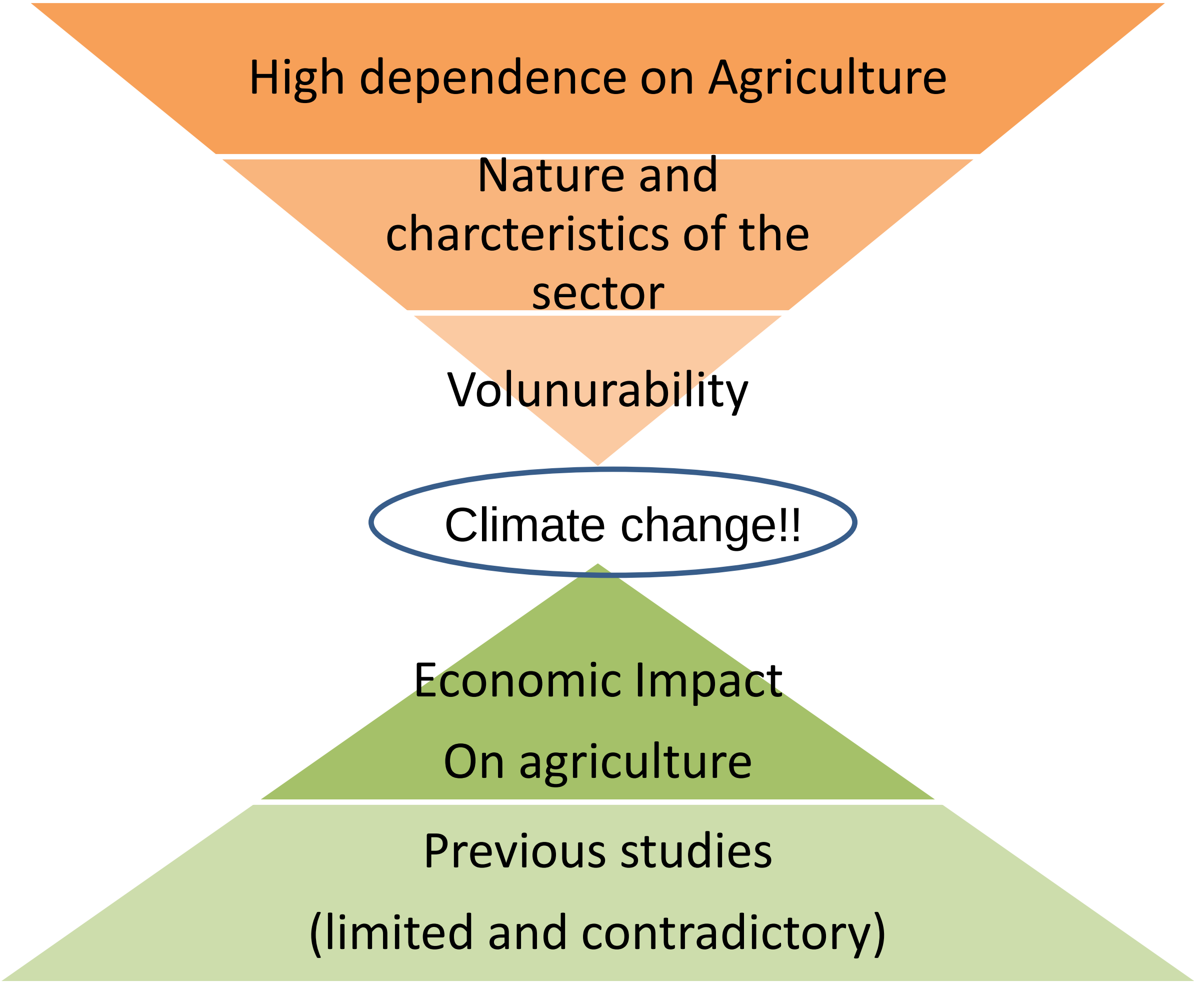
The Economic Impact of Climate Change on Agriculture in Sub-Saharan Africa: A Case of Eastern Africa

By
Goytom Abraha Kahsay
Institute of Food and Resource Economics
University of Copenhagen
PhD fellow

