortality while boys, multiple births, dirt floor houses and the use of pollutant cooking fuels are related positively.*

Key words: Child, Under-five, Mortality, Death, IV-*probit*, Ethiopia

JEL classification: I31, I12

---

[1] Fitsum Zewdu Mulugeta: Junior Research Fellow, Poverty and Human Resource Research Division, Ethiopian Economics Association/Ethiopian Economic Policy Research Institute. E-mail: fitsumz@gmail.com

Acknowledgement

The author would like to acknowledge that the paper has benefited from the comments of the unanimous reviewers. Dr. Degnet Abebaw, Mr. Sinafikeh Asrat and Dr. Catherine Porter also made a valuable contribution through their feedback during the different stages of the paper.

1. Introduction

High levels of infant and child mortality[2] rates are among the typical characteristics of under developed countries. Child mortality is one manifestation of the poor socioeconomic conditions that a certain community or a country in general faces. Infant and child mortality rate is a popular indicator that is commonly quoted on the agendas of public health and international development agencies. The significance of the issue can also be seen from the fact that it is one of the goals of the United Nations' Millennium Development Goals (MGDs)[3] (Mutunga 2007; Desta 2011).

Child mortality is can be considered as a composite index that reflects the environmental, social, economic, health care situation as well as norms and practices of the community (Kumar and File, 2010). Infant and child mortality is dealt as a socioeconomic issue, besides its intrinsic health nature, because most such deaths result from diarrhea, respiratory infections, malaria, measles and other immunizable childhood infections. These types of diseases are easily preventable and curable in high-income economies. This is evident from the visible inverse correlation between the level of development of a society and their infant and child mortality rates Espo (2002).

The fact that most early childhood deaths are preventable diseases imply that improving the living standard and environmental conditions could easily prevent incidence of diseases and significantly reduce deaths. On the other hand, a neglected environment is a threat for the health of both children and adults. According to Iram and Butt (2008), the root of infant mortality is in the uneven distribution of resources or lack of resources. Resources determine an individual's environmental risks, as well as his/her access to resources to deal with those risks. This implies that households with higher income can afford better health care as well as housing and sanitary conditions, such as clean water and toilet facilities. As a result, high income households are more likely to have better health outcomes as compared to low income households.

[2] Infant mortality refers to death before first birthday while child mortality refers to death between the first and fifth birthdays. For the purpose of this paper we will use childhood mortality to refer to under-five mortality (death before the age of 5 as a whole). This grouping follows from the way Goal 4 is framed in the MGDs (UN, 2010)

[3] One of the goals of the MGDs is to reduce the level of child mortality to two-thirds of what it was in the year 1990 by 2015

The 2011 Ethiopian Demographic and Health Survey (EDHS) reported that one in every 17 children dies before seeing their first birthday, and one in every 11 dies before the age of five. According to the same report there has been some improvements compared to what the situation was ten years ago. For instance infant mortality has fallen from 97 to 59 deaths per 1000 live births while under-five mortality has fallen from 166 to 88 per 1000 live births during this period (CSA 2012). These improvements are remarkable, but there still is much more that need to be done in order to achieve lowering this number to the level set by MGD (which is 67 or less under-five deaths per 1000 live births (Desta, 2011)).

The study tries to identify the important socioeconomic factors of child mortality in Ethiopia, bearing in mind that childhood mortality is of socioeconomic issue in addition to its intrinsic health nature. The study uses the latest available EDHS data and by employing rigorous analysis techniques. More specifically we try to identify how maternal, child and environmental characteristics interact with child survival. The findings of the study will provide crucial information for policies and programs targeting child deaths and achieve the goal set by the MGD regarding child mortality.

The third wave of the Ethiopian Demographic and Health Survey (EDHS) conducted in 2011 is used to identify the structural relationship between child mortality and some socioeconomic factors. We employed a discrete choice model to look at this structural relation. In particular, this paper gives due consideration for endogeneity issues and handles them by employing an instrumental variable approach. Accordingly we found several maternal, child specific and household characteristics to significantly affect chances of childhood mortality. Among these factors are maternal education, maternal age at first birth, total number of children, child specific biological controls and access to facilities like toilet, safe water, electricity and information.

The remaining parts of the paper are organized as follows. Section II summarizes some of the literatures in childhood mortality. Then data and descriptive statistics are presented in section III. In section IV we have the model specification to be followed by results and discussion in section V. Finally section IV concludes and gives some recommendation of the study.

2. Literature Review

Schultz (1984) set the theoretical framework for the analysis of childhood mortality as health production function. This function captures the structural relationship between health outcomes and the household's behavioral variables, such as nutrition, breastfeeding and child spacing. In the framework of health production function childhood mortality risks depend on both observed health inputs and unobserved biological endowments on frailty.

Socio-economic variables such as cultural, social, economic, community and religious factors are considered to be exogenous. Biomedical factors like breastfeeding patterns and hygiene are modeled as endogenous and as having direct effect on health outcomes, while socioeconomic factors affect child mortality indirectly since they work through the biomedical factors (Schultz 1984). Several socioeconomic factors have been found to be associated with infant and child mortality in developing countries. However, the relative importance of these socioeconomic factors varies from society to society based on their level of development (Iram and Butt 2008).

Different empirical researches have been conducted using different approaches and country cases to study childhood mortality. Among the studies conducted using data from Ethiopia include Kumar and File (2011), using a cross-tabulation approach on EDHS-2005; Desta (2011), using a *logit* model on EDHS 2000 and 2001; and Essayas (2003) applying the Cox regression model on EDHS-2000. These studies, although tried to look at the same issue as the current study, used older versions of EDHS and did not account for the possible endogeneity issues, especially with that of household size. Failure to account for endogeneity makes it difficult to make a casual inference between factors of childhood mortality and the outcome of childhood mortality, making their results a mere correlation analysis. Fitsum (2010), even if it used a different data (Ethiopian Rural Household Survey – ERHS 2004), it also suffers from similar issue of overlooking endogeneity.

In addition to the studies conducted using data from Ethiopia, the following studies tried to look at the issue using different country cases. Iram and Butt (2008) looked at socioeconomic factors of childhood mortality in Pakistan by fitting a sequential *probit* model using the Pakistan Integrated Household Data (PIHD). In the analysis of the environmental determinants of child mortality in Kenya, Mutunga (2007) fitted Weilbull and Cox models on the Kenyan

Demographic and Health Survey. Ladusingh and Singh (2006) studied place, community education, sex and child mortality in north-east India by applying multivariate logistic model on the Indian National Family Health Survey, while Klaauw and Wang (2004) used similar survey to study child mortality in rural India employing flexible duration model. For the case of child mortality in rural China, Jacoby and Wang (2004) used competing risk model on the Chinese National Health Survey. The study by Gebremariam (2001) focused on one of the major causes of child mortality, diarrhea, using the Eritrean Demographic and Health Survey and by fitting logistic regression.

Among the socio-demographic factors age at first birth, sex of the child , education of the mother, type of birth, birth order, birth interval, household living standard, access to safe water and better sanitation facilities are the most frequently studied ones. For instance very low or very high age of the mother at first birth is associated with higher risk of child mortality (Kumar and File, 2011; Mutunga 2007; Ladusigh and Singh, 2006). Maternal education is also found to be negatively associated with child mortality. There are a number of channels through which the education of the mother works towards reducing the risks of childhood mortality. These include delaying marriage and subsequent pregnancy if the girl stays longer in school, better understanding of how to take a good care of her children as well as better income as a result of increased schooling (Iram and Butt, 2008; Mutunga, 2007; Ladusingh and Singh, 2006; Jacoby and Wang, 2004; Klaauw and Wang, 2004). Mother's working situation, i.e. whether the mother is working or not does not seem to affect infant mortality according to the findings of Essayas (2003), but the results of the same study suggest that disaggregation by type of work will show that children of women working in agriculture and manual work to face a higher risk of mortality than those women in professional/technical/clerical jobs[4].

Factors like sanitation and safe drinking water are also found to be very important. Access to clean water and sanitation facilities turned out to significantly reduce chances of childhood mortality (Kumar and File, 2011; Fitsum, 2010; Mutunge 2007; Ladusigh and Singh, 2006; Jacoby and Wang, 2004; Klaauw and Wang, 2004). According to Desta 2011, children born from unmarried woman, first born children, children born in multiple births and

[4] Esayas (2003) suggests that this difference could be due to the difference in socioeconomic difference than the difference in jobs, hence we decided to use socioeconomic variables as controls than work status itself

children born with less than 18 month birth interval from the previous birth tend to face higher chances of mortality before the age of five than otherwise.

Other variable considered in such studies include household income or wealth, household headship and sex of the child. Household size is also a variable that was frequently considered in these studies. Intuitively it is expected for household size to be directly associated with child mortality, i.e. children born in larger households face a higher chance of childhood mortality. This is because households had to share their limited resources among all the children they have. But the studies reviewed here present a contradicting result to this, children in larger households have a better chance of survival than otherwise (Desta, 2011; Fitsum, 2010; Mutunga, 2007).

The results of the studies reviewed above tried to confirm that health outcomes result from different socioeconomic inputs. Besides confirming this argument, the studies also identify the direction of influence of these socioeconomic factors. The empirical literatures show that socioeconomic and environmental conditions are very important in explaining infant and child mortality in many developing countries. But some of these studies focus on identifying associations rather than casual relation (see for example Kumar and File (2011)) while others fail to account for the possible endogeneity of household size (Desta, 2011; Fitsum, 2010; Mutunga, 2007; Gebremariam, 2001). The fact that household size has a counter intuitive sign, and the fact that it has not been tested or treated as endogenous variable makes it important to take the issue up as a research topic important. Since the datasets used by these studies are also outdated (EDHS 2000 and 2005 by Desta (2011), EDHS 2005 by Kumar and File (2011) and Ethiopian Rural Household Survey (ERHS), 2004 version by Fitsum (2010)), it also justifies the investigation of the issue once more.

3. Data and Descriptive Statistics

The 2011 Ethiopia Demographic and Health Survey (EDHS) was conducted by the Central Statistical Agency (CSA) under the auspices of the Ministry of Health (CSA, 2012). Prior to this, EDHS was conducted twice in the years 2000 and 2005. We use the 2011 EDHS for the purpose of this analysis. The primary objective of the 2011 EDHS is to provide up-to-date information for policy makers, planners, researchers and program managers, which give guidance in the planning, implementation, monitoring and evaluation of population and health programs in the country.

The information obtained from the EDHS, in conjunction with statistical information obtained from the Welfare Monitoring Survey (WMS) and Household Income, Consumption and Expenditure Survey (HICES), will provide critical information for the monitoring and evaluation of the country's development plans and assist in the monitoring of the progress towards meeting the Millennium Development Goals (MDGs). The 2011 EDHS collected information on the population and health situation, covering topics on family planning, fertility levels and determinants, fertility preferences, infant, child, adult and maternal mortality, maternal and child health, nutrition, malaria, women's empowerment, and knowledge of HIV/AIDS (CSA, 2012).

We have adopted three approaches to define childhood mortality and as a result we ended up with three target populations. The first target population includes any child reported by the mother, regardless of their current age, except those who are under five and still alive. In this case we have defined our child mortality variable to take a value of 1 if the child has died before his/her fifth birthday and 0 if the child lived to see his/her fifth birthday. Here, those alive and under five are not considered to be in the target population. The limitation of this approach is that it considers all children that the mother had given birth so far. All the analysis under this definition are referred to as Model 1.But to discredit this approach, looking at the current situation of the household may not be the best approach to identify factors of child deaths that happened many years ago. Hence, we added the following approaches to our analysis.

In an attempt to account for this limitation we have focused our analysis over the ten years period preceding the survey. So our second definition uses similar approach for defining childhood mortality as above but this time by limiting the analysis to those children born during the ten years period preceding the survey. Here we assumed living conditions not to change significantly over ten years period. We call this approach as Model 2 henceforth. On our third attempt to further hone our analysis, we also included a third definition for child mortality by focusing only in the five years period preceding the survey. In the third case we consider all children born within the five year period since the survey, i.e. children under the age of five during EDHS 2011 survey. Those who died are classified as 1 while all children under five and alive are classified as 0 for the variable childhood mortality. Children under-five and alive, which were disregarded in Model 1 and 2, are considered in this case. We call the analysis based on this approach Model 3.

Table 1: Sampling distribution by region

Region	Model (1)		Model (2)		Model (3)	
	Number	Percent	Number	Percent	Number	Percent
Tigray	3,717	10.87	1,143	9.74	1,282	10.13
Afar	3,203	9.37	1,148	9.78	1,271	10.04
Amhara	4,884	14.29	1,288	10.97	1,394	11.02
Oromiya	4,794	14.03	1,683	14.34	1,852	14.64
Somali	2,523	7.38	1,081	9.21	1,154	9.12
Benishangul-Gumz	2,973	8.7	1,028	8.76	1,115	8.81
SNNP	4,952	14.49	1,779	15.16	1,789	14.14
Gambela	2,223	6.5	852	7.26	901	7.12
Harari	1,833	5.36	657	5.6	724	5.72
Addis Ababa	1,382	4.04	407	3.47	428	3.38
Dire Dawa	1,697	4.96	671	5.72	744	5.88
Total	34,181	100	11,737	100	12,654	100

Source: Own computation using EDHS-2011 data

As a result of the above definitions we have 34,181 observations under Model 1, 11,737 under Model 2 and 12,645 under Model 3. The educational status of the mother shows that the majority (more than 70 percent in all the three cases) did not have any education. Some 18 to 24 percent attended primary education, 2 to 3 percent attended secondary and less than 2 percent attended higher education depending upon our definition of the target population. When looking at the situation of access to safe water, around 50 percent reported to have access to improved water source. Access to toilet ranges between 48 and 53 percents depending on how we defined the target population.

Availability of electricity concentrates around urban areas. From our overall respondents, less than 19 percent had electricity. The access to electricity goes as high as 82 percent for urban areas while it remained below 7 percent in the rural. When considering the main materials from which the floor of the houses is made, about 65 percent of the houses have dirt floor.

Table 2: Access to electricity by type of place of residence

Place of Residence	Percentage of households having access to electricity		
	Model (1)	Model (2)	Model (3)
Urban	82.07	78.44	77.60
Rural	6.31	6.45	6.20
Total sample	19.94	18.83	18.14

Source: Own computation using EDHS-2011 data

A typical household has six or seven members while the smallest stand at one while the largest stand at 20. Depending on how we define our target population with regard to child mortality, between 76 to 82 percent of these households are headed by male heads. On the average, women in our sample delivered their first birth when they were around 18 years old, but the minimum age of giving first birth is reported to be 10 years[5] while the maximum is 42 years. The children of interest are composed of almost similar proportion in terms of gender, with 51 percent male and 49 percent female. On the other hand, only 3 percent of them were from multiple births.

The variables considered in the econometric model are summarized in Table 3. The dependent variable is defined to be one if a child dies before the age of five and zero if the child survives to see his/her fifth birthday for the case of Model 1 and Model 2, where Model 1 considers all children ever born in the household while that of Model 2 considers those born in the last 10 years only. For the case of Model 3, childhood mortality is coded 1 if the child is dead and 0 otherwise for all children who are less than 5 years old at the time of the survey.

We have selected our explanatory variables based on previously conducted studies in this area. These explanatory variables are of three type, these are maternal characteristics such as the educational attainment of the mother, her age when giving her first birth and total number of children[6] she has given birth. We included variables such as whether or not the child is from a multiple birth and the child's sex in order to control for child related biological characteristics. Finally, a third group of variables are included in the analysis. The variables in the third category are household characteristics like access to toilet, improved water source, access to electricity, the type of material from which the floor of the house is made of as well as the nature of the cooking fuel used by the household (polluting versus non-polluting).

Unlike the other studies, we did not directly include variables that indicate the income level or the living standard of the household. We choose to exclude this index since we already have included some of the components of the wealth index individually, such as toilet facility, access to safe water, type of materials that the

[5] Even though this seems unrealistic, this is what has been reported by the respondents

[6] The previous studies considered household size instead of total number of children. Here we use total number of children since household size was found to have counter intuitive sign by other studies and we would like to follow a different path to check the reliability of these counter intuitive results by taking a different but very closely related variable (total number of children ever born instead of household size)

residential house is made of, etc. so that our model will not suffer from multicollinearity.

