ETH Yield Gap Analysis

Michiel van Dijk & Tom Morley

27 February 2017

What's new since last time

- A more flexible translog function allowing for interactions between inputs
- Control function approach to incorporate a feedback loop between nitrogen and yield
- numerical methods to find optimum nitrogen use
- GYGA variables

Yield gap framework

@michiel insert picture

Data

- Two (now three) LSMS-ISA surveys 2011 and 2013 (and as of last week 2015)
- But due to a mistake in recording production in 2011 we only analyse 2013
- Yield defined as production/plot size $=>$ different from FAOSTAT/GYGA

Summary statistics

Table 1: Summary statistics

Statistic	N	Mean	St. Dev.	Min	Max
N	2,420	26.62	54.51	1.00	692.30
lab	2,420	173.90	306.90	1.00	2,763.00
area	2,420	0.19	0.35	0.0004	8.12
slope	2,413	12.64	10.32	1.00	84.70
elevation	2,413	1,799.00	405.70	371	2,909
SOC	2,420	7.90	2.21	1.87	17.10
rain_wq	2,420	762.40	97.66	704	1,135
GGD	2,420	7,179.00	827.30	4,922.00	10,392.00
AI	2,420	7,169.00	2,398.00	2,041.00	12,854.00
TS	2,420	1,097.00	340.90	538	2,279
yesN	2,420	0.37	0.48	0	1
irrig	2,401	0.05	0.22	0	1
impr	2,404	0.24	0.42	0	1
extension	2,406	0.34	0.47	0	1
title	2,219	0.50	0.50	0	1

Methodology

- Two stage estimation approach
- First stage Tobit model to avoid feedback loop/confounding
- Second stage stochastic frontiers estimation (bootstrapping SEs -> nonlinear function)
- Third stage policy variables

LSMS-ISA and GYGA

@ Michiel add in the GYGA map

Data issues

- ▶ How to solve missing yield potential data?
  - ▶ Average values are too low, maximum is too high.
- ▶ To calculate feasible yield gap information is needed on the use of inputs when costs are zero:
  - ▶ Level of nitrogen when yield diminishes (per region)
  - ▶ Use of labour and capital at this point
- ▶ Information on nitrogen/fertilizer use can be taken from experimental plot data => literature survey.
- ▶ Use of labour and capital demands assumptions (e.g. 10% increase)

Stochastic frontier analysis

parameter	Basic	GYGA	Basic + r	GYGA + r
(Intercept)	7.46 (1.29)**	7.11 (1.42)**	8.00 (1.28)**	7.75 (1.41)**
log(N)	0.16 (0.17)	0.15 (0.17)	0.07 (0.06)	0.07 (0.06)
log(lab)	0.16 (0.05)**	0.15 (0.05)**	0.15 (0.05)**	0.15 (0.05)**
I(log(N)^2)	0.02 (0.02)	0.02 (0.02)	0.03 (0.01)**	0.03 (0.01)**
I(log(lab)^2)	0.01 (0.01)	0.01 (0.01)*	0.01 (0.01)*	0.01 (0.01)*
log(slope)	-0.22 (0.03)**	-0.23 (0.03)**	-0.20 (0.03)**	-0.21 (0.03)**
elevation	-0.00 (0.00)**	-0.00 (0.00)	-0.00 (0.00)**	-0.00 (0.00)*
log(area)	-0.22 (0.02)**	-0.22 (0.02)**	-0.22 (0.02)**	-0.22 (0.02)**
SOC	0.03 (0.01)**	0.03 (0.01)**	0.03 (0.01)**	0.03 (0.01)**
log(rain_wq)	-0.14 (0.19)	-0.22 (0.21)	-0.18 (0.19)	-0.28 (0.21)
noN	0.27 (0.31)	0.25 (0.31)	NA	NA
irrig	-0.33 (0.09)**	-0.31 (0.09)**	-0.34 (0.09)**	-0.33 (0.09)**
impr	0.37 (0.06)**	0.38 (0.06)**	0.35 (0.06)**	0.35 (0.06)**
crop_count2	0.46 (0.04)**	0.47 (0.04)**	0.46 (0.04)**	0.47 (0.04)**
phdum_gt70	-0.21 (0.09)*	-0.23 (0.10)*	-0.22 (0.09)**	-0.24 (0.10)**
phdum55_2_70	-0.06 (0.07)	-0.07 (0.07)	-0.08 (0.07)	-0.09 (0.07)
log(N):log(lab)	-0.03 (0.01)**	-0.03 (0.01)**	-0.03 (0.01)**	-0.02 (0.01)**
GGD	NA	0.00 (0.00)*	NA	0.00 (0.00)*
AI	NA	0.00 (0.00)	NA	0.00 (0.00)
TS	NA	-0.00 (0.00)	NA	-0.00 (0.00)
rd	NA	NA	-0.08 (0.03)*	-0.08 (0.03)**
sigmaSq	1.94 (0.10)**	1.94 (0.10)**	1.92 (0.10)**	1.91 (0.10)**
gamma	0.81 (0.02)**	0.81 (0.02)**	0.81 (0.02)**	0.81 (0.02)**

Yield gap estimations

Table 3: Relative yield gap

Zone	TEYG_l	EYG_l	EUYG_l	TYG_l	YG_l_Ycor
AMHARA	1564	NA	NA	8232	11923
BENSHANGULGUMUZ	989	NA	NA	11973	16622
DIRE DAWA	998	NA	NA	4132	9947
GAMBELLA	1555	1605	4276	8788	16224
HARARI	744	NA	NA	1603	5938
OROMIYA	1050	NA	NA	9855	13814
SNNP	1182	NA	NA	9663	14253
SOMALI	779	NA	NA	10425	15437
TIGRAY	1255	2179	1350	6420	11204
Total	1151	NA	NA	9195	13481

Closing the yield gap
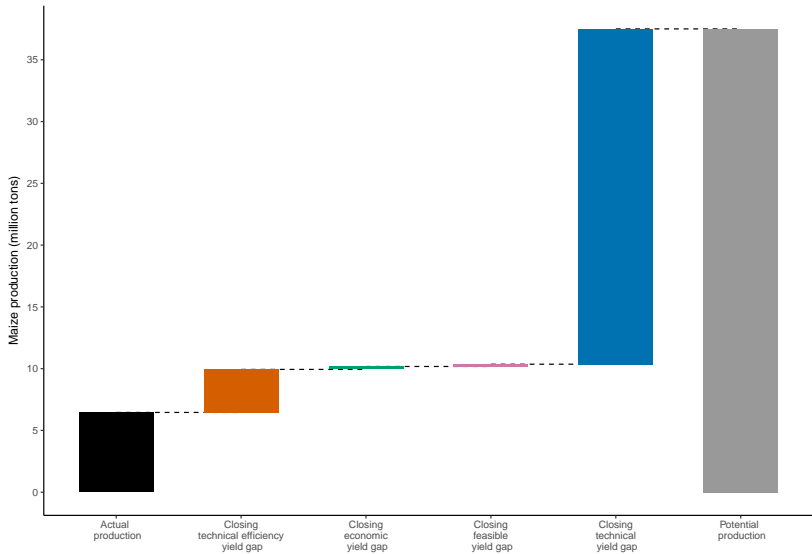
Policies to close the yield gap

@ Michiel policy framework picture

Explaining the TEYG

Next steps

- Calculate elasticities of second stage variables to get a better interpretation
- Think about how to link policies to yield gaps
- finish writing paper
- Third wave of data???? but ..... time constraints