Contents

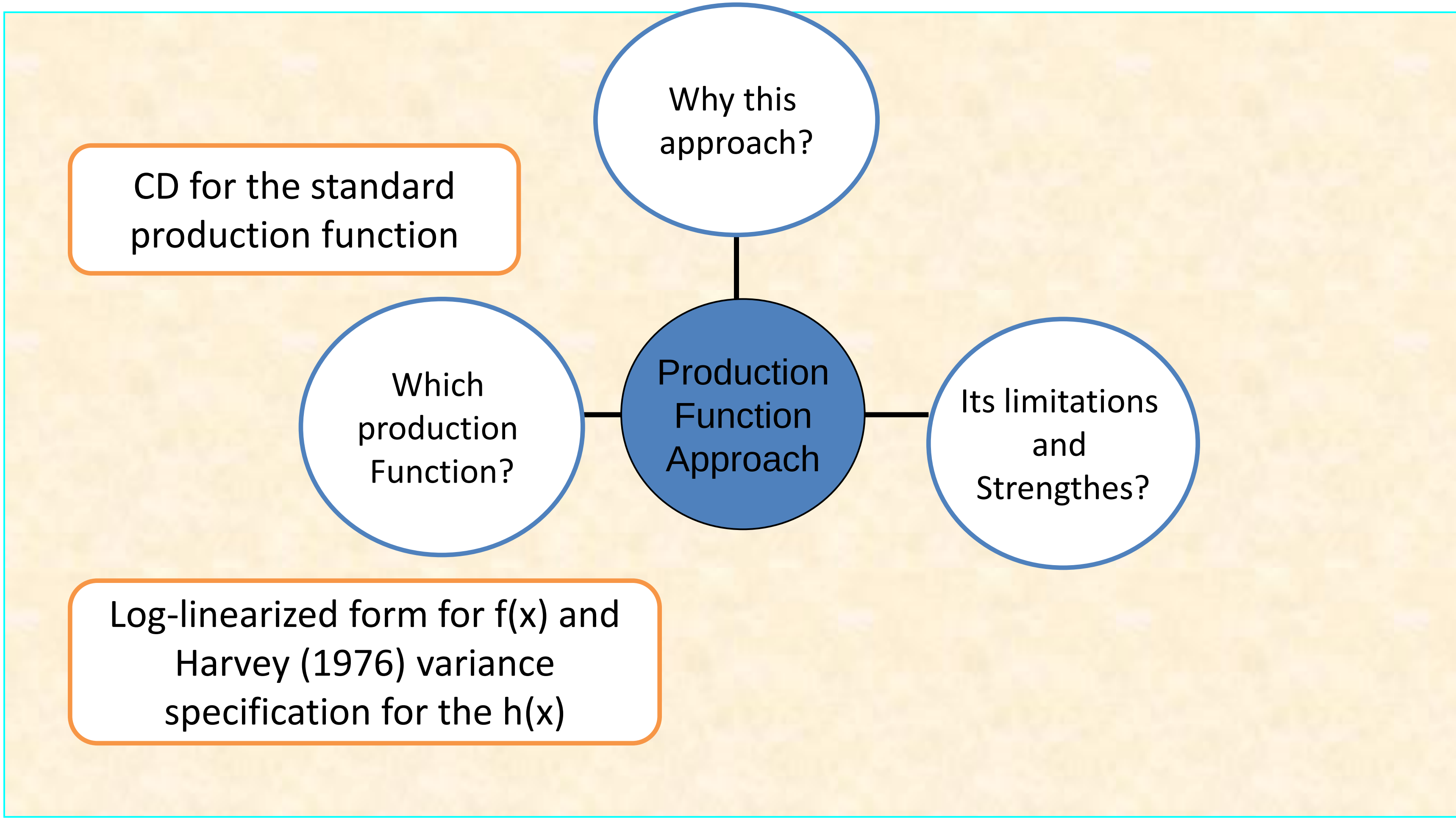
The Problem


High dependence on Agriculture
Nature and charcteristics of the sector
Volunurability
Climate change!!
Economic Impact
On agriculture
Previous studies
(limited and contradictory)

Research Questions

1. The relevance of climate?
2. The estimated and predicted economic impacts?
3. The Risk?
4. Seasonal Climate?
5. Irrigation?

Methodology



Model Specification

Output Function

$y_{it} = f(X_{it};\beta) + T_{it} + \lambda_i + \mu_{it}$,
$\mu_{it} \sim (0, \delta^2)$

What if Heteroscedasticity is Detected?

1. *It can take any form*
2. *It will be corrected using robust standard errors*

$\mu_{it} = h(X_{it};\alpha)^{0.5}\varepsilon_{it}$

$E(y_{it}) = f(X_{it};\beta) + T_{it} + \lambda_i$

Mean Output Function

1. $\hat{\mu}_{it} = y_{it} - f(X_{it};\beta) = h(X_{it};\alpha)^{0.5}\varepsilon_{it}$
2. $\hat{\mu}_{it}^{2} = (y_{it} - f(X_{it};\beta))^2 = (h(X_{it};\alpha)^{0.5}\varepsilon_{it})^2$
3. $v(y_{it}) = h(X_{it};\alpha)\,\varepsilon^2$, $\varepsilon^2 = 1$ and $\varepsilon^{2} \sim (0, 1)$

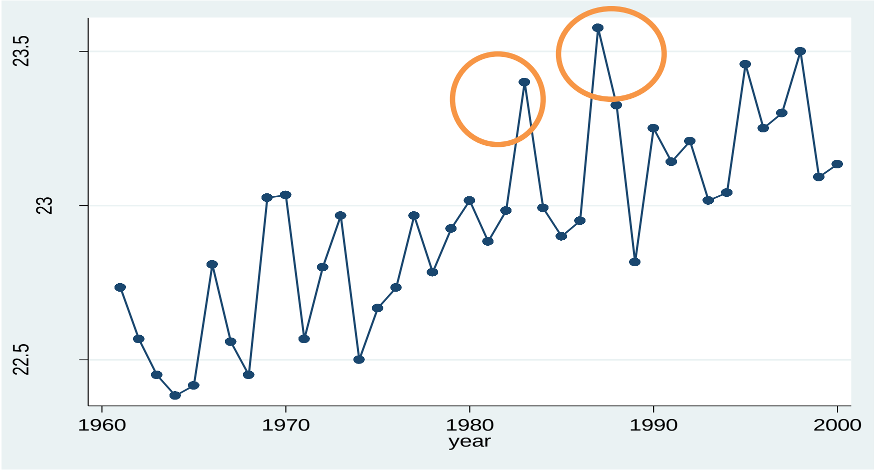
$v(y_{it}) = h(X_{it};\alpha) + T_{it} + \lambda_i$