Table 3: Descriptive statistics of the selected variables*

Variable	Model (1)	Model (2)	Model (3)
Dependent variable			
Child mortality (1=child is dead before age five, 0= otherwise)	0.1860 (0.3891)	0.1841 (0.3876)	0.0734 (0.2608)
Explanatory Variables			
Maternal characteristics			
Total number of children ever born	6.4911 (2.6933)	5.5067 (2.6241)	4.4132 (2.6050)
Education level – no education (1-yes, 0=no)	0.7786 (0.4152)	0.7582 (0.4282)	0.7095 (0.4540)
Education level is primary (1=yes, 0=no)	0.1866 (0.3896)	0.2063 (0.4046)	0.2434 (0.4292)
Education level is secondary (1=yes, 0=no)	0.0220 (0.1466)	0.0237 (0.1521)	0.0318 (0.1756)
Education level is higher (1=yes, 0=no)	0.0128 (0.1123)	0.0118 (0.1082)	0.0153 (0.1226)
Age at first birth	17.8846 (3.5325)	18.5341 (3.6605)	18.9095 (3.7091)
Square of Age at first birth	332.3356 (139.1658)	356.9119 (150.3564)	371.3263 (154.9887)
Child characteristics			
Child is twin (1=yes, 0=no)	0.0269 (0.1618)	0.0364 (0.1872)	0.0298 (0.1700)
Child is male (1=yes, 0=no)	0.5179 (0.4997)	0.5094 (0.4999)	0.5123 (0.4999)
Household characteristics			
Have access to safe water (1=yes, 0=no)	0.5348 (0.4988)	0.5170 (0.4997)	0.5046 (0.5000)
Have toilet facility (1=yes, 0=no)	0.5317 (0.4990)	0.4939 (0.5000)	0.4807 (0.4996)
Have electricity (1=yes, 0=no)	0.1994 (0.3996)	0.1883 (0.3910)	0.1814 (0.3853)
The floor material is dirt (1=yes, 0=no)	0.6294 (0.4830)	0.6543 (0.4756)	0.6533 (0.4759)
Uses cooking fuel that is polluting (1=yes, 0=no)	0.9743 (0.1582)	0.9759 (0.1535)	0.9773 (0.1489)
Has radio (1=yes, 0=no)	0.3926 (0.4883)	0.3714 (0.4832)	0.3770 (0.4847)
Current age of the mother	37.1559 (7.0520)	32.7419 (6.7644)	29.3608 (6.6805)

Source: Own computation using EDHS-2011 data
Note: * Mean values with standard deviations in parenthesis

4. Model Specification

Our dependent variable is a dichotomous variable that takes values 0 and 1 only. We employed the instrumental variable *probit* (IV-*probit)* model since we are interested in identifying the structural relation between childhood mortality and its factors. Given the option of choosing between the *probit* and *logit* models, we selected the *probit* model for this analysis. The main difference between the two models is the functional form they assume. The *probit* model takes the cumulative density function of a normal distribution functional form while *logit* takes that of a logistic function (Cameroon and Trivedi, 2005). We chose the *probit* model simply because the instrumental variable approach is more developed for the *probit* model than the *logit*. Furthermore, according to Cameroon and Trivedi (2005), there is only little difference in the predicted probabilities between the two models.

Structurally, the *probit* model can be described as follows. Let the observed outcome (weather the child is alive or not in this case) be y_i. According to Verbeek (2002), there exists an unobserved threshold level that marks between a child's survival or not to his/her fifth birthday. This underlying latent variable, say y_i^*, is assumed to be a function of several observed personal and socioeconomic factors, say a vector of x_i s, and unobserved characteristics, say ε_i, for individual i. This can formally be expressed as:

$$y_i^* = x_i'\beta + \varepsilon_i \quad \varepsilon_i \sim NID(0, \sigma_\varepsilon^2) \tag{1}$$

If this threshold level is set to zero, without loss of generality, then the *probit* model can be fully described as:

$$y_i^* = x_i'\beta + \varepsilon_i, \quad \varepsilon_i \sim NID(0, \sigma_\varepsilon^2) \tag{2}$$

and

$$y_i = \begin{cases} 1 & if\ y_i^* > 0 \\ 0 & if\ y_i^* \leq 0 \end{cases} \tag{3}$$

This model assumes that $E[\varepsilon \mid x] = 0$, this in other words means that the dependent variables are exogenous. But one of our dependent variables, total number of children ever born, is suspected to be endogenous ($E[\varepsilon \mid x] \neq 0$). The argument for this suspicion is that it is the possibility of two way causation between child mortality and the total number of children that a woman choose to have. When the rate of childhood mortality is high, women or households in general desire larger number of children in order to compensate for the possible deaths. This is to replace for the possible childhood deaths that the household might experience. On the other hand, the number of children that a woman gives birth to over her life time affects child mortality through the burden it puts on mother herself and the household's resources.

According to Cameron and Trivedi (2010) two way causation will make the parameter estimates inconsistent. We used an instrumental variable (IV) approach to solve this problem. The IV approach follows a two stage estimation technique. In the first stage we estimate a model for total number of children ever born (the instrumented or the endogenous variable) using an instrument variable. An IV approach requires the identification of an instrument that is correlated with the instrumented variable (inclusion criteria) but not with the main dependent variable, except through the instrumented variable (exclusion criteria). Formally, this can be expressed as:

Let z_i be an instrument variable for the endogenous variable x_1, then

i. $E[\varepsilon_i \mid z_i] = 0$ and (4)

ii. $\operatorname{cov}(z_i, x_1) \neq 0$ (5)

In our case we identified the current age of the mother to be an instrument for total number of children that she ever had. Obviously, the total number of children that a woman ever had is related to her current age. On the other hand, current age of the mother does not have a direct relation to the chances of her children to survive until their fifth birthday or not. Since we have a just identified case (where the number of instruments is equal to the number of endogenous variables) it is impossible to test the validity of (4) empirically (Cameron and Trivedi, 2010). In order to check for the validity of the instrument we have checked the significance of the instrument in the first stage equation after controlling for all the other regressors. It was not possible to run formal testes such as multicollinearity, relevance and weak instrument after running an IV-*probit* model, but we run a linear IV regression and run some tests to get some idea whether the instrument is weak or not. Our model passed the variance inflation factor test for multicollinearity and the Stock and Yogo test for weak instruments.

We then fitted an IV-*probit* model with robust standard errors and clustered by region using the method of the maximum likelihood estimation technique. We clustered the regression by region to account for some *within* similarities that each region might have following Cameron and Trivedi (2010). We have also tested for the endogeneity of the variable total number of children ever born using Wald test of erogeneity and rejected the null hypothesis that the variable is exogenous at 1 percent level of significance. The following section presents the results of this estimation.

5. Results and Discussion

After rejecting the hypothesis that total number of children ever born is an exogenous variable, we have reported results from an IV-*probit* model estimated using maximum likelihood estimation procedure. The estimation results of factors of under-five mortality, using the three approaches of defining childhood mortality, are presented in Table 4. Most of the significant variables in Model 1 and Model 2 did not appear to give similar results in Model 3. But we kept the same variables as factors in all the three models in order to make their comparisons possible. Total number of children ever born is found to be significant and negative in Model 1 and 2. This means that as the total number of children ever born by the same mother increases, it tends to decrease the chances of childhood mortality of the child under study. This finding, as much as it seems counter intuitive, is in line with the finds of Fitsum (2010) and Mutunga (2007) regarding household size. The reason for this could be due to factors other than total number of children ever born itself, it could be because wealthier households tending to have many children or the child might benefit from the support and care of the extended family. But the fact that this result keeps turning up in different studies that used different data, country and approach calls for deeper investigation in order to get deeper understanding of the case.

Taking no education as a reference, all level of maternal education are found to significantly and negatively affect childhood mortality in Models 1 and 2, while it is only secondary education that has marginal significance with similar negative contribution to childhood mortality in the case of Model 3. As many researches pointed out (see for example Fitsum, 2010; Iram and Butt, 2008; Mutunga, 2007; Ladusingh and Singh, 2006 and Jacoby and Wang, 2004) maternal education contributes towards reducing childhood mortality by preventing girls from becoming mothers prematurely by keeping them in school, improving the quality of care they could give to their children as a result of better knowledge and through improving their living standard as a result of their better earning potential. The effect of education especially that of primary and higher education, is not clearly visible for the case of Model 3. This could partly be because of the quality of the primary education and the fact that primary education cannot keep them away from early marriage, enough to delay premature pregnancy. Since there are very few observations for higher education, it is again difficult to see its effect clearly here. Secondary education on the other hand, though marginal (10 percent significant level), turned out to be an important factor of reducing chances of childhood mortality.

The above argument is also confirmed by the finding that the maternal age at first birth has significant and negative impact on childhood mortality in Model 1 (at 1 percent level of significance) and Model 2 (at 10 percent level of significance). This means that the older the mother is when giving her first birth, the higher the chance of survival for the child. In other words, giving birth at a very early age will increase the chances of childhood mortality. This also supports the findings of Mutunga, (2007) and Ladusingh and Singh (2006). The square of the maternal age at first birth is significant and positive. This indicates that there exist a quadratic relation between chances of childhood mortality and maternal age at first birth. Meaning, very early and much delayed first births contribute significantly to the increased chances of under-five mortality.

Table 4: Estimation results of the IV-*probit* model

Variables	Coefficients		
	Model (1)	Model (2)	Model (3)
Total number of children ever born	-0.0239*** (0.009)	-0.0980*** (0.0100)	-0.0182 (0.0145)
Primary education	-0.117*** (0.017)	-0.122*** (0.0395)	-0.0547 (0.0511)
Secondary education	-0.349*** (0.127)	-0.294*** (0.0952)	-0.267* (0.149)
Higher education	-0.377*** (0.131)	-0.363** (0.182)	-0.376 (0.276)
Maternal age at first birth	-0.0740*** (0.0183)	-0.0635** (0.0297)	-0.0528 (0.0448)
Square of maternal age at first birth	0.00162*** (0.0005)	0.00132* (0.0007)	0.0011 (0.0011)
Multiple birth	0.937*** (0.0712)	1.069*** (0.0801)	0.925*** (0.0940)
Child is male	0.121*** (0.0243)	0.122*** (0.0343)	0.108*** (0.0366)
Has toilet	-0.0498* (0.0293)	-0.0426 (0.0322)	-0.0337 (0.0370)
Dirt floor	0.0428* (0.0247)	0.109** (0.0481)	0.0601 (0.0616)
Has radio	-0.0484** (0.0202)	-0.0094 (0.0377)	-0.0205 (0.0364)
Has access to improved sources of drinking water	-0.0609** (0.0278)	-0.0878*** (0.0331)	-0.0156 (0.0401)
Has electricity	-0.124** (0.0571)	-0.157** (0.0681)	0.00811 (0.0565)
Use polluting cooking fuel	0.310*** (0.0728)	0.424*** (0.0589)	0.217*** (0.0764)
Constant	-0.248 (0.205)	-0.102 (0.331)	-1.086** (0.457)
Observations	34,169	11,729	12,646
Wald test of exogeneity chi^2 (1)	48.89***	64.42***	26.21***
Instrumented variable	Total number of children ever born		
Instrument variable	Current age of the mother		

Source: Own computation using EDHS-2011 data

Notes: Standard errors in parenthesis. Coefficients are significant at *10 percent, ** 5 percent and *** 1 percent.

The biological controls for the child specific characteristic are also found to be significant and consistently positive in all the three cases. The results suggest that children born in multiple births (twins) and boys have higher chance of mortality before reaching five as compared to children resulting from single births and girls respectively. The finding that male children are biologically more disadvantaged than female children is in-line with the findings of Mutunga, (2007) and Ladusingh and Singh (2006); while it contradicts that of Iram and Butt (2008). Mutunga, (2007) also found the higher probabilities of under-five death for twins than single births.

Access to improved water source, toilet, electricity and radio (which is a proxy for access to information) turned out to be negative and significant for Model 1 while it is only access to improved water and electricity that are significant and negative in Model 2. On the other hand households with dirt floor and using polluting sources as main cooking fuel face a higher chance of childhood mortality in the case of Model 1 and 2 while it is only polluting cooking fuel that has similar effect in Model 3.

The differences between the findings of these three approaches (Model 1, 2 and 3) suggest that the way we define childhood mortality affects the result of the analysis. By moving away from the definition in Model 1 to the definitions in Models 2 and 3 we end up dropping some two third of our observations. This resulted in inflating our standard errors, hence loss of significance of some of our variables. But despite being insignificant, all our variables except access to electricity in Model 3 have similar direction of influence as in Model 1. Since our analysis considered factors like access to safe water, floor material of the house, possession of radio and access to electricity, factors which are components of the wealth index and highly correlated with type of place of residence (rural versus urban), we did not directly put wealth index and type of place of residence variables in the analysis.

6. Conclusion and Recommendation

The study tried to identify factors of childhood mortality using the latest round of the Ethiopian Demographic and Health Survey which was conducted in 2011. After fitting an IV-*probit* model to identify the structural relations, our results suggest that total number of children ever born by a mother has a negative and significant effect on a child's chance of mortality. Furthermore, maternal education, age at first birth, access to toilet, safe water, electricity and radio lowers the chance of childhood mortality. On the other hand children resulting from multiple birth, boys, children living in houses of dirt floor and that use polluting cooking fuels face a higher chance of childhood mortality despite some variations on the level of the significance of the variables depending on different approaches of defining childhood mortality.

Given the fact that total number of children ever born has a counter intuitive sign, we recommend further studies to take up the issue and identify how exactly this relation works. Regarding education, the results show that sending and keeping girls in schools will improve the situation of childhood mortality. The government and its development partners should exert their at most effort to ensure universal access to education beyond the primary level as well as to mobilize the society to send their children, particularly girls, to school. Educated mothers will have both the knowledge and the means to give quality care to their children and other members of the family. Programs designed to tackle this problem should also have information, education and communication sub-programs targeted to creating awareness about the problems of early marriage and early pregnancy. This approach could be integrated with motivating parents to send and keep girls in school, as schooling contributes to delaying marriage and pregnancy.

Due to the vulnerability of twins and male children, extra attention must be given while giving care to children both at home and health facilities. Access to radio is also found to be a significant factor. This indicates that access to information contributes inversely to childhood mortality. In other words, better informed households have a better chance of child survival as compared to those otherwise, hence, efforts to inform and educate the community should be strengthened and continue.

In general we suggest that interventions designed to reduce infant and child mortality should pay attention to these socioeconomic factors of childhood mortality along with the preventive and curative healthcare interventions. The nationally representative data gives us a general picture, but interventions should consider the peculiarities of each society and villages while designing and implementing interventions, hence, it is reasonable to conduct specific studies for specific area of intervention.

References

Cameron, A.C, and Trivedi, P.K. 2005. *Microeconometrics Methods and Applications*, Cambridge University Press, New York

_______. 2010. *Microeconometrics Using Stata*, Stata Press, College Station, Texas

CSA. 2000. 'Ethiopia Demographic and Health Survey 2000' Central Statistical Authority, ORC Macro Calverton

_______. 2005. 'Ethiopia Demographic and Health Survey 2005' Central Statistical Authority, ORC Macro Calverton

_______. 2012 'Ethiopia Demographic and Health Survey 2005' Central Statistical Agency, Addis Ababa, Ethiopia and ICF International, Calverton, Maryland, USA

Desta M. 2011. 'Infant and Child Mortality in Ethiopia: The Role of Socioeconomic, Demographic and Biological Factors In the Previous Five Years Period of 2000 and 2005.' MSc. Thesis, Lund University

EEA. 1999/2000. 'Annual Report on the Ethiopian Economy' Ethiopian Economic Association. Vol I, ed. Befekadu Degefe and Berhanu Nega

Esayas, M. 2003. 'Mother's Work Status and Infant Mortality in Ethiopia: A Study Based on Demographic and Health Survey,' Master's Thesis, Addis Ababa University

Espo, M. 2002. 'Infant Mortality and its Underlying Determinants in Rural Malawi', Dissertation, University of Tamper Medical School.

Filmer, D. and Pritchett, L. 1999. 'The Effect of Household Wealth on Educational Attainment: Evidence from 35 Countries.' Population and Development Review 25(1): 85-120.

Fitsum Z. M. 2010. 'Socioeconomic Determinants of Infant and Child Mortality in Rural Ethiopia', Proceeding of the 7th International Conference on the Ethiopian Economy, edt. Getnet Alemu, vol. 2: pp. 113-136

Gebremariam, W. 2001. 'Diarrheal Morbidity among Young Children in Eritrea: Environmental and Socioeconomic Determinants', *Journal of Health Population Nutrition* 19(2): 83-90

Iram, U. and Butt, M. S. 2008. 'Socioeconomic Determinants of Child Mortality in Pakistan: Evidence from Sequential Probit Model', *International Journal of Social Economics*, Vol. 35 No. 1/2, 2008

Jacoby, H. and Wang, L. 2004. 'Environmental Determinants of Child Mortality in Rural China: A Competing Risks Approach', World Bank Policy Research Working Paper 3241, March 2004

Klaauw, B. and Wang, L. 2004. 'Child Mortality in Rural India', World Bank

Kumar, P. and File, G. 2011. 'Infant and Child Mortality in Ethiopia: A statistical Analysis Approach', Ethiop. J. Educ. & Sc. Vol. 5 No 2

Ladusingh, L. and Singh, C.H. 2006. 'Place, Community Education, Gender and Child Mortality in North-East India', Population, Space and Place 12, 65-76

Mutunga, C. J. 2007. 'Environmental Determinants of Child Mortality in Kenya', World Institute for Development Economics Research, United Nations University, Research Paper No. 2007/83.

Pass, C. Lowes, B. Davies, L. 2005. *Collins Dictionary of Economics*, 4th ed.

Pritchett, L. and Summers, L.H. 1996. 'Wealthier is Healthier' *Journal of Human Resources* 31(4):841-868.

Schultz, T. 1984. 'Studying the Impact of Household Economic and Community Variables on Child Mortality', Population and Development Review, 10: 215-35

UN. 2010. 'The Millennium Development Goals Report,' United Nations, New York

Verbeek, M. 2002. *A Guide to Modern Econometrics*, John Wiley & Sons Ltd.

Wang, L. 2002. 'Determinants of Child Mortality in Low-Income Countries: Empirical Findings from Demographic and Health Surveys', World Bank

FOREIGN EXCHANGE RATIONING AND WHEAT MARKETS IN ETHIOPIA[1]

Paul Dorosh[2] and Hashim Ahmed[3]

Abstract

In spite of remarkable growth in Ethiopia's agricultural production and overall real incomes (GDP/capita) from 2004/05 to 2008/09, prices of major cereals (teff, maize, wheat and sorghum) have fluctuated sharply in both nominal and real terms. International prices of cereals also fluctuated widely, particularly between 2006 and 2008. However, the links between Ethiopia's domestic cereal markets and the international market are by no means straightforward. Among the major staples, only wheat is imported or exported on a significant scale. And frequent changes in trade and macro-economic policies, movements in international prices and fluctuations in domestic production have at times eliminated incentives for private sector imports of wheat.

From July 2005 to March 2007, private sector wheat imports were profitable and domestic wheat prices closely tracked import parity prices. Then, from April 2007 to May 2008, good domestic harvests coincided with increase international wheat prices, so private sector wheat imports were no longer profitable. Most recently, rationing of foreign exchange for imports effectively stopped private sector wheat imports beginning in about April 2008. Partial equilibrium analysis shows, however, that government imports and sales in 2008-09 effectively increased domestic supply and lowered market wheat prices. These sales at the low official price also implied that recipient households, traders and flour mills enjoyed a significant subsidy. Allowing the private sector access to foreign exchange for wheat imports or auctioning government wheat imports in domestic markets would eliminate these rents and generate additional government revenue, while having the same effect on market prices as government subsidized sales.

Keywords: Exchange Rate Policy, Agricultural Trade, Price Stabilization, Food Policy

JEL Classification: O240 Development Planning and Policy: Trade Policy; Factor Movement; Foreign Exchange Policy
Q180 Agricultural Policy; Food Policy

---

[1] An earlier version of this paper was presented at the Ethiopian Economics Association, Seventh International Conference on the Ethiopian Economy, 26 June, 2009 in Addis Ababa. We wish to thank participants of that conference, as well as Nick Minot, Shahidur Rashid and Alemayehu Seyoum Taffesse, for helpful comments and suggestions. Thanks also to Senseshaw Beyene and Eyasu Tsehaye for their research support. The views expressed in this paper are solely those of the authors and do not represent the official positions of their respective institutions.

[2] International Food Policy Research Institute, IFPRI, p.dorosh@cgiar.org
[3] Ethiopian Development Research Institute, EDRI

1. Introduction

Ethiopia enjoyed remarkable growth in agricultural production and overall real incomes (GDP/capita) from 2004/05 to 2008/09, due to a combination of factors, including good weather, increased efforts in agricultural extension, increased usage of fertilizer, and foreign capital inflows that funded major increases in private and public infrastructure investments. In spite of these developments, prices of major cereals (teff, maize, wheat and sorghum) have fluctuated dramatically in both nominal and real terms. International prices of cereals also fluctuated dramatically, particularly between 2006 and 2008. However, among Ethiopia's major cereals, only for wheat is international trade a major source of supply (or demand).

Nonetheless, the links between Ethiopia's domestic wheat market and the international market are by no means straightforward. Frequent changes in trade and macro-economic policies, movements in international prices and fluctuations in domestic production have at times eliminated incentives for private sector imports of wheat. In particular, after major external shocks to Ethiopia's economy (including increases in world prices of fuel in 2007 and early 2008) exacerbated foreign exchange shortages, access to foreign exchange for imports was restricted (rationed) in March 2008 to avoid excessive drawdown of foreign exchange reserves. As a result, the private sector was not able to freely import wheat, even though high domestic prices relative to international prices made imports potentially very profitable. Instead, the government imported wheat commercially in mid-2008 (in addition to food aid inflows) to increase total supplies and stabilize rising domestic cereal prices.

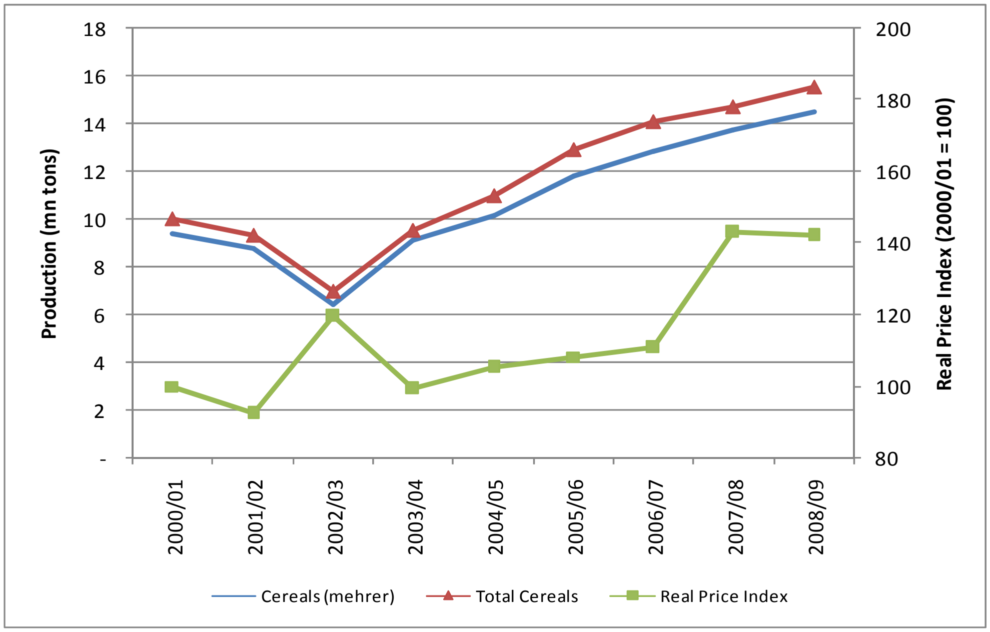
This paper examines these developments in Ethiopia's wheat markets, including the links between international and domestic prices for wheat, the implications of foreign exchange rationing (that effectively stopped private sector wheat imports), and the effects of sales of government wheat imports in 2008-09. Section 2 discusses the evolution of production and prices of cereals in Ethiopia, from 2000/01 to 2007/08, a period characterized by substantial increases in production accompanied by an upward trend in real prices. The macro-economic setting is described in Section 3, focusing particularly on developments from 2004/05 to 2008/09. Section 3 also includes a discussion of the basic analytical framework for assessing the effects of rationing on real exchange rates and domestic prices of wheat. Section 4 focuses on wheat markets, and includes a decomposition of nominal and real wheat prices over time and a description of the various wheat trade regimes that determined the relationship between domestic and international wheat prices in recent years. This section also includes results of a partial equilibrium analysis of the effects of alternative levels of government import sales on domestic prices and consumption of wheat. The final section summarizes the results of the analysis and presents policy implications.

2. Cereal Production, Availability and Prices

Cereal production has increased rapidly in Ethiopia since the 2002/03 drought year (Table 1 and Figure 1). Production in that year was only 7.0 million tons, more than 30 percent below the previous peak of 10.0 million tons achieved in 2000/01. With good rains, production recovered in 2003/04, and by 2004/05, production had reached 10.96 million tons, 9.3 percent greater than the 2000/01 harvest. Although growth in production decelerated from 18 percent in 2005/06 to only 5 and 6 percent in 2007/08 and 2008/09, respectively, production in 2008/09 was still 55 percent higher than in 2000/01. Overall, production grew by an average of 5.6 percent per year between 2000/01 and 2008/09. Sorghum (7.7 percent per year), teff (7.1 percent per year) and wheat (6.2 percent per year) increased at the fastest rates; barley and maize increased by 4.9 and 3.3 percent per year, respectively.

Almost all of the increase in production in the four major cereals (teff, wheat, maize and sorghum) was due to increases in smallholder meher season production, which accounted for 93.2 percent of total production in 2007/08. Production of the four major cereals by large farms (2.5 percent of production in 2007/08) increased by 5.5 percent per year between 2000/01 and 2008/09, essentially the same rate as that of small farms in the meher season (5.6 percent per year). Production of major cereals (mostly maize) in the belg season grew much faster (13.3 percent per year), but still accounted for only 4.2 percent of total annual production in 2007/08.

Figure 1: Production and Real Prices of Major Cereals in Ethiopia, 2000/01 to 2007/08



Source: Calculated from Central Statistical Authority (CSA) production data and Ethiopian Grain Trading Enterprise (EGTE) wholesale price data for Addis Ababa. 2008/09 belg season production is estimated.

Table 1: Cereal Production in Ethiopia, 2000/01 to 2007/08

Crop Year	2000/01	2001/02	2002/03	2003/04	2004/05	2005/06	2006/07	2007/08	2008/09	2000/01 -08/09
Ethiopian Calendar Year	1993	1994	1995	1996	1997	1998	1999	2000	2001	
Teff	1,764	1,658	1,450	1,692	2,044	2,379	2,511	3,027	3,063	7.1%
Wheat	1,738	1,571	1,192	1,740	2,377	2,683	2,720	2,572	2,820	6.2%
Maize	3,428	3,147	2,101	2,830	2,994	3,569	4,426	4,337	4,439	3.3%
Sorghum	1,585	1,608	1,089	1,826	1,758	2,200	2,379	2,726	2,876	7.7%
Barley	1,115	988	816	1,098	1,388	1,639	1,483	1,470	1,635	4.9%
Other cereals	394	364	350	362	400	472	541	562	682	7.1%
Total Production	10,024	9,337	6,998	9,548	10,961	12,944	14,059	14,694	15,515	5.6%
(percentage change)		-7%	-25%	36%	15%	18%	9%	5%	6%	---
Four Major Cereals										
Meher Small Farms	8,072	7,499	5,377	7,659	8,404	9,823	10,993	11,807	12,303	5.4%
(percentage change)		-7%	-28%	42%	10%	17%	12%	7%	4%	---
Belg Small Farms	224	311	330	247	550	761	761	538	538	11.5%
(percentage change)		39%	6%	-25%	122%	38%	0%	-29%	0%	---
Large Farms	218	175	125	182	219	248	280	317	358	6.4%
(percentage change)		-20%	-28%	45%	21%	13%	13%	13%	13%	---
Total	8,514	7,985	5,832	8,088	9,173	10,833	12,035	12,662	13,198	5.6%
(percentage change)		-6%	-27%	39%	13%	18%	11%	5%	4%	---

Source: Calculated from CSA data.

Note: Belg production data for 2008/09 are estimates. Large farm production data for 2007/08 and 2008/09 are estimates.

In spite of these increases in production (and net supply[27]), however, both the nominal and real prices of major cereals rose between 2003/04 and 2007/08, with especially large price increases in 2007/08 (Tables 2 and 3; Figures 2 and 3). From 2003/04 to 2006/07, the average real price of the four major cereals (teff, wheat, maize and sorghum)[28] rose by 12 percent; including 2007/08, the real price increase was 45 percent. The average real price of the four cereals actually declined slightly (by 1 percent) in 2008/09, though.

Table 2: Nominal Wholesale Prices of Major Cereals in Addis Ababa (birr/quintal)

		Nominal Prices		
	Teff	**Wheat**	**Maize**	**Sorghum**
Oct97-Sept98	229.9	175.6	102.8	196.9
Oct98-Sept99	252.5	196.6	126.7	180.7
Oct99-Sept00	272.8	205.8	121.8	203.0
Oct00-Sept01	244.3	149.1	68.1	163.1
Oct01-Sept02	216.4	128.2	69.6	136.9
Oct02-Sept03	252.3	198.0	136.7	205.9
Oct03-Sept04	249.0	172.0	113.7	162.1
Oct04-Sept05	259.0	185.1	146.0	198.1
Oct05-Sept06	324.8	241.5	143.6	241.6
Oct06-Sept07	406.9	283.6	159.7	313.1
Oct07-Sept08	650.6	472.5	369.1	507.1
Oct08-Sept09	869.8	527.1	362.3	625.6
		Annual Change		
	Teff	**Wheat**	**Maize**	**Sorghum**
1997/98 - 1998/99	9.8%	11.9%	23.2%	-8.2%
1998/99 - 1999/00	8.0%	4.7%	-3.8%	12.3%
1999/00 - 2000/01	-10.4%	-27.6%	-44.1%	-19.7%
2000/01 - 2001/02	-11.4%	-14.0%	2.2%	-16.0%
2001/02 - 2002/03	16.6%	54.4%	96.3%	50.4%
2002/03 - 2003/04	-1.3%	-13.1%	-16.8%	-21.3%
2003/04 - 2004/05	4.0%	7.6%	28.4%	22.2%
2004/05 - 2005/06	25.4%	30.5%	-1.7%	22.0%
2005/06 - 2006/07	25.3%	17.4%	11.2%	29.6%
2006/07 - 2007/08	59.9%	66.6%	131.1%	62.0%
2007/08 - 2008/09	33.7%	11.5%	-1.9%	23.4%
2004/05 - 2008/09	**235.8%**	**184.8%**	**148.1%**	**215.8%**

Source: EGTE data.

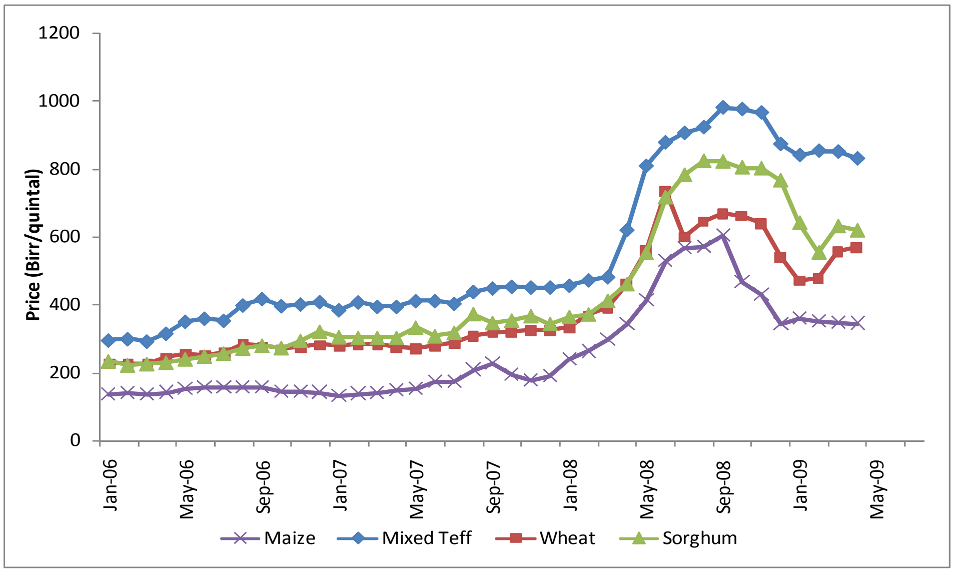
Nonetheless, the steady increases in real cereal prices that accompanied significant increases in per capita cereal supply from 2003/04 to 2007/08 remain a puzzle. Rapidly increasing domestic demand is one major factor. Population

[27] There is very little external trade in teff, maize and sorghum, so net availability is essentially determined by production less seed use and losses. For wheat, external trade is significant, particularly food aid imports which averaged 630 thousand tons per year over this period. However, food aid plus government commercial imports in 2007/08 (about 700 thousand tons) was not much different than food aid in 2001/02 (630 thousand tons).

[28] The real price index reported here is a 2007-08 production-weighted average of the four major cereals.

growth averaged 2.8 percent per year and per capita incomes grew by 7.1 percent per year between 2003/04 and 2006/07. Assuming a (high) income elasticity of demand for cereals of 1.0 on average, total cereal demand would increase by 10.1 percent per year (21.2 percent over two years). This figure is still significantly less than the 31.2 percent increase in cereal production over this period, however, suggesting that real prices should have fallen significantly rather than rising by 5 percent. The surge in real cereal prices in 2007/08 is even more puzzling, though it may have been due in part to expectations of a possible poor harvest or reduced levels of imports (after the start of foreign exchange rationing in March 2008).