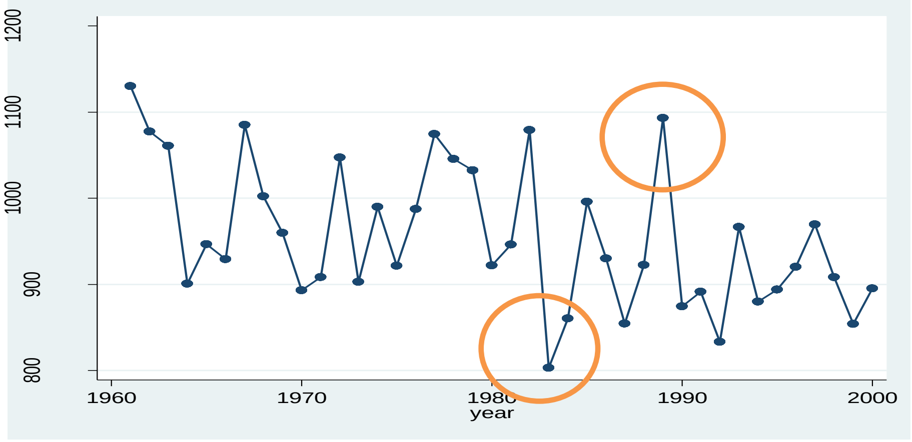
Variance Output Function

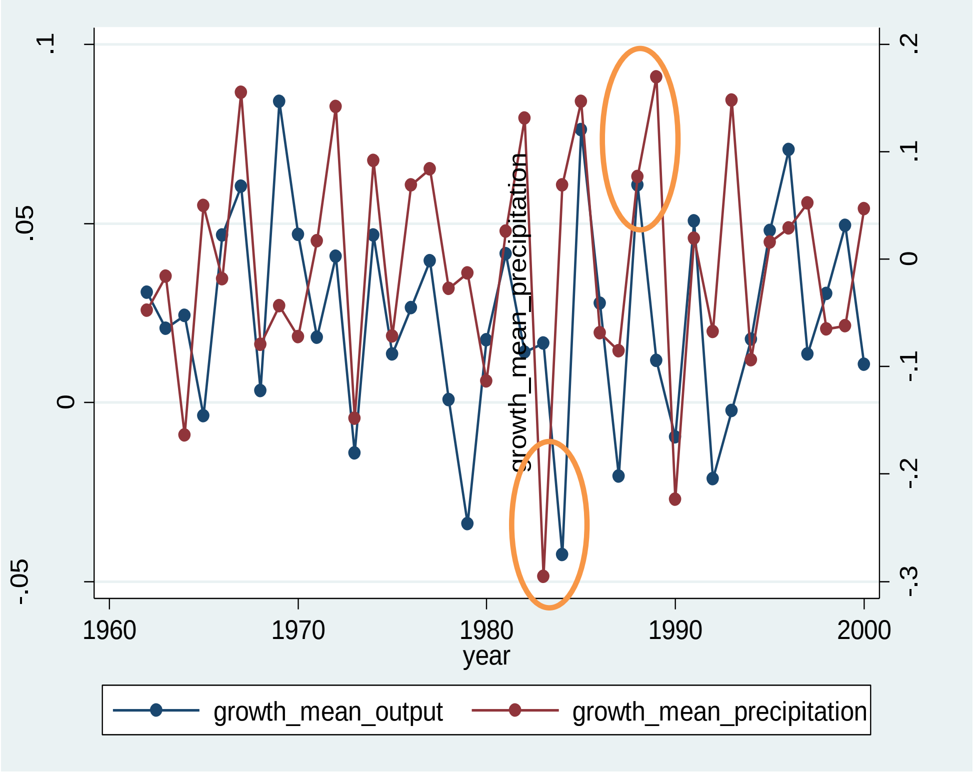
Estimation strategy

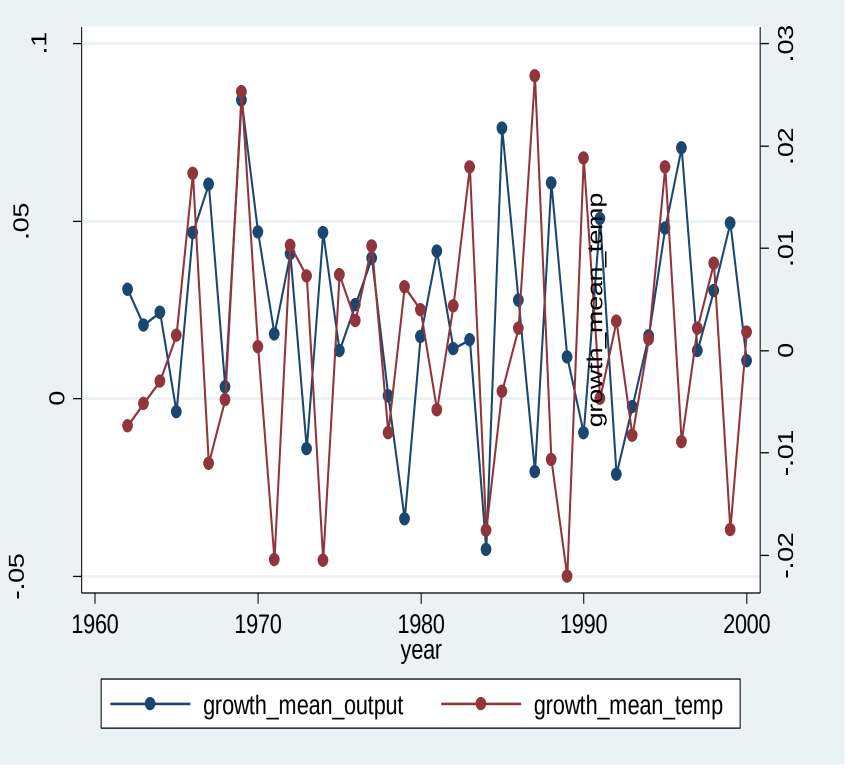
1. Fixed Effects model
2. Three stage estimation

Trend Analysis1









Analysis of Results

Pre-estimation Tests

- Panel Unit Root Tests(LLC)

Result

- The Standard Production Results
- The Stochastic Production Results

Post-estimation Tests

- Autocorrelation
- Multicollinearity tests
- Hausman Specification test
- Likelihood-ratio tests

Heteroscedasticity
Endogeneity tests
Multicollinearity tests
Test of Linear ristrictions

Variables	Standard production function ln(output)
Constant	2.104
	(6.880)
ln(capital)	0.031
	(0.049)
ln(labour)	0.429**
	(0.180)
ln(land)	0.313***
	(0.059)
ln(livestock)	0.180**
	(0.065)
ln(fertilizer)	0.055***
	(0.015)
ln(irrigation)	0.022
	(0.039)
ln(temp)	-0.466
	(0.596)
ln(precip)	0.181***
	(0.059)
T	0.003
	(0.004)
ln(temp1)	-0.411*
	(0.205)
ln(precip1)	-0.092*
	(0.050)
ln(temp1)x(lnprecip1)	0.111**
	(0.050)
Observations	463
Countries	12
F-test	265.82***
R_square	0.861

The hypothesis tests(Molua,2008)

Model	χ^2	**P value**
The relevance of irrigation(base model)	2.80	0.0941
The relevance of Climate variables(model	40.22	0.0000

The Cost!