Figure 2: Wholesale Prices of Cereals in Addis Ababa, 2006-09



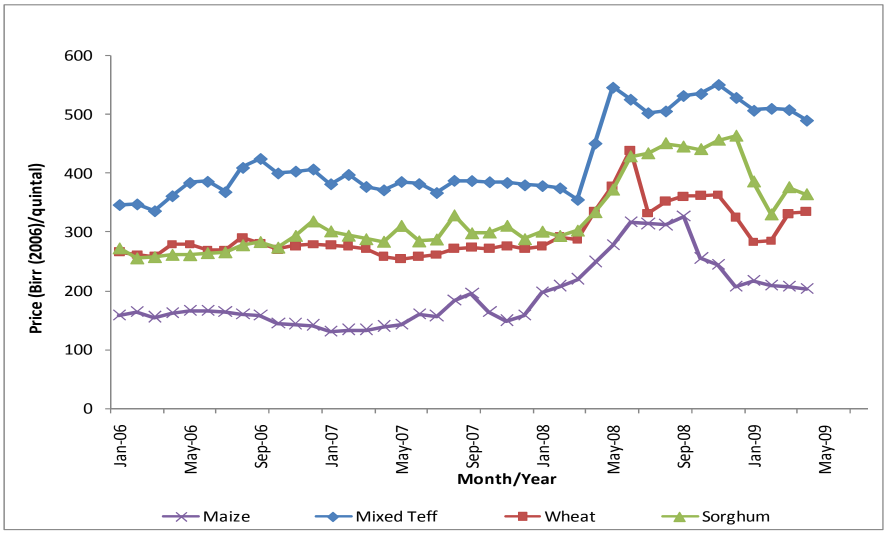
Source: EGTE data.

Table 3: Real Wholesale Prices of Major Cereals in Addis Ababa (birr (2006)/quintal)

	Real Prices			
	Teff	Wheat	Maize	Sorghum
Oct97-Sept98	388.6	297.2	173.8	332.7
Oct98-Sept99	397.6	309.0	198.4	285.1
Oct99-Sept00	417.3	315.1	186.3	310.4
Oct00-Sept01	405.7	246.8	112.6	270.2
Oct01-Sept02	368.5	217.5	117.7	232.3
Oct02-Sept03	363.7	285.5	196.6	296.7
Oct03-Sept04	343.0	236.7	156.3	223.0
Oct04-Sept05	325.4	232.1	182.9	248.1
Oct05-Sept06	361.7	269.7	160.5	270.1
Oct06-Sept07	387.4	269.9	151.1	297.6
Oct07-Sept08	440.8	321.3	244.2	343.5
Oct08-Sept09	507.8	307.2	211.0	364.8
	Annual Change			
	Teff	Wheat	Maize	Sorghum
1997/98 - 1998/99	2.3%	4.0%	14.1%	-14.3%
1998/99 - 1999/00	5.0%	2.0%	-6.1%	8.9%
1999/00 - 2000/01	-2.8%	-21.7%	-39.5%	-13.0%
2000/01 - 2001/02	-9.2%	-11.9%	4.5%	-14.0%
2001/02 - 2002/03	-1.3%	31.2%	67.0%	27.7%
2002/03 - 2003/04	-5.7%	-17.1%	-20.5%	-24.8%
2003/04 - 2004/05	-5.1%	-1.9%	17.0%	11.2%
2004/05 - 2005/06	11.1%	16.2%	-12.3%	8.9%
2005/06 - 2006/07	7.1%	0.1%	-5.9%	10.2%
2006/07 - 2007/08	13.8%	19.1%	61.6%	15.4%
2007/08 - 2008/09	15.2%	-4.4%	-13.6%	6.2%
2004/05 - 2008/09	**56.0%**	**32.3%**	**15.4%**	**47.1%**

*Real prices calculated using the national consumer price index as a deflator (December 2006=100).

Figure 3: Real (Dec 2006) Wholesale Prices of Cereals in Addis Ababa, 2006-09



Source: Calculated from EGTE data and CSA consumer price index.

Considering only the wheat market, the supply and demand calculations appear more consistent with the observed 30 percent increase in real prices from 2000/01 to 2007/08. During this period, population increased by a total of 21 percent and wheat production rose by 52 percent, but per capita availability of wheat increased by only 14 percent since wheat imports changed little. Given the large increase in per capita incomes over this period and a positive income elasticity of demand for wheat, it is likely that per capita demand increased faster than per capita supply. Thus higher real prices of wheat are broadly consistent with main supply and demand factors, but further analysis is needed to explain the magnitude of the real price trends.

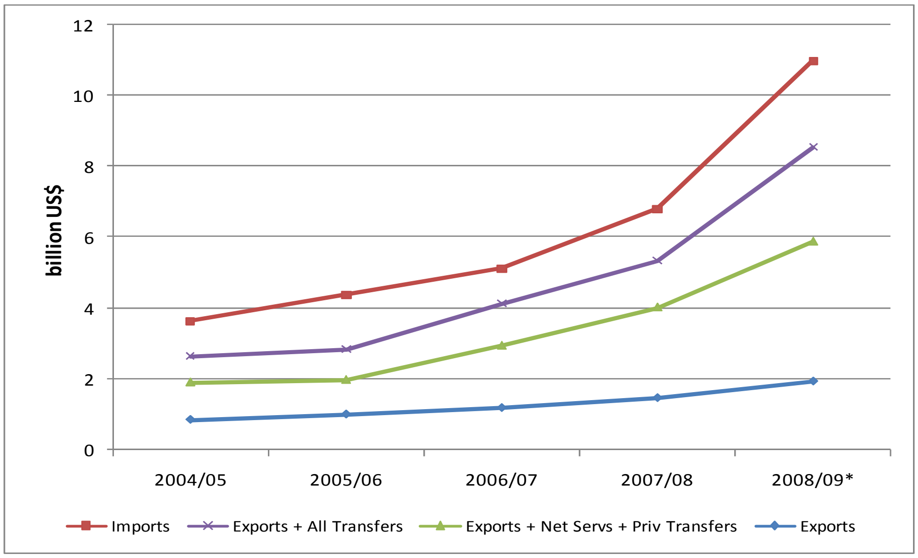
3. The Macro-Economic Setting: The Investment Boom and Foreign Exchange Rationing

The sharp increase in cereal production in recent years coincided with rapid overall economic growth, as well. Between 2004/05 and 2007/08, government policies of expanded domestic credit to finance private investment and increased foreign borrowing to finance public investment contributed to sustained economic growth in excess of 10 percent per year.

However, increased investment implied increased demand for imports (and for foreign exchange), since private (and public) sector investors had access to foreign exchange to finance imported intermediate and capital goods. Merchandise imports surged by 87 percent (US$3.2 billion) between 2004/05 and 2007/08. Half of this increase in merchandise imports was financed by a $1.6 billion increase in annual private transfers; public transfers and merchandise exports also each increased by $0.6 billion (Figure 4).

Moreover, higher world prices, increased domestic credit, foreign capital inflows, changes in expectations and other factors contributed to a surge in domestic inflation, which reached an annual rate of 65 percent in 2007/08 (July 2007 to July 2008). Subsequently, inflation slowed substantially as money supply growth was reduced and a good 2008 meher harvest helped reduce cereal prices. Nonetheless, since nominal exchange rates had changed little relative to the US dollar (and also relative to a basket of currencies of Ethiopia's major trading partners, the high cumulative inflation over the July 2004 to July 2008 period resulted in a real exchange rate appreciation of 34 percent that reduced price inventives for exports and for production of import substitutes (Figure 5).[29]

Figure 4: Ethiopia Imports and Import Financing, 2004/05 to 2007/08



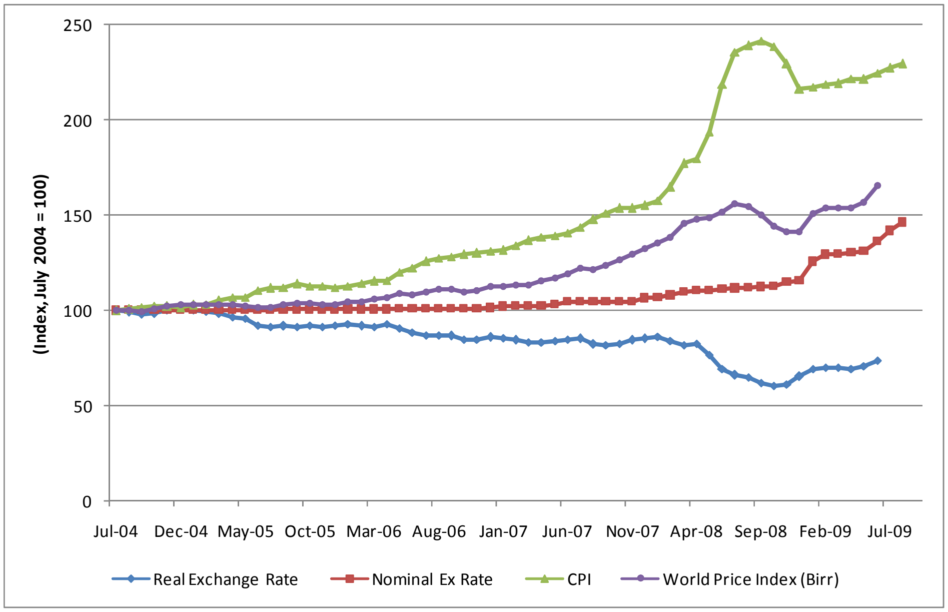
Source: National Bank of Ethiopia data; Dorosh, Robinson and Ahmed (2009).

Up until early 2008, there had been sufficient foreign exchange to finance increased imports. However, growth in foreign exchange earnings slowed in mid-2007 while import demand surged ahead. From the end of June 2007 to the end of March 2008, foreign exchange reserves fell by US$381 mn. (equivalent to 13 percent of the value of merchandise imports in that period). Then, in early 2008, with foreign exchange reserves near zero and import demand in excess of supply of foreign exchange at the prevailing official exchange rate, foreign exchange controls were put in to restrict the effective demand for imports.[30]

[29] Nominal depreciation of the Birr (from 9.83 to 11.39 Birr/US$) between July 2008 and June 2009 helped reduce real appreciation of the birr to 26.3 percent, but this still represented a major reduction in incentives for production of tradables (export goods and import substitutes) since July 2004.

[30] See Dorosh and Ahmed (2009) for an analysis of the macro-economic and distributional effects of foreign exchange rationing relative to an alternative policy of allowing the Birr to depreciate to restore balance between supply and demand for foreign exchange.

Figure 5: Ethiopia Nominal and Real Exchange Rates, 2004-2009

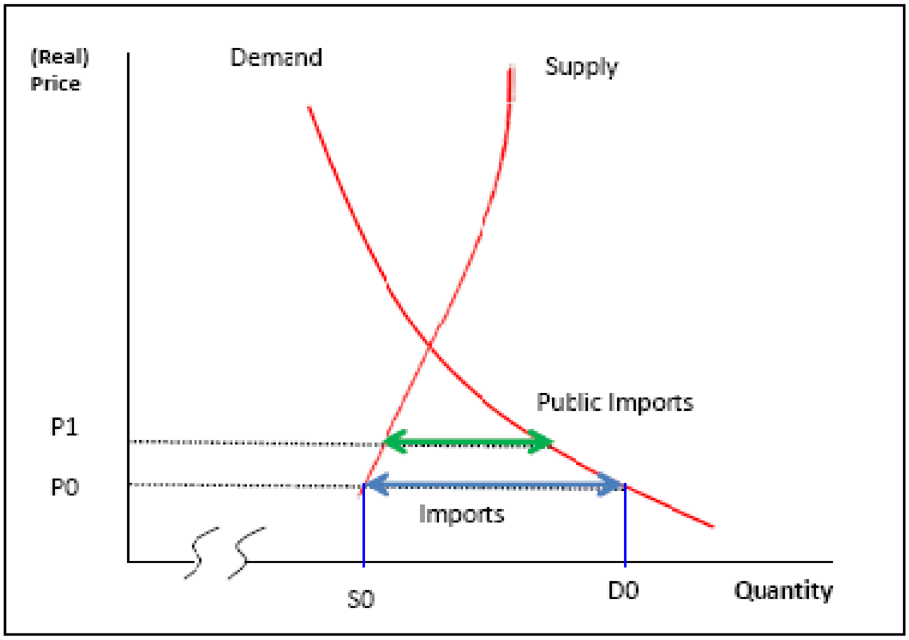


Source: EDRI data; Dorosh, Robinson and Ahmed (2009).

Note: In this figures an appreciation of the real exchange rate is denoted as a decrease in the index.

The effects of foreign exchange rationing on the price of any specific importable goods depends not only on the overall rationing of foreign exchange (which determines the real exchange rate), but also the size of the ration of foreign exchange for the particular importable good. Figure 6 illustrates the case of wheat. If the amount of the ration is less than the amount of wheat that would be imported in the absence of rationing (D0 minus S0), then the market clearing price will rise (in this case from P0 to P1). This same analysis applies whether the rationed amount is imported by the public sector or the private sector, (apart from possible income effects on demand arising from additional incomes of households receiving rationed or subsidized imports). Thus, restrictions on foreign exchange have direct implications for domestic prices of tradable goods, including key food imports like wheat.

Figure 6: Price Effects of Restrictions on Wheat Imports



Source: Authors.

4. Domestic Wheat Price Formation and International Trade

From 2000 to 2009, wheat markets in Ethiopia have been governed by several different regimes of price determination (Box 1). From mid-2000 through 2004, domestic prices of wheat in Addis Ababa were generally below import parity levels but above export parity levels, thus providing little incentive for private imports or exports of ordinary wheat (Table 4 and Figure 7). Domestic prices were on average 24 percent below import parity levels in this period, in part because food aid inflows helped to depress prices to the benefit of net wheat consumers and the detriment of net wheat producers.[31]

Then, from early 2005 to early 2007, domestic prices of wheat (wholesale, Addis Ababa) tracked import parity prices, as private sector wheat imports constituted the marginal supply of wheat in Ethiopia, given levels of domestic production and food aid inflows. Thus, from 2004/05 through 2006/07, domestic prices of wheat were on average only 0.8 percent higher than import parity prices (Table 4).

[31] See Rashid, Assefa and Ayele (2008) for estimates of price distortions in Ethiopian agriculture.

Box 1: Wheat Market Regimes in Ethiopia, 2000 to 2009

Regime 1: January 2000-June 2005: Domestic wheat prices were generally between import and export parity

- Given levels of official imports (including food aid), there was little incentive for private sector imports of ordinary wheat
- Domestic prices were determined by domestic supply (including official imports) and demand

Regime 2: July 2005-March 2007: Domestic wheat prices were generally at import parity levels

- Private sector imports adjusted to equate total supply and domestic demand at the import parity price

Regime 3: April 2007- May 2008: Domestic wheat prices were again below import parity

- Given sharp increases in world prices, private sector imports were not profitable

Regime 4: June 2008 – May 2009: Domestic wheat prices were above import parity

- Restrictions on foreign exchange for imports prevented private imports from taking advantage of profitable import opportunities

Source: Authors.

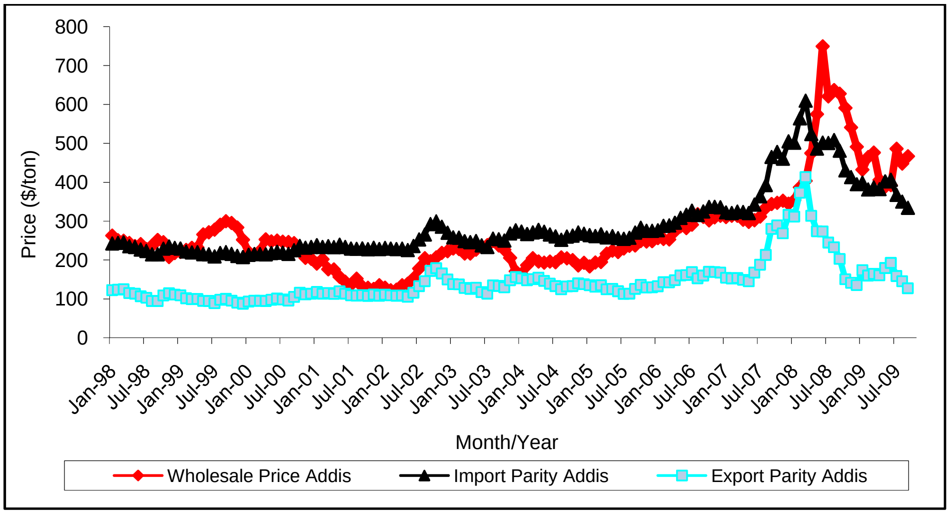
Table 4: Domestic and Import Parity Prices of Wheat in Ethiopia, 1998 – 2009

Crop Year (October-September)	White Wheat Wholesale Addis (Birr/kg)	Exchange Rate (Birr/$)	White Wheat Wholesale Addis ($/ton)	Wheat Import Parity Addis ($/ton)	Nominal Protection Coefficient (percent)
1998-99	1.97	7.87	248.9	221.8	12.7%
1999-00	2.06	8.30	248.0	215.9	15.0%
2000-01	1.49	8.52	175.4	233.5	-24.9%
2001-02	1.28	8.69	147.5	239.7	-38.9%
2002-03	1.98	8.72	227.1	257.7	-11.3%
2003-04	1.72	8.78	195.8	266.1	-25.8%
2004-05	1.85	8.83	209.6	262.4	-17.7%
2005-06	2.42	8.86	272.5	297.8	-3.6%
2006-07	2.84	9.06	313.1	348.7	-2.1%
2007-08	4.73	9.60	489.3	510.1	8.2%
2008-09	5.27	11.39	465.2	387.3	40.5%
Ave. 2000-01 - 04-05	1.66	8.71	191.1	251.9	-23.7%
Ave. 2005/06 - 07-08	3.33	9.17	358.3	385.5	0.8%

* Average of data from October 2008 through April 2009.