Scenario0	Temperature	Precipitation	SD of temperature	SD of precipitation	
A B C	1.3C^0 2.3C^0 3.3C^0	15% 10% -10%	5% 10% 15%	7.5% 10% 15%	+0.8% -1.1% -6.5%

The Predicted Cost by 2050

Scenario	NPI (in billions of international dollars)	Percentage deviation	Confidence interval	
			Lower bound	Upper bound
Base	71.4	…….	70.79	72.01
A	72.6	+1.7%	71	73.21
B	72.3	+1.3%	71.69	72.91
C	68.3	-4.5%	67.69	68.91

Further Analysis

Where is the Price 

Seasonal Climate 

	Stochastic production function	
Variables	Mean output function	Variance output function
	ln(output)	ln(output)
Constant	2.536 (2.183)	.070 (.211)
ln(capital)	.030* (.016)	.000*** (.000)
ln(labour)	.428*** (.045)	.000*** (.000)
ln(land)	.324*** (.060)	.00001*** (.000)
ln(livestock)	.176 (.031)	.000*** (.000)
ln(fertilizer)	.055*** (.007)	.000*** (.000)
ln(irrigation)	.022 (.014)	-.0001*** (.000)
ln(temp)	-.582 (.401)	-.002 (.002)
ln(precip)	.209*** (.050)	.00004*** (.000)
T	.003*** (.001)	.002*** (.0001)
ln(temp1)	.024 (.040)	.008** (.003)
ln(precip1)	-.059 (.036)	-.0002*** (.000)
ln(temp1)x(lnprecip1)	....	....
Observations	463	463
Countries	12	12
F-test	246.24***	337.23***

Anyways they are almost Zero!!!

Are they Really Risk y inputs?

1. Rajsic and Weersink 2008).
2. Fufa and Hassan(2003)
3. Holst et al. (2010)
4. Just and Pope(1978, 79)
5. Cabas et al.(2010)

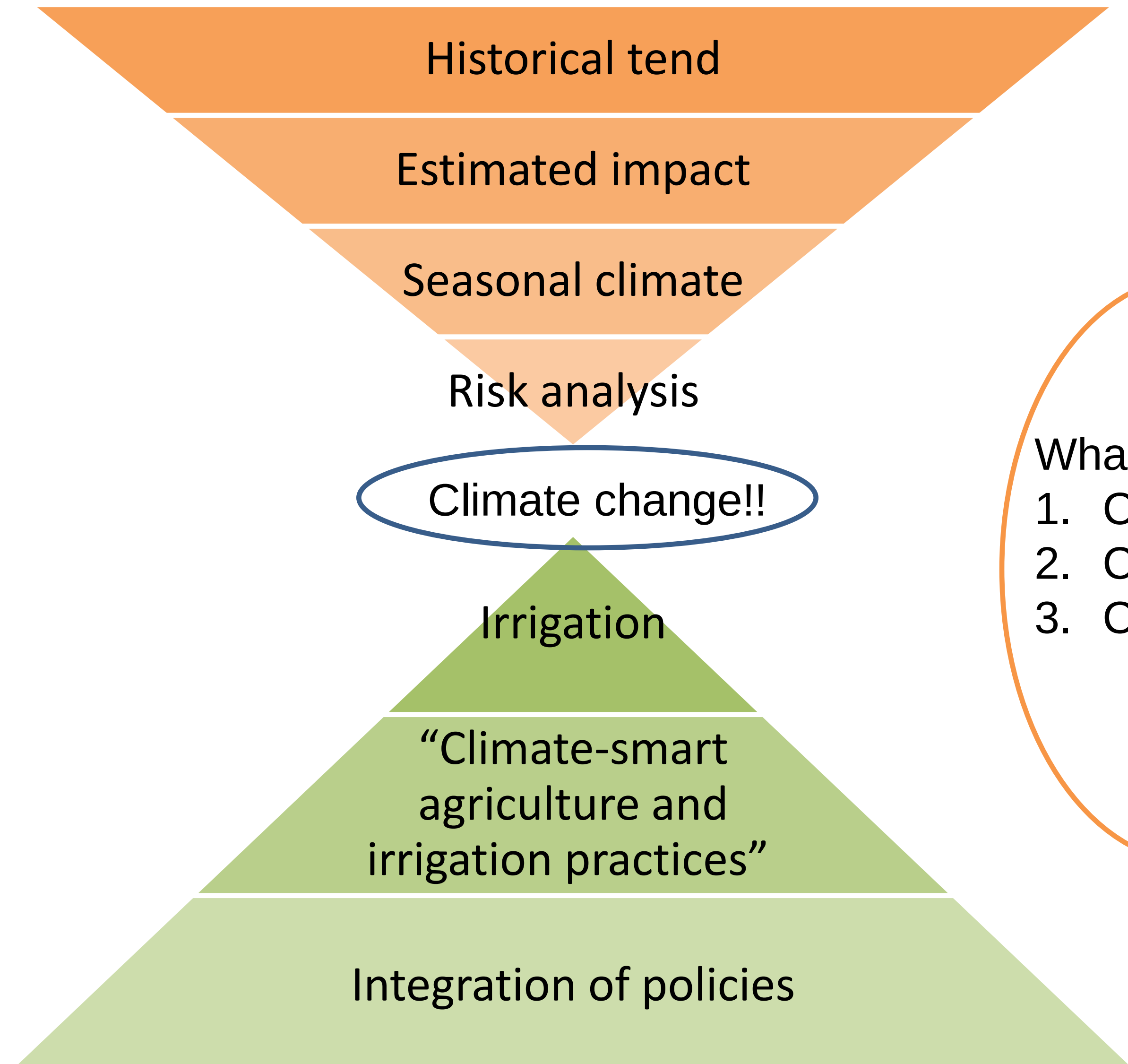
Comparisons

Previous studies in Africa

Previous studies in East Africa

- Barrios et al.(2008),
- Seitz and Nyangena (2009)

Conclusion



What is Next 

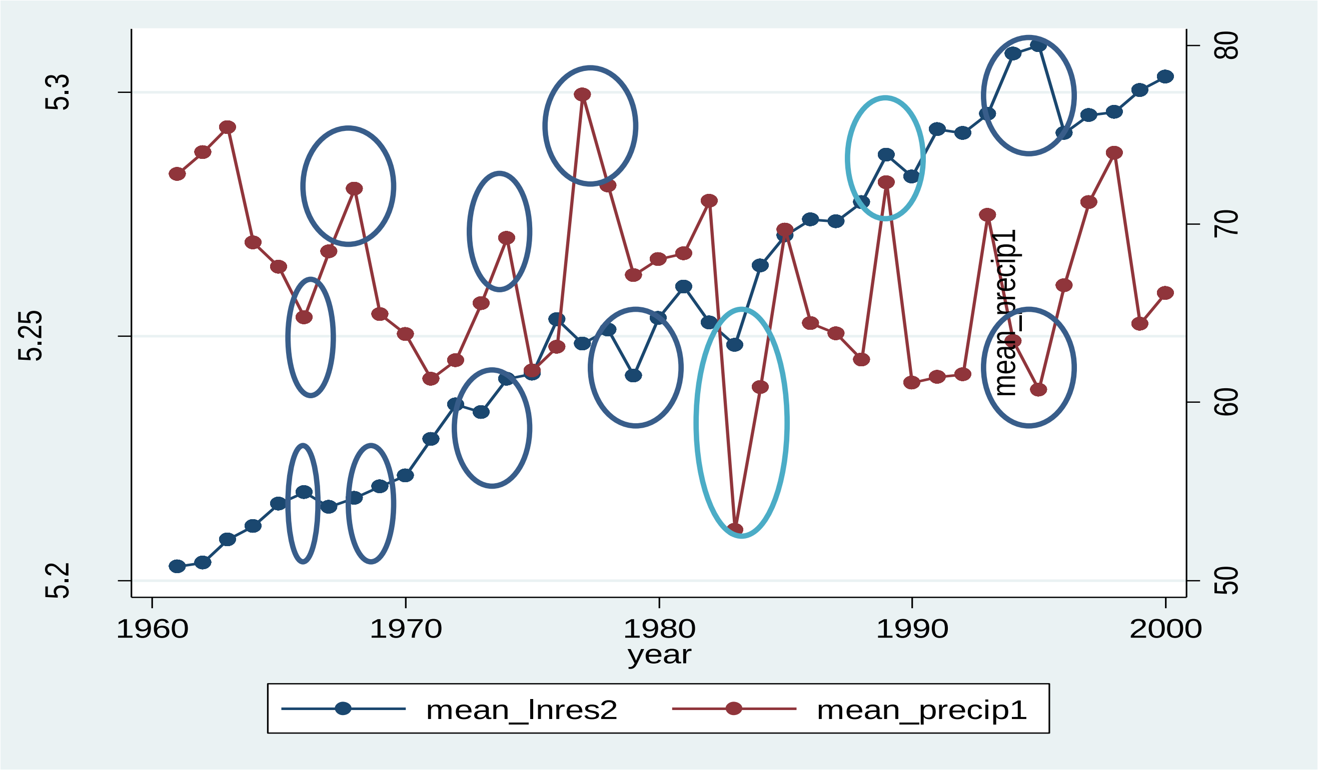
1. Crop specific
2. CGE
3. CBA

Comments

1. Approach to measure the impact of climate change
2. The esimation strategy
2. The variables included to capture climate change

Thank You!!

Discussion



5.3
5.25
5.2
80
70
60
50
mean_precip1
1960
1970
1980
1990
2000
year
mean_lnres2
mean_precip1

Coefficients, sign and magnitude

1. Barrios et al.(2008)

- Rainfall , 0.221, temp=0.75 up to 1.4

2. Seitz and Nyangena (2009)

- Rainfall, 0.17 to 0.2 , 0.2 up to 1.8, var temp 0.44 up to 2, 0.21, var precip around 0.14 up to 0.31

3. Molua(2008)

- Rain fall=-0.087, 0.17 temp=-0.065. 0.07 both are negative since they are variance (deviation variables) Molua(2008)

Climate Anomaly

- These rainfall series were constructed by assimilating measurements of rainfall from meteorological stations across the world into 0.5 degree latitude by 0.5 degree longitude grids covering the land surface of the earth. Each grid-box was then assigned to the appropriate country in order to calculate a measure of rainfall for each by using the weighted mean of the values of all grid boxes within a country. This procedure resulted in comparable mean annual measures of annual rainfall and temperature for each `country', given in millimetres and centrigrades, respectively. Where a grid box was located across more than one country, the grid box was assigned to the country with the largest stake, except where a country would otherwise have been left without any grid box. Weighting was essential since the spatial areas represented by each grid box differ in latitude.

Confidence interval

1. The predicted ln(output) has a confidence interval between 13.80 and 14.03 for the base model regression and between 13.78 and 14.05 for the base model regression by dropping the insignificant variables
2. The confidence interval for the 2050 levels

Alternative confidence intervals using standard error θ^0 from a regression with all the variables of the base model included imply that we are 95% confident that the predicted agricultural output will be between 67.2 and 78, 66.9 and 77.7, 62.9 and 73.7 for scenario A, B, and C respectively. Moreover, the predicted output will be between 66 and 76.8 when we keep the climate variables at their long term average.