Source: Authors' calculations from Ethiopian Grain Trading Enterprise (EGTE) data.

Figure 7: Domestic, Import and Export Parity Prices of Wheat in Ethiopia, 1998 – 2009



Source: Authors' calculations from Ethiopian Grain Trading Enterprise (EGTE) data.

Note: Import and export parity figures are calculated using U.S. Hard Red Winter Wheat Price (fob Gulf of Mexico) plus international shipping (estimated at US$30/ton for December 2008) and domestic handling and transport from Djibouti to Addis (estimated at approximately 1,350 Birr/ton in December 2008).

Since mid-2007, however, domestic wheat prices have NOT been determined by international prices. World prices (import parity Addis Ababa) were higher than domestic prices from mid-2007 through March 2008. Thus, during this period, there were very little imports of ordinary wheat by the private sector as private imports of ordinary wheat were not profitable.

However, when poor rains in many parts of Ethiopia in early 2008 led to a failure of the *belg* season harvest and concerns about adequacy of rainfall for planting of the upcoming 2008 *meher* crops (harvested in October-December), domestic prices rose sharply.[32] Private imports of wheat were apparently again profitable, but restrictions on foreign exchange for imports of wheat (and other goods) were imposed in March 2008.

As a result, import parity did not provide a ceiling on domestic prices of wheat. Instead, domestic wheat prices rose above world prices beginning in May 2008, reflecting the inability or unwillingness of private importers to take advantage of the profitable trade opportunity. Factors such as lack of access to foreign

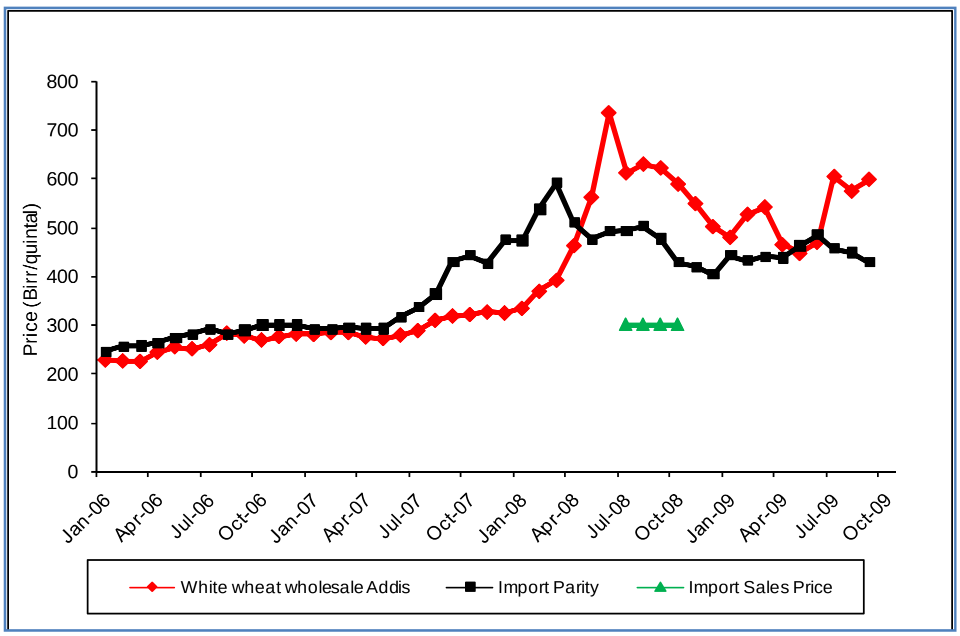
[32] The *belg* harvest accounts for about 15 percent of annual maize production, but less than 2 percent of annual teff, wheat and sorghum production.

exchange, policy uncertainty related to government imports and domestic sales, and concern over possible seizure of private stocks all likely contributed to this lack of private sector import supply response.

In lieu of private sector imports, government policy in mid-2008 was to contract for its own imports of wheat and then sell the wheat at fixed prices in the domestic market (generally 300 Birr/quintal, only about half of the wholesale price of wheat in Addis Ababa market), (Figure 8).

Most of this wheat (55 percent) was sold to flour mills; 23 percent of the subsidized wheat was sold to consumers and 18 percent of the wheat was sold to cooperatives (Table 5). Overall, less than 2 percent of the wheat (8,100 tons) was sold to traders, and none after September 2008, due to concerns that traders did not pass on the huge implicit subsidy to consumers.

Figure 8: Wholesale, Import Sales Prices of Wheat (Addis Ababa, Birr/quintal)



Source: Authors' calculations from Ethiopian Grain Trading Enterprise (EGTE) data.

Table 5: Ethiopia: Distribution of Public Wheat Imports, June 2008 to April, 2009 ('000 tons)

	Flour mills	Cooperatives	Traders	Consumers	Others	Total
June	-	-	-	-	-	9.2
July	-	-	-	-	-	9.7
August	65.1	9.7	3.8	16.2	-	94.8
September	70.9	15.9	4.4	30.3	-	121.5
October	32.9	19.4	-	12.1	2.1	66.4
November	42.6	12.6	-	12.6	2.9	70.6
December	20.4	7.9	-	7.0	0.4	35.8
January '09	14.6	4.0	-	2.1	4.4	25.1
February	8.8	4.5	-	5.1	0.0	18.4
March	19.2	11.9	-	19.7	3.1	53.9
April	9.8	9.4	-	14.2	0.1	33.5
Total	284.3	95.3	8.1	119.2	13.1	538.9
Total (Aug-Apr)	284.3	95.3	8.1	119.2	13.1	520.0
Share	54.7%	18.3%	1.6%	22.9%	2.5%	100.0%

Source: Ethiopian Grain Trading Enterprise (EGTE) data.

Simple partial equilibrium wheat market analysis suggests that the announcement of the wheat imports and the subsequent government wheat sales accounted for the real price decline (see Table 6).[33] Given wheat production in 2007-08 of 2.31 million tons and a 17 percent adjustment for seed, feed and wastage, net wheat production was 1.91 million tons. Adding approximately 400 thousand tons of food aid and net public stock changes gives a total net wheat supply of 2.31 million tons. Average wheat consumption per month is thus about 192 thousand tons per month, and using this average for the July to October 2008 period (i.e. the four-month period before the major *meher* season wheat harvest), wheat consumption would be 192 thousand tons per month x 4 months = 770 thousand tons.

Injecting an additional 200 thousand tons of wheat on the market over the four month (July – October) period, as in Simulation 1, increases net supply by 26 percent. Using an own-price elasticity of demand of -0.35, simulated market prices fall by 48 percent in real terms (i.e. adjusted for overall inflation in the

[33] The equations used for this partial equilibrium analysis are given in Appendix 1. This methodology is a one-commodity simplified version of the multi-market model outlined in Dorosh, Dradri and Haggblade (2009), used for Zambia. See Braverman and Hammer (1986) and Sadoulet and de Janvry (1995) for a detailed description of multi-market models. A first version of this analysis of the potential price impact of injections of government wheat imports on the domestic wheat market was first presented in Gabre-Madhin, Dorosh and Kulkarni (2008). See Diao et al. (2007) and Rashid et al. (2009) for more detailed multi-market model analyses for Ethiopia.

CPI). Using an own-price elasticity of demand of -0.8, real market prices fall by 25 percent. Sales of 300 thousand tons (almost exactly what was actually sold) would lower real market prices by 61 percent with the more price-inelastic demand (elasticity of -0.35) and 34 percent with the more elastic demand (elasticity of -0.8).[34]

Table 6: Partial Equilibrium Estimates of Impacts of Government Wheat Sales in 2008

Simulation	1a	1b	2a	2b
Assumptions				
Government imports (mn tons)	0.200	0.200	0.300	0.300
Elasticity of wheat demand	-0.35	-0.80	-0.35	-0.80
Base Data				
Wheat Supply (mn tons/month)	0.192	0.192	0.192	0.192
4 month wheat supply (mn tons)	0.770	0.770	0.770	0.770
Results				
% change in net supply	26%	26%	39%	39%
New wheat price (Birr/quintal)	349	506	264	447
% change wheat price	-48.3%	-25.1%	-61.0%	-33.7%
Reference				
Actual real price decline: June-Oct	-19.7%	-19.7%	-19.7%	-19.7%
Expected seasonality (2% per month)	8.2%	8.2%	8.2%	8.2%
Actual real price relative to expected price	-25.8%	-25.8%	-25.8%	-25.8%

Source: Authors' calculations.

Note: The wholesale market of price of wheat in Addis Ababa in mid-August 2008 was 675 Birr/quintal.

Sales of government imported wheat reduced real wheat prices in domestic markets from July through October, but not by as much as initially expected, as market wheat demand ultimately proved to be quite price-elastic. Announcement of planned imports of 157,500 tons of wheat and disbursements to millers and wholesale traders contributed to a 12 percent fall in wholesale wheat prices in Addis in July 2008 relative to the June 2008 price (24 percent in real terms). Wheat prices rose slightly in real terms in August, but averaged about 20 percent below June 2008 real price levels from August through October 2008. October 2008 real prices were 26 percent below a projected real price without the import

34 Ultimately, the government distributed 292.4 thousand tons of commercially imported wheat through various channels from July through October 2008.

intervention (the June price plus an estimated 2 percent per month real seasonal price rise), somewhat less than the 33.7 percent decrease in simulation 2 using an elasticity of demand of -0.8.

Two factors likely accounted for the smaller than expected real price decline. First, wheat millers may not have milled all the wheat received or sold all the wheat flour produced by October 2008. Second, imported wheat is not a perfect substitute for locally produced wheat, so increases in imported wheat quantities would likely have smaller effects on prices of locally produced wheat than on prices of domestic sales of imported wheat.

Nonetheless, sales at below-market prices implied huge rents (excess profits) for traders and millers who were able to purchase wheat at 300 Birr/quintal and sizeable income transfer to poor households who were able to purchase government wheat directly. (If the cooperatives sold the wheat at market prices, they would also reap huge rents. Otherwise, the value of these rents would be passed on to consumers as a subsidy.) The total value of these rents and subsidies reached about 900 million Birr (about US$90 million), (Table 7).

Table 7: Ethiopia: Subsidy on Government Wheat Sales, August-October, 2008

	Quantity Sold ('000 tons)	Sales Price (Birr/ton)	Market Price (Birr/ton)	Subsidy (Birr/ton)	Total Subsidy mn Birr	Total Subsidy mn $
August	94.8	3,000	6,600	3,600	341	34.3
September	121.5	3,000	6,375	3,375	410	40.5
October	66.4	3,000	5,375	2,375	158	15.5
Total (Average)	282.7	3,000	6,215	3,215	909	90.4

Source: Authors' calculations from EGTE data.

Following the 2008 *meher* harvest, domestic wheat prices fell sharply, but nonetheless have still remained above import parity levels in spite of a 16 percent depreciation of the birr relative to the US dollar and a 27 percent reduction in the international price of wheat (fob US Gulf) from October 2008 to April 2009. Thus, the divergence between international and domestic prices remained.

Nonetheless, in real terms, domestic wholesale prices in Ethiopia in 2008-09 were at essentially the same level as in 1999-2000 and 2000-01 (Table 8). Real domestic prices have increased by 17.7 percent since 2003-04, the year after the major drought of 2002-03, however. International wheat prices (cif Djibouti), which rose steeply in 2007-08 have again returned to more normal levels and in 2008-09 were only 19.5 percent higher than in 2003-04. If not for the real exchange rate depreciation of 13.7 percent during this period (and changes in transport costs), import parity prices would have increased by a similar amount in real terms.

Table 8: Real Domestic and Import Parity Prices of Wheat in Ethiopia, 1998 – 2009

	Import Price CIF Djibouti ($/ton)	Real Import Price CIF Djibouti ($2004/ton)	Real Exchange Rate (Jly 2004=100)	Real Import Price CIF Djibouti (Birr 2004/qntl)	Real Import Parity Price (Birr 2004/qntl)	Real Wholesale Price (Birr 2004/qntl)
1998-99	161.8	186.2	94.3	149.1	204.4	228.9
1999-00	155.9	180.7	95.2	146.7	203.2	233.4
2000-01	173.5	209.1	95.8	181.9	244.8	182.8
2001-02	179.7	215.2	99.4	196.8	262.7	161.1
2002-03	197.7	220.4	94.7	185.2	241.2	211.5
2003-04	204.6	208.8	99.8	183.6	238.8	175.3
2004-05	195.2	190.9	96.4	161.0	216.3	172.0
2005-06	222.5	209.6	90.4	163.2	218.6	199.8
2006-07	262.4	232.4	84.2	167.0	222.1	199.9
2007-08	401.5	307.2	77.2	204.6	258.4	238.0
2008-09*	272.4	223.6	67.9	133.9	190.0	227.6
Ave 00-01 to 04-05	190.1	208.9	97.2	181.7	240.7	180.6
Ave 05/06 to 07-08	295.4	249.7	83.9	178.3	233.0	212.6
03/04-08/09 %change	**55.4%**	**19.5%**	**-13.7%**	**-1.9%**	**-3.2%**	**17.7%**

* Real exchange rate data are from October 2008 through June 2009.

Source: Authors' calculations from Ethiopian Grain Trading Enterprise (EGTE) data.

Note: Import and export parity figures are calculated using U.S. Hard Red Winter Wheat Price (fob Gulf of Mexico) plus international shipping (estimated at US$30/ton for December 2008) and domestic handling and transport from Djibouti to Addis (estimated at approximately 1,350 Birr/ton in December 2008).

5. Conclusions

Wheat price formation regimes have changed several times between 2000 and 2009: For most of this period, domestic prices have not been determined by international border prices. Given foreign exchange rationing starting in March 2008, private sector wheat importers have had restricted access to foreign exchange. Domestic wheat prices have been above wheat import parity prices since May 2008, indicating that it would be profitable for private traders to import wheat if they had access to foreign exchange at the official exchange rate.

The partial equilibrium analysis in this paper shows, however, that government imports and sales in 2008-09 effectively increased domestic supply and lowered market wheat prices. These sales at the low official price also implied that recipient households, traders and flour mills enjoyed a significant subsidy. Allowing the private sector access to foreign exchange for wheat imports or auctioning government wheat imports in domestic markets would eliminate these rents and generate additional government revenue, while having the same effect on market prices as government subsidized sales.

Although government imports and sales reduced market prices from their extremely high June 2008 levels, market prices still averaged 36 percent above import parity prices from July to October, 2008. Inhibiting private sector imports through foreign exchange rationing thus resulted in lower wheat imports, higher wheat prices, lower wheat consumption, and reduced welfare for net wheat consumers.[35] Depreciation of the nominal and real exchange rates from December through June 2009 substantially reduced the gap between domestic wholesale market prices and import parity and thus the negative effects of foreign exchange rationing on net wheat consumers. Restoring a liberalized trade regime would likely completely eliminate the gap between import parity and domestic wholesale prices, while allowing the private sector to respond to future production shocks with timely imports.

[35] The efficiency and distributional effects of foreign exchange rationing go far beyond the wheat sector, however. See Dorosh, Robinson and Ahmed (2009) for an economy-wide analysis of these impacts.

Annex 1: Model Equations

Supply

$S = \underline{X} * (1\text{-loss}) + MPRIV + (\underline{GOVSALE} - \underline{GOVPURCH}) + \underline{\Delta STOCKS}$

Demand (Consumption)

$D = C = C0 * (P/P0)^{ed} (\underline{Y}/Y0)^{ey}$

Equilibrium

$S = D$

Trade

Under free trade: $MPRIV = C - \underline{X} * (1\text{-loss}) - (\underline{GOVSALE} - \underline{GOVPURCH}) - \underline{\Delta STOCKS}$

$P = \underline{PM}$

Under quotas: $MPRIV = \underline{MPRIV}$

P is endogenous

Variable names

C = wheat consumption
C0 = base level of wheat consumption
D = total wheat demand
GOVPURCH = government domestic wheat purchases
GOVSALE = government domestic wheat sales and distribution
MPRIV = net private wheat imports
P = wheat price
P0 = base wheat price
S = total wheat supply
ΔSTOCKS = changes in private sector wheat stocks
X = wheat production
Y = household income

Parameter names

ed = own price elasticity of demand for wheat
ey = income elastity of demand for wheat
loss = combined rate of storage loss and use as animal feed

References

Braverman, Avishay and Jeffrey S. Hammer. 1986. "Multimarket Analysis of Agricultural Pricing Policies in Senegal", Chapter 8 in Singh, Indermit, Lyn Squire and John Strauss (eds.), *Agricultural Household Models: Extensions, Applications and Policy*. Baltimore MD: Johns Hopkins University Press.

Diao, Xinshen, Belay Fekadu, Steven Haggblade, Alemayehu Seyoum Taffesse, Kassu Wamisho and Bingxin Yu. 2007. "Agricultural Growth Linkages in Ethiopia: Estimates Using Fixed and Flexible Price Models", Discussion Paper 695. Washington, D.C.: International Food Policy Research Institute.