- However, it is important to note that the values of θ^0 in both alternatives can't be directly used in constructing the confidence interval of the above prediction. This is because the prediction is made only by taking the coefficients of the significant variables without dropping the insignificant ones. This is because the insignificant variables from the base model can't be dropped as they are found to be jointly significant as proved by F test and a model with these variables is justified using the likelihood ratio test. Moreover, dropping the insignificant variable would result in omitted variables bias and hence biased prediction.

JP Postulates

Postulate 1. *Positive production expectations* $[E(y) > 0]$.

Postulate 2. *Positive marginal product expectations* $[\partial E(y)/\partial X_i > 0]$.

Postulate 3. *Diminishing marginal product expectations* $[\partial^2 E(y)/\partial X_i^2 < 0]$.

Postulate 4. *A change in variance for random components in production should not necessarily imply a change in expected output when all production factors are held fixed* $[\partial E(y)/\partial V(\varepsilon) = 0$ possible$]$.

Postulate 5. *Increasing, decreasing, or constant marginal risk should all be possibilities* $[\partial V(y)/\partial X_i \gtreqless 0$ *possible where* $V(y) = E[y - E(y)]^2]$.

Postulate 6. *A change in risk should not necessarily lead to a change in factor use for a risk-neutral (profit-maximizing) producer* $[\partial X_i^*/\partial V(\varepsilon) = 0$ *possible where* X_i^* *is the optimal input level*$]$.

Postulate 7. *The change in the variance of marginal product with respect to a factor change should not be constrained in sign* a priori *without regard to the nature of the input* $[\partial V(\partial y/\partial X_i)/\partial X_j \gtreqless 0$ *possible*$]$.

Postulate 8. *Constant stochastic returns to scale should be possible* $[F(\theta X) = \theta F(X)$ *for scalar* $\theta]$.

The Mean Output Regression

```
. nl(lnoutput={bo} + {b1}*lncapital + {b2}*lnlabor + {b3}*lnland1 + {b4}*lnlivestock + {b5}*lnfertilizer1 + {b6}*lnirrigation + {b7}
> *lntemp + {b8}*lnprecip + {b9}*lntemp1 + {b10}*lnprecip1 + {b11}*t  + {b13}*_Icountry_2 + {b14}*_Icountry_3 + {b15}*_Icountry_4 +
> {b16}*_Icountry_5 + {b17}*_Icountry_6 + {b18}*_Icountry_7 + {b19}*_Icountry_8 + {b20}*_Icountry_9 + {b21}*_Icountry_10 + {b22}*_Ic
> ountry_11 + {b23}*_Icountry_12) [aweight=12.900373]
(sum of wgt is 6192.179)

Iteration 0:  residual SS =  6.072138
Iteration 1:  residual SS =  6.072138
```

Source	SS	df	MS
Model	392.516869	22	17.8416759
Residual	6.07213815	457	.013286954
Total	398.589007	479	.832127363

Number of obs =	480
R-squared =	0.9848
Adj R-squared =	0.9840
Root MSE =	.1152691
Res. dev. =	-735.4552

lnoutput	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
/bo	-1.390092	2.058309	-0.68	0.500	-5.435016	2.654831
/b1	.0778855	.0171093	4.55	0.000	.0442629	.111508
/b2	.4847885	.0491852	9.86	0.000	.3881313	.5814457
/b3	.3420253	.0622298	5.50	0.000	.2197333	.4643173
/b4	.1448534	.0329268	4.40	0.000	.0801467	.2095601
/b5	.0032612	.0021532	1.51	0.131	-.0009703	.0074926
/b6	.0376011	.0145873	2.58	0.010	.0089345	.0662676
/b7	-.6150154	.4307065	-1.43	0.154	-1.461426	.2313953
/b8	.1486419	.0531344	2.80	0.005	.0442239	.2530599
/b9	.0678268	.0417913	1.62	0.105	-.0143001	.1499537
/b10	-.0237543	.0380974	-0.62	0.533	-.0986221	.0511135
/b11	.0030276	.0011716	2.58	0.010	.0007251	.00533
/b13	.4326165	.108147	4.00	0.000	.2200894	.6451435
/b14	.3479112	.2131994	1.63	0.103	-.0710615	.7668839
/b15	1.444784	.4339386	3.33	0.001	.5920212	2.297546
/b16	-.6877483	.2705492	-2.54	0.011	-1.219423	-.1560735
/b17	.6203319	.3878098	1.60	0.110	-.1417798	1.382444
/b18	.0628545	.3762085	0.17	0.867	-.6764587	.8021677
/b19	-.3905649	.347157	-1.13	0.261	-1.072787	.2916571
/b20	-.8675494	.4001719	-2.17	0.031	-1.653955	-.0811442
/b21	-.0691657	.4818279	-0.14	0.886	-1.016039	.8777074
/b22	-1.155059	.570012	-2.03	0.043	-2.275229	-.0348897
/b23	-.8592614	.5972923	-1.44	0.151	-2.033041	.3145186