Dorosh, Paul A., Simon Dradri and Steven Haggblade. 2009. "Regional Trade, Government Policy and Food Security: Recent Evidence from Zambia", *Food Policy* 34: 350-366.

Dorosh, Paul, Sherman Robinson and Hashim Ahmed. 2009. "Economic Implications of Foreign Exchange Rationing in Ethiopia". ESSP2 *Discussion Paper*. Addis Ababa: IFPRI.

Ethiopian Development Research Institute (EDRI). 2009. Real Exchange Rate Indices for Ethiopia. Trade and Macroeconomics Division Policy Note.

Gabre-Madhin, Eleni Z., Paul A. Dorosh and Bharat Kulkarni. 2008. "Price Stabilization and Commodity Exchanges: The Case of an Open Market Operation for Wheat in Ethiopia", (July 29, 2008). Addis Ababa, Ethiopia (processed).

Rashid, Shahdur, Nicholas Minot, Ruth Vargas Hill, Alemayehu Seyoum Taffesse, Gezahegn Ayele and Dawit Alemu. 2009. Cereal Availability Study in Ethiopia for 2008 Final report submitted to the Joint Research Center (JRC) of the European Union, March (processed).

Rashid, Shahidur, Meron Assefa and Gezahegn Ayele. 2008. "Distortions to Agricultural Incentives in Ethiopia" (Chapter 8) in Kym Anderson and William Masters (eds.). Distortions to Agricultural Incentives in Africa, pp. 231-52. Washington, D.C.: World Bank.

Sadoulet, Elisabeth and Alain de Janvry. 1995. *Quantitative Development Policy Analysis*. Baltimore, MD: Johns Hopkins University Press.

THE IMPACT OF PROXIMITY TO URBAN CENTER ON CROP PRODUCTION CHOICE AND RURAL INCOME: EVIDENCES FROM VILLAGES IN WOLLO, ETHIOPIA

Seid Nuru Ali[1] and Holger Seebens[2]

Abstract

This article attempts to demonstrate how proximity to urban centers influences households' decision to allot their agricultural land to the production of either staple crops or high value cash crops. By applying fractional logit estimation technique on data collected from villages in Wollo of the Amhara Regional State in 2006, it has been found that proximity to urban centers, access to road, and education of the head of the household determine the crop choice in favor of the production of high value cash crops. While the purely liquid wealth positively affects the allocation of land for the production of cash crops, the direction of the impact of livestock on crop choice is found to depend on the particular location of the activities in relation to urban (market) centers. The pattern of crop choice has been translated into a variation in the level of per capita income across villages. Households operating in those villages located far from urban centers with no access to road are found to be the poorest among the villages covered by the study.

Key Words: Location, Crop Choice, Rural Income, Fractional Logit

JEL Classification: D13, Q12, Q15

[1] Senior fellow at the Ethiopian Economic Policy Research Institute, Ethiopian Economics Association, Addis Ababa

[2] Assistant professor at Department for Agricultural Economics and Rural Development, University of Göttingen

Acknowledgment: We are grateful to Prof. Stephen Klasen and the anonymous referees.

1. Introduction

Early thinking on the relation between location and crop choice dates back to the 19th century owing to von Thünen who first formally theorized the importance of location in shaping the duality between the rural and urban economy. In his 'Isolated State', Thünen portrays an economy that consists of an urban center surrounded by homogenous agricultural land which differs only in terms of distance from the urban center. Agricultural produces from the land around a town are transported to the town for trading. Crop choices depend on the cost effectiveness of each crop in terms of transportation. According to Thünen's portrayal, in the inner ring around the town, crops which are costly to transport (such as vegetables) are produced. At the outer annulus of the rings, crops involving lower transport costs (such as grain) are grown (Samuelson, 1983; Fujita and Thisse, 2002).

Looking beyond Thünen's model, observations show that the decision of households located in the outer annulus to produce grains may not be entirely driven by price incentives but could also be an outcome of their desire to be self-sufficient in staple crops in order to smooth consumption. Assuming other factors being constant, high value cash crops can be costly to bring to the market if they are grown in locations far from urban centers. Inherent to different distances of villages from the urban centers is, thus, an unequal distribution of income.

Recent literature on crop choice focuses on uncertainties arising from weather conditions and price shocks. It has been widely argued that various forms of uncertainties contribute to the subsistent nature of many rural areas in developing countries (Dillon and Scandizzo, 1978; Fafchamps, 1992; Dercon, 1996; Ayalew, 2003). In response to this, rural households have developed different strategies to cope with the risk associated with agricultural production. Diversification has been conceived as a feasible insurance strategy, although often implies lower returns. Price fluctuations can be compensated if households cultivate a wide portfolio of crops, among which staple crops—tending to be more stable in terms of prices—constitute an important safety measure. In particular, poor and risk-averse households tend to ensure self-sufficiency in staple crops leading to the limitation of diversification to only different kinds of staple crops.

The impact of risk on crop choice may vary across locations. Even in periods of stable and high prices for cash crops, households' decision to engage in the production of cash crops depends on transportation costs, which in turn depend on the distance of the particular plot or village from the market. For instance, markets for logs and lumber of eucalyptus are well established in urban centers of Ethiopia. However, households living far from urban centers do not grow eucalyptus trees in significant magnitude even on their marginal land because eucalyptus growers living closer to urban centers outbid in the market. One reason for this is higher transportation cost. Distance to markets has thus an important influence on the development prospects of remote villages. Decisions by households to allocate the bulk of land to the production of less valued staple crops results in low surplus and low incomes, implying that the incidence of poverty is likely to increase with distance away from urban centers.

This article attempts to look into how the location of an agricultural activity in relation to markets in urban centers affects the production of high value cash crops using cross sectional data collected from six villages in Wollo, Ethiopia. It also shows the associated disparity in income by location. The remaining part of the paper is organized as follows. Section 2 highlights descriptive facts from surveyed villages. Section 3 presents a simple theoretical framework. Section 4 deals with the econometric analysis. Section 5 concludes.

2. Crop Choice in Selected Villages of Ethiopia by Location

According to the 2007 population census, Ethiopia's population is estimated to be about 88 million people in 2011. The population is growing at a rate of 2.8 percent annually. About 84 percent of the population makes a living from subsistent agriculture accounting for 43 percent of GDP (MoFED, 2011). The country exhibits one of the lowest rates of urbanization where only 16 percent dwells in urban centers. As a result, size of arable land per household decreases making the land issue critical in transforming the Ethiopian economy. Average land size in the country hovers around 0.97 to 1.2 hectare per household. This is equivalent to a mere 0.2 hectare per head (CSA, 2010, 2011).

Ethiopian farmers mainly focus on the production of staple crops except for coffee for which an already established international market exists. One of the major factors for this is believed to be a poorly developed transport network and low demand from the urban center. According to the national data from Central Statistical Agency (CSA), in 2005, 84.3 percent of rural households in Ethiopia, excluding nomadic areas, live on crop and livestock production. In 2011, about 87 percent of the total production of major crops is accounted by cereals. If we exclude teff[3] which is both staple and cash crop, the share of the mainly staple crops in the major crop production is as high as 70 percent. Pulses which are predominantly cash crops have a share of 9.5 percent (CSA, 2005, 2011).

2.1 Location

The study covers six villages in four locations in Wollo, eastern part of the Amhara Regional State. The survey was conducted in the year 2006. The villages were systematically selected based on their increasing distance from major and district towns. The survey also accounts for agro-ecological differences. About 252 households were randomly selected from the villages. The distance between the reference district town and the nearest villages to the town is about 4 kilometers. The farthest village is 20 kilometers away from the nearest district town. Proximity to major towns is also considered. The major towns that are taken as references are Dessie and Woldiya. Dessie is the capital of South Wollo Zone (one of the eleven administrative Zones in the Region) and has an estimated population of 169,000. Woldiya is the capital of North Wollo Zone with an estimated population of 43,000. The two towns are 120 kilometers apart along the main Addis Ababa – Mekele road. District towns include Kutaber and Mersa.

One of the villages covered by the study called Alasha is located in Kutaber district some 12 kilometers from Dessie. The nearest district town to Alasha is Kutaber with an estimated population of 5,000. Two major attributes of the village compared to other survey areas in terms of location are (i) it is the nearest village to major urban centers, and (ii) it is located in the highland plateau characterized by a relatively cool climate.

The other study site, Mersa Zuria area, includes three villages intercepting the district town Mersa on either side of the Dessie-Woldiya road. Mersa has an estimated population of 6,500. The villages have easy access to the market primarily due to their proximity to the major Addis Ababa-Mekele road via Mersa and Woldiya. Besides, the villages are nearer to the district town, Mersa, and the Zone town Woldiya. Among the three villages, Buhoro has significant access to irrigation partly due to availability of tributary rivers.

The third study site is Girana. It is located about 7 kilometers east of the Addis Ababa-Mekele road. There is a gravel road linking the village to the major

[3] *Teff* is an indigenous grass growing in Ethiopia which is used to make Ethiopian staple bread called '*injera*'.

highway. The major attribute of the village is that it has some tributaries which allow for irrigating a significant part of land. Moreover, there is weekly open market in the village attracting people from the surrounding villages.

Among the villages covered by the study, Habru-Ligo has the farthest distance from both urban centers and major roads, and even lacks feeder road. Individuals have to travel a minimum of three hours back and forth on foot on difficult terrains to work on their land. About 25 to 30 percent of the land possessed by the villagers is irrigable.

2.2 Land size and crop choice

The average land size per household ranges from 0.61 hectare in Alasha area to about 1 hectare in Mersa Zuria area. Although Alasha and Kullie have similar distance from district towns, per capita land size in Alasha is lower than in Kullie and even less than that of Menentela which is closest to the next district town. The pattern is similar in term of per capita land size where Alasha has the lowest with 0.13 hectare and Mersa Zuria has the highest with 0.27 hectare. Girana and Habru-Ligo have a roughly equal size of per capita land holding which is about 0.14 hectare.

Table 1: Location and Land Size by Village, 2006

	Distance from District Town (in km)	Distance from Major Towns (in km)		Land Size per Household (in hectare)	Proportion of Land Allotted for Purely Cash Crop and Eucalyptus (%)
		Dessie	Woldiya		
Alasha	7	12	geographically remote	0.613	7.9
Mersa Zuria				1.020	18.2
Menentela	4	94	20	0.822	12.2
Kullie	7	97	25	1.000	9.4
Buhoro	8	98	20	1.160	28.3
Girana	15	75	50	0.666	19.9
Habru-Ligo	20	85	60	0.643	0.9

In terms of land allocation, Buhoro exhibits the highest share of land allotted for the production of cash crops (about 28 percent) while Habru-Ligo has the lowest share which is less than 1 percent. Major cash crops produced are sugarcane, fruits (orange, papaya, guava), coffee, and vegetables. The staple crops include sorghum of various varieties, and teff in villages other than Alasha. Teff is used both as a cash crop and staple food due to its high value in urban markets as it is the major staple for the urban population. During periods of poor harvest, households usually sell their teff and buy other cheaper staple crops such as sorghum for household consumption. However, since teff has low productivity compared to sorghum and maize, households in the study areas allot only a small portion of their limited land for the production of this crop unlike other regions which are endowed with large land size and specialize in the production of the crop on a large scale. Households in Alasha area produce wheat, barley, oats, and pulses.

2.3 Patterns of income

Data on the level of income of households by source has been collected from the villages under study. Among those villages, Mersa Zuria area is relatively affluent with a per capita income of 1830 Birr. This is well above the average per capita national income of about 1300 Birr recorded in 2005 (NBE, Annual Report 2006). Buhoro with a relatively better access to irrigation is specialized in cash crop production. Unlike other villages, 47.5 percent of its income comes from cash crops. The peasants' involvement in the production of high value cash crops in the area is reflected by the fact that about 48 percent of their income comes from 28 percent of their land. Kullie and Menentela, where irrigable land is lacking, the highest share of their income is derived from commercial livestock farming. About 24 percent of household income in Menentela area and 26 percent of the income in Kullie come from livestock farming.

Habru-Ligo has the lowest per capita income (about 520 Birr) among the villages covered by the survey. A typical rural farmer in Habru-Ligo earns just 23 percent of what a typical farmer in Buhoro earns. Though the village has irrigable land, cash crop production is not very common. Peasants in the area do not invest in commercial livestock even though the village is well endowed with suitable conditions for animal husbandry. Households raise cattle, goats and sheep mainly as a buffer stock.

Table 2: Sources of Income of Households by Village, 2006

	Per capita income (in Birr)	Source of Household Income and their Contribution to Total Income (%)							
		SStaple crops	CCash crops	EEucalyptus	WWage	Remittance from Abroad	Remittance from Towns	Rural Enterprise	SSale of Animal
Alasha	934	53.7	2.9	18.3	1.6	2.1	1.6	3.1	12.8
Mersa Zuria	1830	31.2	23.7	9.6	5.3	7.6	2.6	0.0	16.8
Menentela	1545	35.3	0.8	10.4	6.7	15.1	5.0	0.0	24.0
Kullie	1079	47.5	0.6	3.7	8.4	8.3	0.0	0.0	25.5
Buhoro	2298	22.5	47.5	11.2	3.1	2.3	2.0	0.0	8.7
Girana	1087	45.6	20.3	0.1	6.3	14.8	0.6	2.4	3.7
Habru-Ligo	520	86.3	3.8	0.1	1.8	2.2	0.0	0.7	3.2

Source: Own computations from the survey data

Besides crop production, villagers operating nearer to urban centers allot more plots of land for fast growing trees in particular eucalyptus than those located far from urban centers. This partly depends on the type of slope and soil fertility of the plot of land possessed by peasants. In Alasha, hilly and marginal land which is held by peasants privately is largely covered by eucalyptus forests which have demand from urban centers for purposes of construction and energy supply. About 18 percent of household income in Alasha comes from the sale of logs of eucalyptus. In Menentela and Buhoro, between 10 and 11 percent of household income is derived from selling eucalyptus.

3. Theoretical Framework on Location, Crop choice and Rural Income

3.1. Background

We model a Thünen type of environment where rural households make a living from income that is generated from their farming activities. Households dwell and operate at different distance from urban centers. Each household consists of working household members who maximize a joint utility function. Labor time is optimally allocated between agricultural activities and off-farm income generating activities, most importantly employment in the urban centers. However, to make the analysis tractable, the household is assumed to consist of a single individual only.

Agricultural activities involve mainly crop production and animal husbandry. Crop production, which is the mainstay of rural households, involves various items of products, of which the production technologies may differ. We restrict our attention to two major activities, namely, production of stable crops and production of cash crops. In fact, about 74 percent of the income of households in the villages covered by the survey comes from crop cultivation.

The household produces crops by combining land and other inputs such as labor, animal draft power, fertilizer and pesticides. Part of the staple crop and a significant share of the cash crop have to be sold to purchase manufactured goods for consumption. A household not producing sufficient staple crops thus falling short of home consumption has to purchase additional food from the market using the proceeds from the sale of cash crops.

The decision to produce a particular item depends on the relative distance of the activity from the town. Moreover, unlike the Thünen's rings, the land surrounding the town needs not to be uniform so that villages at the same distance from town specialize in different crops. In what follows, we attempt to analyze how location affects the decision of a household to allot a plot of land for either staple or purely cash crops.

3.2 Production technologies and costs

Land is a limited resource. As a result, households rationally decide to invest in high value crops that maximize income per unit of land. Cash crops are preferred not necessarily because they give high yields per unit of land but because they fetch high market value, most importantly in urban centers. Some cash crops such as coffee are not consumed for their nutritional values. Other crops such as vegetables are highly perishable. Staple crops on the other hand give more security to the household against low prices because the household can still use staple crops for own consumption.

The production of the two crops requires factors such as land and labor. We further assume that labor is not a binding constraint for agricultural production. The household is assumed to have a single unit of labor and a single plot of land that can be allotted to the production of cash crops and staple crops. Let l^c and l^s represent the shares of land for cash and staple crops, respectively, so that $l^c + l^s = 1$. Using l^c portion of land, the household produces q^c units of cash crops to be sold at price p^c in urban centers. The remaining land ($l^s = 1 - l^c$) is used to produce

q^s units of staple crop. Part of this crop will be consumed at home and any surplus is sold at the market at a price of p^s.

The production function of the two types of crops that relate the output per labor q^j to a fraction of a unit of land l is, therefore, given by:[4]

$$q^c = A^c f\left(l^c\right)$$

$$q^s = A^s g\left(l^s\right) \tag{1}$$

where q^j denotes output per unit of labor and A^c and A^s are the levels of technology required to produce cash and staple crops, respectively. The production functions are assumed to fulfill the standard conditions:

$$f'\left(l^c\right) > 0\ , \quad f''\left(l^c\right) < 0\ ;$$

$$g'\left(l^s\right) > 0\ , \quad g''\left(l^s\right) < 0.$$

where $f'(.)$, $g'(.)$ and $f''(.)$, $g''(.)$ refer to the first and second order derivates of the production function with respect to land respectively. The technology required to produce staple crops, A^s, is considered a numéraire to which the technology A^c can be compared. Thus, A^s is set to unity so that $q^s = g\left(l^s\right)$.

It is assumed that the decision to produce cash crops also depends on the technical know-how about the production of the particular cash crop. An individual might be a quick innovator in terms of acquiring new technology if he

[4] Practically, some cash crops such as coffee, orange, and pawpaw have maturity period of two to five years. There are also some crops such as vegetables and oilseeds with a maximum maturity period of one year. Ayalew (2003) noted this issue and has taken the opportunity cost of land in terms of yield of annual crops as a result of longer maturity period of coffee trees into account in his model. However, it is customary in the area under study that the land under permanent cash crops can at the same time be used for the production of annual crops until the cash crops grew to a full-fledged tree. Thus, it is not harmful to continue the analysis without considering the opportunity cost of land due to long gestation period of permanent crops.

has some formal education. The technological parameter in the production function of the cash crop is given by:[5]

$$A^c = A_0^c e^{\psi E} \qquad (2)$$

where A^c_0 is some indigenous knowledge of the technology, E is level of education (say in years of schooling), and ψ is a parameter.

Given prices of cash crop and staple crops, the total monetary value of these crops is given by:

$$y = A_0^c e^{\psi E} p^c f\left(l^c\right) + p^s g\left(l^s\right) \qquad (3)$$

The household incurs production costs for each crop. Costs of production of each crop are proportional to land allotted to the production of the crops. Let w^c and w^s represent factor prices per unit of land. The associated cost of production of cash and staple crops are given by $w^c l^c$ and $w^s l^s$.

The household also incurs transportation costs for both crops. We further assume that direct cost of transportation is the same for each crop. However, the cost of transportation varies depending on the amount of crop the household wants to sell. Household sell small shares of the staple crop because most of it is produced for home consumption. We assume that all cash crops produced by the household are sold[6] and let n denote the share of staple crop that is marketable. Then, the total transportation cost with k unit price of transportation is given by $kq^c r$ and $knq^s r$, where r is the distance between the village and the urban center.

The household also faces cost due to the perishable nature of each crop. We define an index that measures the degree of the perishable nature of each crop in connection to transporting the surplus to the market. Let r be the distance of the plot from the market place and $r^i_{\max}$ denote the maximum distance of the i^{th} crop

[5] The adoption of the technology once it is available is assumed to evolve exponentially according to $A^c = A_0^c e^{gt}$ where g is the rate of innovation and t is time required to acquire the technique. The rate of growth of technology is assumed to be a function of education over time, $g = \psi E.$

[6] This assumption is only to make the analysis simple. Practically, part of the cash crops produced by the household is consumed by the household even though it might be in small proportion compared to staple crops.

beyond which the crop cannot be sold at the market due to its perishable nature. Then, the index for the i^{th} crop is given by:

$$r^{*i} = \frac{r}{r^{i}_{\max}} \tag{4}$$

where:

$$r^{*i} = \begin{cases} 0 & if \ r = 0 \\ 1 & if \ r \geq r^{i}_{\max} \end{cases} \qquad \text{so that} \qquad r^{*}_{i} \in [0,1].$$

If the crop produced at distance r is perishable, then it loses a value of r^{*i} monetary units per unit of crop. If almost all cash crops produced and n fraction of the staple crop are intended to be sold at their respective prices, and if all staple crops are not perishable, then the associated total cost incurred can be summarized by:

$$C = \left(q^{c} + nq^{s}\right)kr + r^{*c}p^{c}q^{c} + w^{c}l^{c} + w^{s}l^{s} \tag{5}$$

Given the revenue function in Equation (3) and the cost function in Equation (5), the profit π of the household is, therefore, given by:

$$\pi_{t} = A_{0}^{c}e^{\psi E} f\left(l^{c}\right)\left(p^{c} - kr - r^{*c}p^{c}\right) + g\left(l^{s}\right)\left(p^{s} - nkr\right) - w^{c}l^{c} - w^{s}l^{s} \tag{6}$$

3.3 The problem of the household

The household maximizes profit according to:

$$\max_{l^{c}} \pi = A_{0}^{c}e^{\psi E} f\left(l^{c}\right)\left(p^{c} - kr - r^{*c}p^{c}\right) + g\left(l^{s}\right)\left(p^{s} - nkr\right) - w^{c}l^{c} - w^{s}l^{s} \tag{7}$$

Taking the first order derivatives with respect to proportion of land under cash crop, the first order condition is:

$$\frac{d\pi}{dl^c} = A_0^c e^{\psi E} f'\left(l^c\right)\left(p^c - kr - r^{*c}p^c\right) - g'\left(l^s\right)\left(p^s - nkr\right) - w^c + w^s = 0$$

This can be rearranged to give:

$$p^c A_0^c e^{\psi E} f'\left(l^c\right) = \left[A_0^c e^{\psi E} f'\left(l^c\right)\left(kr + r^{*c}p^c\right) + w^c\right] + \left[g'\left(l^s\right)\left(p^s - nkr\right) - w^s\right] \tag{8}$$

The left hand side of Equation (8) is the value marginal product of land in the production of cash crops. The first term of the right hand side of the equation in square brackets is the marginal cost of producing and selling cash crops. The term in the second square bracket denotes the opportunity cost of production of cash crops at the net margin. In general, this condition says that an optimum allocation of the available plot of land between cash and staple crops ensures that the marginal product of land in the production of cash crops equals the foregone value of the marginal product of staple crops net of marginal costs of production in the alternative use plus direct marginal costs.

It can be shown that the second order derivative of the profit function with respect to plot of land allotted for the production of cash crops is negative.

$$\frac{d^2\pi}{d\left(l^c\right)^2} = A_0^c e^{\psi E} f''\left(l^c\right)\left(p^c - kr - r^{*c}p^c\right) + g''\left(l^s\right)\left(p^s - nkr\right) < 0$$

By the assumption of diminishing returns to scale, $f''\left(l^c\right)$ and $g''\left(l^s\right)$ are negative. The household produces cash crop if his optimization condition ensures that unit profits are greater than unit costs so that $p^c > \left(kr + r^{*c}p^c\right)$ and sells his staple crop if $p^s > nkr$. This implies that the second derivative is negative. Thus, the sufficient condition for maximization of profit is met. Note that the second order derivative becomes positive if r^{*c} is unity, that is if $r \geq r^c_{max}$. Nonetheless, at $r^{*c} = 1$, the household has no incentive to produce any cash crop as it would intuitively mean that all cash crops that have to be transported will be spoiled before they reach the market.

3.4 Comparative static analysis

In this section we examine the impact of varying the distance of producers to the urban centers and the level of education on land allocation decision. The first order condition can be re-written in the form of an implicit function *F(.)*:

$$F\left(l^c; r, r^*, E, p^c, p^s, w^c, w^s, k, n\right)$$
$$= A_0^c e^{\psi E} f'\left(l^c\right)\left(p^c - kr - r^{*c} p^c\right) - g'\left(l^s\right)\left(p^s - nkr\right) - w^c + w^s = 0 \qquad (9)$$

By totally differentiating the implicit function, we have:

$$dF = \left\{A_0^c e^{\psi E} f''\left(l^c\right)\left(p^c - kr - r^{*c} p^c\right) + g''\left(l^s\right)\left(p^s - nkr\right)\right\} dl^c$$
$$+ \left\{k\left[A_0^c e^{\psi E} f'\left(l^c\right) - ng'\left(l^s\right)\right]\right\} dr$$
$$+ \left\{A_0^c \psi e^{\psi E} f''\left(l^c\right)\left[p^c - kr - r^{*c} p^c\right]\right\} dE + \ldots = 0$$

Holding other exogenous variables constant, the change in l^c in response to a change in distance from the market is given by:

$$\frac{dl^c}{dr} = \frac{k\left[A_0^c e^{\psi E} f'\left(l^c\right) - ng'\left(l^s\right)\right]}{J}$$

where:

$$J = \left\{A_0^c e^{\psi E} f''\left(l^c\right)\left(p^c - kr - r^{*c} p^c\right) + g''\left(l^s\right)\left(p^s - nkr\right)\right\}.$$

Basically, J is the second order derivative of the profit function with respect to l^c which is negative. In the numerator, $f'\left(l^c\right)$ and $g'\left(l^s\right)$, are positive by assumption. We assume that the marginal product under cash crop production $\left(A_0^c e^{\psi E} f'\left(l^c\right)\right)$ is greater than the n fraction of the marginal productivity of land for the production of staple crop, $\left(ng'\left(l^s\right)\right)$. This implies that the term in the numerator is greater than zero. Hence, we have:

$$\frac{dl^c}{dr}=\frac{k\left[A_0^c e^{\psi E} f'\left(l^c\right)-ng'\left(l^s\right)\right]}{J}<0$$

That is, a unit variation in location across plots in relation to markets in the direction away from such markets leads to a decline in the share of land under cash crop production. Similarly, the direction of the impact of the index for the perishable nature of a cash crop can be shown to be negative. The higher the index (i.e. the more perishable the crop is), the less proportional land to be allotted for the production of the particular cash crop.

$$\frac{dl^c}{dr^*}=\frac{p^c A_0^c e^{\psi E} f'\left(l^c\right)}{J}<0.$$

The direction of the impact of other exogenous variables can be determined as well. For instance, the effect of education on crop choices can be shown to favor the allocation of more land for the production of cash crop. After totally differentiating Equation (9) and rearranging we get:

$$\frac{dl^c}{dE}=\frac{-A_0^c\psi e^{\psi E}\left\{f'\left(l^c\right)\left[p^c-kr-r^*p^c\right]\right\}}{J}>0.$$

which is positive. As it has been shown already, J is less than zero, while in the numerator, the term in the square bracket is positive. That is, for the household to engage in the production of cash crops, the unit price p^c must be greater than the unit costs associated with transport. This holds even without considering other costs of production. The negative sign multiplying the whole numerator turns it to negative giving rise to the overall expression to be greater than zero. The result can be interpreted such that an increase in the level of education, say by a year of schooling, increases the proportion of land under cash crop cultivation.

4. Empirical Analysis

4.1. The model

The theoretical framework that has been considered in Section 3 suggests that a household's decision to allot a plot of land to cash crop production in an attempt to maximize household income is by and large a function of, among others, distance from the market (usually urban centers), and level of education. There are, however, other factors which are deemed to be important in affecting crop choice. These include access to irrigation scheme, climatic conditions, wealth of the household, input availability, soil type, and others. Some cash crops such as sugarcane are water intensive and its production presupposes availability of irrigation scheme. Areas with irregular rainfall may not specialize in cash crop production. Moreover, wealthier households are highly likely to afford relatively higher initial investments in cash crops. A model that can accomodiate some of these factors for given prices p^c and p^s, and costs, can be given by:

$$L^c = f\left(R_i, AR_i, E_i, DI, DC, W, DR\right) \tag{10}$$

Where L^c = proportion of land allocated for cash crop mostly fruits, vegetables and stimulants, R_i = distance of the plot from market centers, AR_i = access to road, E_i= level of education of the agent, DI_i= dummy for access to irrigation, DC = dummy for climate, W_i= wealth of the household, and DR_i= dependency ratio. It is expected that R, and DR would affect L^c negatively while other variables except DC affect it positively. The impact of climate on allocation of land for cash crops depends on the particular cash crop, whereas in the Ethiopian context, areas with cold climate tend to specialize less on cash crops.

In this section, we test the hypothesis that proximity to urban centers influences crop choice by applying a fractional logit model. In a second step, we estimate an income function using land under cash crops and staple crops as explanatory variables.

In the crop choice model, the dependent variable is land under cash crop in proportion to total land size. The explanatory variables include distance from urban centers, access to roads linking to urban centers which in this case is measured by the inverse of the distance from road accessible by vehicle in kilo meter, total land endowment, level of education of the head of the household, a dummy for climate, and a dummy for whether a household possesses irrigable land. Size of own plot, and size of land used under share cropping arrangements are also considered.

Obviously, OLS procedures are not appropriate when the dependent variable is a ratio bounded between 0 and 1. Running OLS on a fractional dependent variable would entail similar problems as it does in the linear probability model for strict binary cases (Wooldridge, 2002). One of the drawbacks of this approach is that predicted values of OLS estimates would not necessarily lie in the [0,1] interval. The other important advantage of using fractional logit model over OLS is that the first accounts for possible non-linear relationship in the model.

A common approach to model dependent variables which are bounded between 0 and 1 is a logistic transformation where the log-odds ratio is modelled as a linear function of a set of independent variables. Unfortunately such procedure does not account for data that includes the limits 0 and 1. Moreover, it is not possible to recover the predictions for the dependent variable without some simplifying assumptions. In our case, though a value of 1 is rare, there are a number of households who do not allot their plots for cash crop at all. One way out could be to proceed with such transformation by giving an extremely small number for values equal to zero and a near unity number for values of 1. This is, however, arbitrary which may lead to undesirable results (Wooldridge, 2002).

Papke and Wooldridge (1996) based on the results of Gourieroux, Monfort, and Trongen (1984) and McCullagh and Nelder (1989) suggested as an alternative the Generalized Linear Model (GLM) that makes use of quasi-maximum likelihood estimation procedures.

The notion of the GLM is that a regression model can be decomposed into a random component with expected value and variance of the dependent variable, a systematic component that is predicted by covariates, and a link function that relates the systematic component to the random component. For classical regression models, the random component is assumed to be distributed normal and the link function is an identity in the sense that the random and systematic components are identical (McCullagh and Nelder, 1989).

What makes GLM more relevant is that the normality assumption on the distribution of the random component could come from any function of the exponential family, and the link function could be any monotonic differentiable function (McCullagh and Nelder, 1989).

Given the dependent variable l_i and the vector of the various explanatory variables x, where $0 \leq l^c \leq 1$. Then, for all i:

$$E\left(l_i^c\right) = x_i\beta \tag{11}$$

In this case, the random component, $E\left(l_i^c\right)$, is expected to have a value of μ so that $0 \leq \mu \leq 1$, and, unlike the linear regression model, the random component could have a distribution different from normal. It might rather have a binomial distribution which is from the exponential family.

More importantly, the link function cannot be assumed to be identity because the systematic component ($x_i\beta$) does not ensure the condition that the random component, $E\left(l_i^c\right)$, lies between 0 and 1. Hence, the link function that relates $E\left(l_i^c\right)$ and ($x_i\beta$) could be given by:

$$E\left(l_i^c \mid x_i\right) = g\left(x_i\beta\right) \tag{12}$$

where g (.) is a link function satisfying the condition that $0 \leq g(.) \leq 1$.

Gourieroux, Monfort, and Trongen (1984) showed that quasi-maximum likelihood estimators (QLME)[7] are consistent as long as the likelihood function is in the linear exponential family and that the link function under (12) holds. Papke and Wooldridge (1996) suggested the random component to be Bernoulli for it being easy to maximize. For the link function, we use the logistic distribution as suggested by McCullagh and Nelder (1999).

Thus, for $l_i^c \sim Bernoulli$ with a logistic link function, we have:

[7] Quasi-maximum likelihood estimators, also known as pseudo-maximum likelihood estimators, are methods which maximize probability distributions which do not necessarily contain the true distribution.

$$g\left(x_i\beta\right) \equiv \Lambda\left(x_i\beta\right) = \frac{e^{x_i\beta}}{\left[1+e^{x_i\beta}\right]} \tag{13}$$

The Bernoulli likelihood function is given by:

$$f\left(l_i^c / x_i;\beta\right) = \left[\Lambda\left(x_i\beta\right)\right]^{l_i^c}\left[1-\Lambda\left(x_i\beta\right)\right]^{1-l_i^c}, \text{ where } l_i^c \in \left[0,1\right].$$

This can be transformed to give:

$$L\left(\beta\right) = l_i^c \log\left[\Lambda\left(x_i\beta\right)\right] + \left(1-l_i^c\right)\log\left[1-\Lambda\left(x_i\beta\right)\right], \tag{14}$$

The other model considered in this section is the income function of rural households. The estimable model is given by:

$$y_r = f\left(L_i^c, L_i^s, N, O, DI, E_h, DR\right) \tag{15}$$

where y_r = household per capita income from crop production, L_i^c = land under cash crop, L_i^s = land under staple crop, N = labor, O = number of oxen, DI = dummy for availability of irrigable land, Eh = education level of the head of the household, and DR = dependency ratio. The model in Equation (15) is estimated by OLS.

4.2 The data and estimation results

As it has been introduced in Section 2, the data used for this study is the household survey data collected from six villages in Wollo, the Amhara Regional State. The survey was conducted in the year 2006. The villages were systematically selected based on their distance from major towns. 252 households were randomly selected from the villages.

In the crop choice model, distance from town is approximated by the distance in kilometer between what is thought to be 'centroid' of the village to the nearest district town. Distance from road is the distance in kilometer of the village from the nearest road accessible by vehicles. We defined access to road as the inverse of the distance from the nearest road accessible by vehicle.