Parameter bo taken as constant term in model & ANOVA table

The variance Ouput Function

```
. xtreg lnabsresidual capital land1 labor livestock irrigation fertilizer temp precip temp1 precip1 t, fe

Fixed-effects (within) regression               Number of obs      =       480
Group variable: country1                        Number of groups   =        12

R-sq:  within  = 0.8923                         Obs per group: min =        40
       between = 0.0057                                        avg =      40.0
       overall = 0.0097                                        max =        40

                                                F(11,457)          =    344.07
corr(u_i, Xb)  = -0.9279                        Prob > F           =    0.0000
```

lnabsresid~l	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
capital	1.06e-07	1.99e-08	5.32	0.000	6.69e-08	1.45e-07
land1	6.21e-06	5.91e-07	10.49	0.000	5.04e-06	7.37e-06
labor	1.18e-07	2.94e-07	0.40	0.689	-4.60e-07	6.95e-07
livestock	-8.31e-11	5.95e-11	-1.40	0.163	-2.00e-10	3.39e-11
irrigation	-.0000276	.0000106	-2.60	0.010	-.0000485	-6.73e-06
fertilizer	7.80e-08	1.59e-08	4.91	0.000	4.68e-08	1.09e-07
temp	-.0005181	.0009923	-0.52	0.602	-.0024681	.0014319
precip	.0000154	2.44e-06	6.32	0.000	.0000106	.0000202
temp1	.005125	.0017479	2.93	0.004	.0016902	.0085598
precip1	-.0000728	.0000268	-2.71	0.007	-.0001255	-.0000201
t	.0012469	.0000504	24.73	0.000	.0011478	.0013459
_cons	-.1513791	.1026558	-1.47	0.141	-.3531149	.0503568
sigma_u	.18503829					
sigma_e	.00613659					
rho	.99890136	(fraction of variance due to u_i)				

```
F test that all u_i=0:     F(11, 457) =   578.86            Prob > F = 0.0000
```

Appendix 9: IPCC projections for Eastern and Southern Africa from a set of 21 global models in the CMIP3 under the medium emission scenario (A1B) in the 2080 to 2099 from the 1980 to 1999 levels

		Temperature Response (°C)						Precipitation Response (%)						Extreme Seasons (%)		
Region[a]	Season	Min	25	50	75[32]	Max	T yrs	Min	25	50	75	Max	T yrs	Warm	Wet	Dry
EAF	DJF	2.0	2.6	3.1	3.4	4.2	10	-3	6	13	16	33	55	**100**	**25**	1
	MAM	1.7	2.7	3.2	3.5	4.5	10	-9	2	6	9	20	>100	**100**	**15**	4
12S,22E	JJA	1.6	2.7	3.4	3.6	4.7	10	-18	-2	4	7	16		**100**		
to	SON	1.9	2.6	3.1	3.6	4.3	10	-10	3	7	13	38	95	**100**	**21**	3
18N,52E	Annual	1.8	2.5	3.2	3.4	4.3	10	-3	2	7	11	25	60	**100**	**30**	1
SAF	DJF	1.8	2.7	3.1	3.4	4.7	10	-6	-3	0	5	10		**100**	**11**	
	MAM	1.7	2.9	3.1	3.8	4.7	10	-25	-8	0	4	12		**98**		
35S,10E	JJA	1.9	3.0	3.4	3.6	4.8	10	-43	-27	-23	-7	-3	70	**100**	1	**23**
to	SON	2.1	3.0	3.7	4.0	5.0	10	-43	-20	-13	-8	3	90	**100**	1	**20**
12S,52E	Annual	1.9	2.9	3.4	3.7	4.8	10	-12	-9	-4	2	6		**100**	4	**13**

Source: IPCC, 2007

Estimation of the JP stochastic production function (consistency and Efficiency)

1. JP + Stochastic View

$$y_t = f(Z_t, \alpha) + h(Z_t, \beta)\varepsilon_t,$$
$$E(\varepsilon_t) = 0, \quad E(\varepsilon_t^2) = 1, \quad E(\varepsilon_t \varepsilon_\tau) = 0, \quad \text{for} \quad t \neq \tau, \tag{8}$$

where

$$\ln f(Z_t, \alpha) \equiv (\ln Z_t)'\alpha \equiv z_t'\alpha, \tag{9}$$

$$\ln h(Z_t, \beta) \equiv (\ln Z_t)'\beta \equiv z_t'\beta, \tag{10}$$

$$Z_t = Z(X_t).$$

- The problem that JP described

$$\ln y_t = \ln f(Z_t, \alpha) + \ln [1 + \varepsilon_t h(Z_t, \beta)/f(Z_t, \alpha)].$$

Hence, where

$$X = \begin{bmatrix} z_1' \\ \vdots \\ z_T' \end{bmatrix}, \qquad V = \begin{bmatrix} u_1 \\ \vdots \\ u_T \end{bmatrix},$$

and

$$u_t \equiv u_t(\varepsilon_t) \equiv \ln [1 + \varepsilon_t h(Z_t, \beta)/f(Z_t, \alpha)],$$

one finds that the OLS estimate α^* has expectation

$$E(\alpha^*) = \alpha + (X'X)^{-1}X'E(V) \neq \alpha,$$

JP and OLS in the second stage

Theorem 2. *Suppose the conditions of Theorem 1 holds -- that the ε_t are identically and independently distributed and fourth moments of the ε_t distribution exist. Then consistent estimates of β except for the constant term are obtained by ordinary least-squares (OLS) regression of* $\ln|\hat{u}_t|$ *on* z_t *[where $\hat{u}_t$ is given by* (12) *and $\hat{\alpha}$ is the α estimate from Theorem* 1] *if the true parameter vector β_0 is an interior point of Ω. Furthermore, the OLS estimator $\hat{\beta}$ is asymptotically normal, and a consistent estimator of the covariance matrix of $\hat{\beta}$ is given by the usual OLS covariance estimator.*

It is also apparent that a consistent estimator for the constant term coefficient in h is obtained by adding 1.2704 to the one produced by the methods of Theorem 2.

Others

- *Traxler et al.,1995 both f(x) and h(x) quadratic with exp of h(x)*
- *Bernardo et al.,1995; log f(x) and linear for h(x) with exp of h(x) + fixed effects*
- *Barnwal and Kotani, 2010; linear f(x) and h(x) with exp of h(x) + fixed effects*
- *Attavanich and McCarl, 2011; linear for both f(x) and g(x) with exp of h(x)*
- *Ligeon et al., 2008 quadratic for both f(x) and h(x) with exp of h(x)*
- *Eggert and Tveterås, 2001; quadratic f(x) and h(x) wuth exponential h(x) + fixed effects*
- Tveteras et al.(2008); generalized leaontief for both f(x) and h(x) with exp of h(x) + fixed effects
- *Chavez, M. Daniela, 2003 Quadratic for the f(x) and translog and leontief for h(x)*
- *Svend Rasmusse, 2006 log-linear form of CD for f(x) and CD for h(x)*
- *Edward Kato, et al., 2009 CD f(x) and h(x)*
- *Benjamin Crost and Bhavani Shankar, 2008 CD f(x) and h(x) + fixed effects*

Bernardo et al.,1995

For the stochastic component (equation [3]), it is assumed that the logarithm of the variance of weight gains is a linear function of the exogenous variables energy supplement, initial calf weight, and pounds of available forage. This is referred to as multiplicative heteroskedasticity because different components of the variance are related multiplicatively (Judge et al., 1988). In addition, time effects are assumed fixed (which implies the use of year dummy variables).

The specified model is:

$$ln\,GN = \beta_0 + \sum_{t=1}^{T-1} \delta_t D_t + \beta_1 \ln INWT + \beta_2 \ln PF + \beta_3 \ln EN \qquad [4]$$

$$+ \beta_4 \ln(PF) \ln(EN) + e_i$$

with

$$e_i = \varepsilon_i \, h^{1/2}(D_t, INWT, EN, PF, \alpha), \qquad [5]$$

where GN denotes daily rate of weight gain (kg/head/grazing day), INWT is the initial calf weight (kg/head), EN is daily quantity of energy supplement fed (Mcal/grazing day), PF is the level of forage availability (kg/steer day), and the D_t's are year dummy variables. ε_i is normally distributed with $E(\varepsilon_i) = 0$ and $E(\varepsilon_i^2) = 1$.

The error term, e_i, is normally distributed with mean zero and variance

$$h(D_t, INWT, EN, PF, \alpha) = \exp(\alpha_0 + \alpha_1 INWT + \alpha_2 EN + \alpha_3 PF + \sum_{t=1}^{T-1} \gamma_t D_t).$$

Benjamin Crost and Bhavani Shankar, 2008

A problem of the cross-sectional approach described above is that it does not control for unobserved heterogeneity. By using an OLS-regression in the first step, one implicitly assumes that all parameters of the model are identical for all observations and that the error-term υ is the only source of variation. If this assumption is violated in practice, the estimates of the first-stage regression may be biased and inconsistent. The first-stage residuals are then no longer consistent estimates of the true error, which leaves the estimates of the second-stage regression inconsistent as well.

Fortunately, the availability of panel data – which contains several observation of the same unit at different points in time – allows us to control for some of the unobserved heterogeneity. The first analysis of production risk using panel data was conducted by Griffiths and Anderson (1982). They develop a model with random-effects for each unit of observation and each time-period, thereby, controlling for unobserved heterogeneity. They then use prior information on the covariance structure of the errors to obtain more efficient estimates of the parameters. Their model, however, assumes that the unobserved heterogeneity is random, in the sense that the household – and time-effects are uncorrelated with the other explanatory variables in the model. If this assumption is violated, the model's estimates will be biased and inconsistent.

$$Y_{it} = f\left(X_{it}, \beta\right) + h\left(X_{it}, \gamma\right)^{1/2} \upsilon_{it} + \lambda_t + \alpha_i$$

where the subscript $i \in \{1,\ldots,n\}$ denotes the household and the subscript $t \in \{1, 2\}$ denotes the year in which the observation was made. We choose Cobb-Douglas functional forms for both $f(\cdot)$ and $h(\cdot)$. The first-stage estimates of β can easily be obtained, for example by a dummy-variable regression (i.e. an OLS-regression that contains a dummy-variable for each household and for one of the two years). This regression yields consistent estimates of β, α_i and λ_t.

We can then calculate consistent estimates of ε and use them in a second-stage that is analogous to the one described above:

$$\hat{\varepsilon}_{it}^2 = h\left(X_{it}, \gamma\right) \sigma_\upsilon^2 + \eta_{it}$$

One might suspect that the unobserved heterogeneity is not only confined to the expected-mean function $f(\cdot)$, but also has an effect on the variance function $h(\cdot)$. Theoretically, one could control for this by adding household-specific fixed effects to the second-stage regression as well. Unfortunately, this is not feasible in our case, since we only observe two growing seasons and including fixed effects would leave us with too few degrees of freedom to obtain meaningful results.

So far, we have defined the risk-effect of inputs very narrowly as the effect on the variance of the idiosyncratic error ε_{it}. Inputs may, however, also affect the covariance between ε_{it} and the year-specific error λ_t. It is, for example, possible that farmers prefer traditional varieties which underperform in good years but do reasonably well in bad years to modern varieties which have a higher expected yield but can be prone to very low yield in bad years. Also, a widespread adoption of risk-reducing technologies may reduce the variance of the season-specific error itself.

Barnwal and Kotani, 2010

Three stage FGLS procedure is applied to estimate the parameters of equation (2). Log yield variance regression, in the second stage, adjusts standard errors appropriately taking first stage yield variation into account. F-value is less than 0.1 in Log yield variance regression for Kharif as well as Rabi, which suggests existence of heteroskedasticity (See $Prob > F$-values of Log yield variance regression in table 7). The final stage FGLS estimates parameters for Yield mean regression using the square root of variance predictions from the second stage as inverse of weights. Variance equation takes a non-linear (logarithmic) form and assures positive predicted variances, whereas Yield mean regression is linear in all dependent and independent variables. The final estimates of the stochastic function parameters with Kharif and Rabi rice yield as dependent variable are shown in Table 7.

Table 7 shows the estimated value of coefficients for Log yield variance (second stage) and Yield mean (third stage) regressions. Log yield variance takes Log of variance of the residuals from the first stage as dependent variable and corresponding part of the table provides information about effect of climatic factors on the yield variability. Here, the interpretation of positive coefficient will imply that a higher yield variance is expected with an increase in the corresponding explanatory variable, keeping all other factors constant.