Table 1: List of Variables used in the Estimation

Variable	Mean	Standard Deviation	Min	Max
Land under cash crop (ratio to the total)	0.12	0.15	0	1
Town-Distance	11.86	5.51	4	20
Distance from Road	3.39	3.80	1	10
Access to Road (inverse of distance)	0.69	0.39	0.1	1
Dummy Irrigation	0.42	0.50	0	1
Dummy Climate (=1 if Dega)	0.29	0.46	0	1
Education - Head				
Years of Schooling	2.06	2.99	0	11
Primary (1-6)	0.43	0.50	0	1
Junior Secondary (7-8)	0.06	0.23	0	1
Senior Secondary (9-12)	0.03	0.17	0	1
Total Own Land in hectare	0.72	0.39	0	2.5
Land Leased in for share cropping (LSC1)	0.21	0.38	0	3
Land Leased out for share cropping (LSC2)	0.02	0.10	0	0.75
Dependency Ratio	0.77	0.80	0	4
Permanent Cash Income (per capita)	30.92	157.57	0	1600
Value of Livestock (per capita)	1220.42	1349.76	0	9250
Per Capita Income (logs)	6.68	0.85	0.37	8.77
Land under cash crop	0.11	0.14	0.00	0.50
Land under staple crop	0.72	0.41	0.09	2.25
Labor	2.33	1.02	1	7
Oxen	1.61	1.25	0	9
Cattle (other than oxen)	2.47	2.47	0	12
Dummy Rural Enterprise	0.19	0.40	0	1

The dummy variable for availability of irrigation scheme takes a value of 1 if the village has access to irrigation facilities (modern or traditional) at a significant scale and 0 otherwise. The dummy for climate assumes a value of 1 if the village has cold (dega) climate which in this case ranges from 2600-2800 meters above sea level in elevation and 0 if it has moderate (woina-dega) climate. The elevation in the latter category ranges from 1400 meter for Girana to 1800 meter for Habru-Ligo.

Per capita cash income and per capita value of livestock[8] are included to capture the impact of wealth on crop choice. To account for liquidity constraints, we include per capita value of permanent cash income which includes pensions, permanent remittances, and salaries from long-term off-farm employments. Value

[8] Similarly, Dercon (1996), and Kurosaki and Fafchamps (2002) used the value of livestock as a proxy for liquid wealth in their crop choice model.

of livestock is the sum of the average market price of cattle, goats, sheep, and camels.

Livestock ownership may have two opposing impacts on crop choice. On one hand, livestock serve as buffer stock against risk in which case it favors the allocation of more land for cash crop production. On the other hand, livestock farming might be a competing activity to cash crop production. The relative importance of the two effects depends on village specific factors such as distance from urban centers. To disentangle the two effects, we used an interaction variable of distance from urban centers and value of livestock.

For the educational attainment of the head of the household, years of schooling by level (primary, junior secondary and senior secondary levels in which the head has attended some classes) were considered. The maximum year of schooling is 11 years. A dummy is used for each level where a value of 1 denotes some education at the respective level and 0 otherwise. The omitted category is 'never attended any of these levels'. Own land is the size of plot in hectares that belongs to the household. Size of land under sharecropping arrangements is also included as well as a dummy for whether a household has some plots of land that is adapted to irrigation irrespective of whether the plot is irrigated during the survey period. Many households implanted irrigation schemes but do not necessarily irrigate their plots depending on the season and the type of crop.

A potential source of endogeneity bias arises from liquid assets. Non-agricultural cash income is exogenous because pensions, remittances, and compensations for long term off-farm activities may not be expected to be affected by crop choice decisions. However, the simultaneity problem may arise in the case of value of livestock. Dercon (1996) reports simultaneity between crop choice and value of livestock. On the other hand, Kurosaki and Fafchamps (2002) find that liquid assets and livestock are predetermined and conclude that these variables are exogenous.

In our case we applied a Hausman test to check whether value of livestock is exogenous[9]. The instruments used were total land size, number of oxen, and labor. The test does not support the null that value of livestock is endogenous.

[9] We estimated an auxiliary regression where per capita value of livestock was regressed on total land, labor, and oxen.

$$\underset{(3.68)}{PCVL = 848.03} + \underset{(3.50)}{719.36Land} + \underset{(5.46)}{349.80Oxen} - \underset{(-4.07)}{304.89Labor}$$

We estimated the crop choice model by including the residual of the auxiliary regression along with the per capita value of livestock (Wooldridge 2002). We found that the coefficient of the residual term was not statistically significant indicating that the case of simultaneity is not supported.

4.2.1 Results for crop choice model

The results for the land allocation model are shown in Table 2. In most cases, slopes of the GLM estimates and OLS parameter estimates are not very different both in terms of magnitude and their statistical significance. The results show that proximity to town, access to road, education of the head, ownership of liquid assets and access to irrigation scheme are significant for predicting household crop choices. Rural households under study who operate nearer to urban centers tend to allot more land for the production of cash crops while those households who operate far from urban centers tend to allocate much of their land for the production of staple crops (grains). This might be due to the fact that rural peasants nearer to urban centers have a greater advantage in terms of transportation cost and information about the market. These results support the argument that for crop choices the location of the village relative to the next market matters.

Table 2: GLM Estimation of Land Allocation Decisions

Dependent Variable: Share of Land Allotted to Cash Crop							
	GLM Estimates				OLS Estimates		
	Coefficient		Slope				
Distance-Town	-0.180	[-5.92]***	-0.013	(-6.22)	0.011	(-4.12)	[-4.50]
Access to Road	1.689	[5.64]***	0.125	(5.42)	0.091	(3.20)	[3.21]
Dummy Irrigation	0.787	[2.51]**	0.062	(2.56)	0.067	(3.02)	[2.66]
Dummy Climate	-1.585	[-3.78]***	-0.094	(-3.97)	-0.140	(-4.47)	[-3.41]
Cash Income	0.0009	[3.21]***	7×10^{-5}	(2.89)	0.0002	(3.07)	[2.33]
Livestock (Value)	-0.0005	[-3.08]**	-4×10^{-5}	(-3.08)	-4×10^{-5}	(-2.82)	[-3.56]
VLS×r	5×10^{-5}	[3.21]***	3.4×10^{-6}	(3.26)	3.5×10^{-6}	(2.33)	[3.11]
Education-Head							
Primary (1-6)	0.391	[2.43]***	0.030	(2.31)	0.032	(1.86)	[1.95]
Junior Sec. (7-8)	0.904	[2.65]**	0.094	(1.94)	0.103	(2.94)	[2.43]
Senior Sec.(9-12)	0.398	[0.97]	0.035	(0.83)	0.048	(0.98)	[1.15]
Total Own Land	0.185	[0.89]	0 .014	(0.91)	0.013	(0.59)	[0.58]
LSC1	-0.206	[-1.07]	-0.015	(-1.05)	-0.038	(-1.62)	[-2.01]
LSC2	0.817	[1.80]*	0.061	(1.79)	0.144	(1.69)	[1.59]
Dependency Ratio	-0.243	[-1.89]*	-0.018	(-1.85)	-0.022	(-2.14)	[-2.09]
Intercept	-1.375	[-2.48]**	-		0.201	(3.55)	[3.44]
N	252						
R^2						0.39	
$\bar{R}^2$						0.35	
Joint Stability					F(14,237)	10.59	21.48
Heteroscedasticity					χ^2 (1) =	26.50	

**** Significant at 1% level, ** Significant at 5% level, * Significant at 10% level.*
Figures in brackets are t-ratios and those in square brackets are robust t-ratios.

The irrigation dummy is significant and positive. Irrigation may have two impacts. First, most cash crops which have high demand in the urban market require a sustainable supply of water. As it has been indicated in Section 2.3, major cash crops that are produced include sugarcane, and fruits whose production is water intensive. Secondly, availability of irrigation scheme gives households the opportunity to produce more than once within a year. This in turn secures them to shift into the production of staple crops with low gestation period during a risk of falling prices of cash crops such as vegetable.

In the case of liquid asset, estimation results without the interaction variable (VLS×r), value of livestock was found to be insignificant while permanent cash income reveals a positive and significant coefficient. Upon the introduction of the interaction variable, both permanent cash income and value of livestock were significant the latter having a negative coefficient. The interaction variable itself has a positive and significant coefficient.

It can be shown from the coefficients of value of livestock and interaction variables that within about 18 kilometers radius from market centers, the rivalry effect of cash crop production and livestock farming dominates[10]. Beyond 18 kilometers radius, the role of livestock as a buffer stock against risk dominates in that households with more livestock tend to allot land for cash crop. One explanation for positive association between cash crop production and value of livestock might be that remote villages have significant land that is not arable but which can be used for livestock farming. Hence, livestock farming does not necessarily compete with crop production in terms of land use.

In general, education of the head is positively associated with a higher probability of allocating more land to cash crops. Education on primary and junior secondary levels has positive impact. However, additional schooling to senior secondary schooling does not have much influence on the household's decision to allot more land to cash crops. The negative sign of the dummy for climate shows that highlanders of the villages under survey do not allot much land to cash crop compared to lowlanders. The coefficients and slopes for total own land, and land under sharecropping arrangements are not statistically significant. Land leased out in the form of share cropping arrangements is significant only at 10 percent level of significance.

[10] We calculated the threshold distance (= 18 km) by differentiating the land allocation equation with respect to value of livestock and set to zero. We used the slope coefficients of the GLM estimates for this purpose.

Lastly, the dependency ratio (proportion of members of a household below the age of 10 and above the age of 65 to the active labor force) is found to be significant only at 10 percent level in the case of GLM estimation but significant at 5 percent in the case of OLS estimates. Households with a higher share of dependants might be more risk averse and hence do not tend to allot more land for cash crop as they prefer food security.

4.2.2 Results for incomes function

To investigate whether distance predicts income we use annual per capita income in Birr from agricultural activities, in particular cash and staple crop production as the dependent variable. On the right hand side we include the distance variables along with size of land under cash crop and staple crops as separate variables as well as a number of further controls. Head counts are used for oxen. In the case of labor, a sort of adult equivalent labor is used. Household members aged 16 and above are given a weight of 1 while those in the age of 10 to 15 are given a weight of 0.5. Some variables which were used as determinants of land allocation decision are also used in estimating the income function. The rationale of including the variables which were used as determinants of land allocation decision (dummy for irrigation scheme, and education) is to see their direct effect on income apart from their impact on it through land allocation decision.

Results are summarized in Table 3. The null for constant variance under the Breusch-Pagan test for heteroscedasticity was rejected at 5 percent level. However, there was little change in the standard errors between the OLS and robust estimates causing no change in significance of coefficients at 5 percent level. The estimates revealed that coefficient for land under cash crop was significantly greater than that of the land under staple crop reflecting that the marginal product of land under cash crop is greater compared to its alternative use of staple crop production. More importantly, distance from the nearest urban center is found to significantly predict the level of per capita income of households. It shows that, other things being equal, households operating far from urban centers tend to have lower per capita income compared to those households nearer to towns.

Table 3: OLS Results of Rural Per Capita Income: Land being instrumented

Dependent Variable: Per capita Household Income (in logs)			
Covariates	Coefficients	t-ratios	
Land under cash crop(Estimated)	1.13	(2.13)	[3.11]
Land under staple crop (Estimated)	0.46	(3.42)	[3.22]
Labor	-0.10	(-2.35)	[-2.47]
Oxen	0.16	(4.12)	[4.21]
Dummy for Irrigation	0.36	(3.20)	[3.67]
Distance from Town	-0.03	(-3.15)	[-3.31]
Access to Road	0.32	(2.20)	[2.56]
Education - Head			
Primary (1-6)	0.02	(0.27)	[0.28]
Junior (7-8)	0.17	(0.86)	[0.79]
Secondary (9-12)	-0.06	(-0.23)	[-0.43]
Dummy for Rural Enterprise	0.31	(0.89)	[3.30]
Dummy food for Work	0.09	(3.04)	[1.01]
Intercept	6.11	(24.00)	[23.72]
N		252	
R^2		0.49	
$\bar{R}^2$		0.47	
F(12, 239)		19.28	29.43
RESET: F(3, 236)		1.28	
Heteroscedasticity: $\chi^2(1)$		4.28	

Figures in brackets are t-ratios and those in square brackets are robust t-ratios

5. Conclusions

In this article, we investigated the interaction between distance to markets and crop choice in Ethiopia. We found that proximity to urban centers and access to roads increases the share of land allotted to cash crop production. Shorter ways of bringing the produce to the market imply lower transaction costs and consequently better returns. Another channel through which market proximity may affect crop choices is better access to information about prices or new technologies. Furthermore, households located closer to urban centers with access to road but who do not have irrigable land tend to invest in commercial livestock farming and fast growing trees such as eucalyptus to be sold in urban centers. This translates into uneven levels of per capita income among villages: a typical household in the richest village nearer to urban center has a per capita income more than 4 times that of a typical household who lives in the remotest village among those covered by the study.

Estimation results of the income function of rural household show that size of plots under cash crops and staple crops are significantly related to higher incomes. The coefficient of land under cash crop is by far greater than that of land under staple crop. Distance from the nearest urban center is found to be significant and negative in the incomes function implying that level of per capita income varies over such distances where the households with relative proximity to urban centers are better off.

In conclusion, strong linkages to the urban sector matter for the development prospects of rural areas. Policies that target on supply bottlenecks in the agricultural sector might not be successful without vibrant urban centers which constitute sustainable demand for marketable surplus. In a rural economy such as that of Ethiopia which is characterized by fragmented and static urban enclaves, encouraging township could be considered as a priority. Moreover, enabling rural households to have access to road and better information networking, expanding purposeful education, developing irrigation schemes, introducing new varieties of high yield cash crops including for cold climate zones might help rural households better cope up with shocks and enable them to create surplus that would serve as a basis for agrarian transformation.

References

Ayalew, Daniel. 2003. Essays on Household Consumption and Production Decisions under Uncertainty in Rural Ethiopia, Ph.D. Dissertation, Katholieke Universiteit Leuven.

Becker, S. Gary. 1993. *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education*, The University of Chicago Press.

CSA: Central Statistical Agency of the Federal Democratic Republic of Ethiopia, Annual Statistical Abstract, Various Issues.

Dercon, Stefan. 1996. "Risk, Crop Choice and Savings: Evidence from Tanzania," Economic Development and Cultural Change, Vol. 44, No. 3, pp. 485-514.

Dillon, John L. and Pasquale L. Scandizzo. 1978. "Risk Attitudes of Subsistence Farmers in Northern Brazil: A Sampling Approach," *American Journal of Agricultural Economics*, Vol. 60, No. 3, pp. 425-435.

Fafchamps, Marcel. 1992. "Cash Crop Production, Food Price Volatility, and Rural Market Integration in the Third World," *American Journal of Agricultural Economics*, pp. 90-99.

Fafchamps, Marcel and Agnes R. Quisumbing. 1999. "Human Capital, Productivity, and Labor Allocation in Rural Pakistan," *The Journal of Human Resources*, Vol. 34, No. 2, pp. 369-406.

Fujita, Masahisa and Jacques-Francois Thisse. 1986. "Spatial Competition with a Land Market: Hotelling and Von Thunen Unified." The Review of Economic Studies, Vol. 53, No. 5, pp. 819-841.

_______. 2002. *Economics of Agglomeration: Cities, Industrial Location, and Regional Growth*, Cambridge University Press.

Gourieroux, C. A. Monfort and A. Trognon. 1984. "Pseudo Maximum Likelihood Methods: Theory," *Econometrics*, Vol. 52, No. 3, pp. 681-700.

Haddad, Lawrence, John Hoddinott, and Harold Alderman. 1997. *Intrahousehold Resource Allocation in Developing Countries*. The Johns Hopkins University Press.

Jones, Charles I. 2003: *Introduction to Economic Growth*, New York, W.W. Norton and Co. Second Edition.

_______. 2004. Growth and Ideas, NBER Working Paper No. 10767.

Kurosaki, Takashi and Marcel Fafchamps. 2002. "Insurance Market Efficiently and Crop Choices in Pakistan," *Journal of Development Economics*, Vol. 67, No. 2, pp. 419-453.

McCullagh, P. and J. A. Nelder. 1999. *Generalized Linear Models; Monographs on Statistics and Applied Probability* 37, Second Edition, Chapman & Hall/CRC

MoFED. 2011. Ministry of Finance and Economic Development, National Income Account

Nerlove, Mark L. and Efraim Sadka. 1991. "Von Thunen's Model of the Dual Economy." *Journal of Economics*, Vol. 54, N0. 2, pp. 97-123.

Papke, Leslie E. and Jeffery M. Wooldridge. 1996. "Econometric Methods for Fractional Response Variables with an Application to 401 (K) Plan Participation Rates." *Journal of Applied Econometrics*, Vol. 11, No. 6, pp. 619-632.

Pindyck, Robert S. and Daniel L. Rubinfeld. 1998. *Econometric Models and Econometric Forecasts*; Forth Edition, Irwin Mc Graw-Hill

Ray, Debraj. 1998. *Development Economics*, Princeton University Press.

Samuelson, Paul A. 1983. "Thunen at Two Hundred." *Journal of Economic Literature*, Vol. 21, No. 4, pp. 1468-1488.

Wooldridge, Jeffery M. 2002. *Econometric Analysis of Cross Section and Panel Data*; The MIT Press, Cambridge, Massachusetts.