Ethiopia Yield Gap Analysis

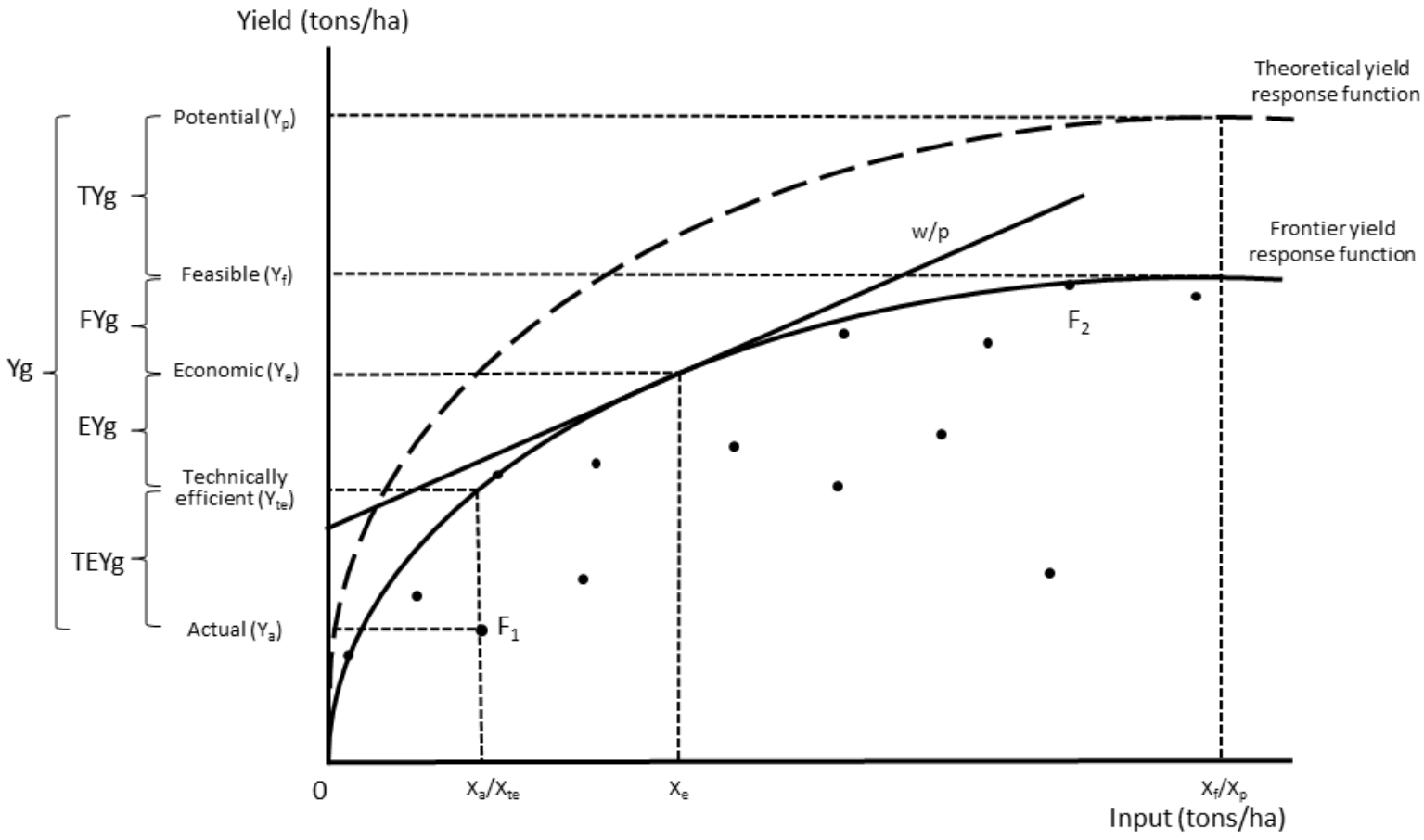
Michiel van Dijk & Tom Morley

27 February 2017

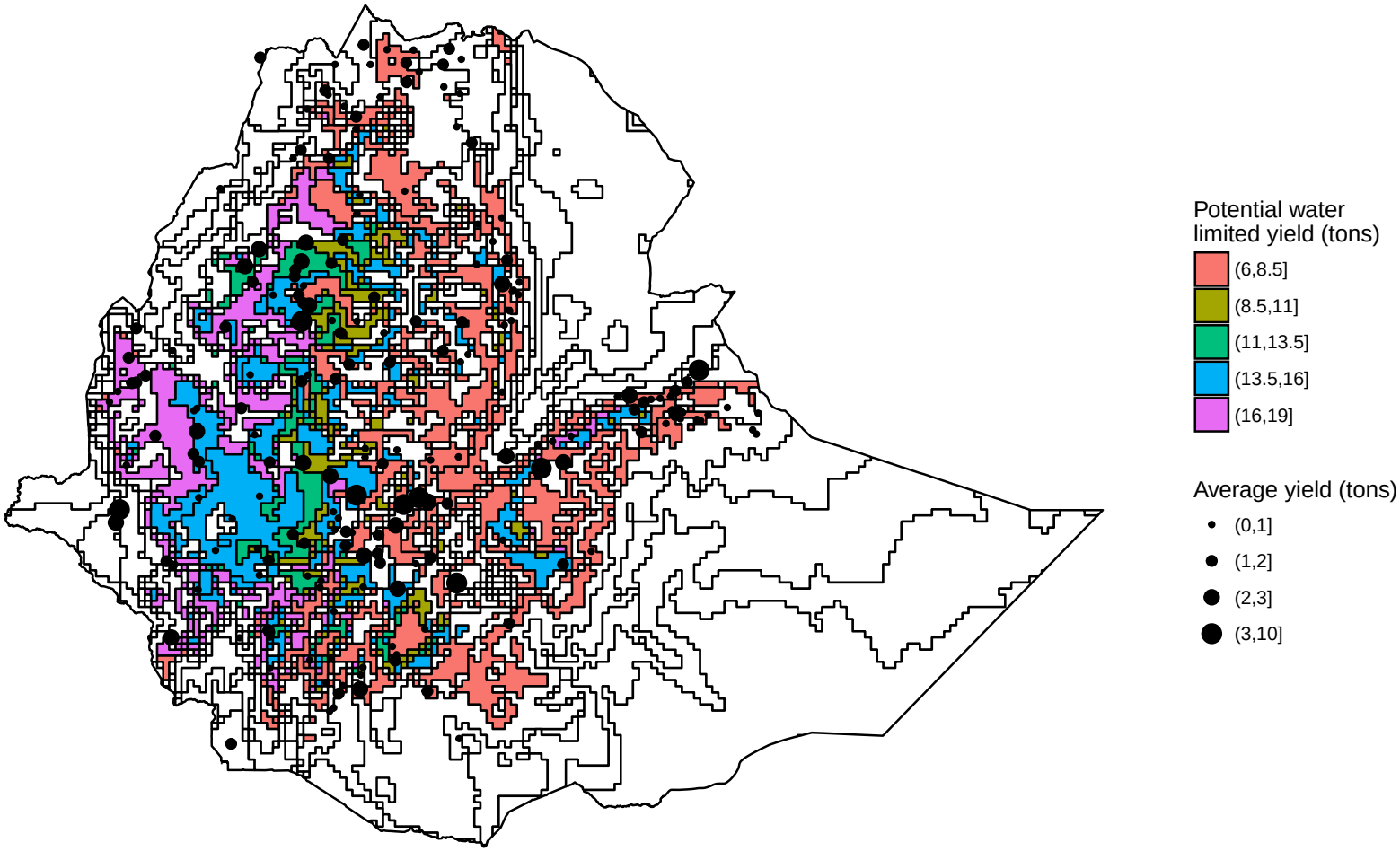
What's new since last time

- 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 environmental control variables

Yield gap framework



LSMS-ISA and GYGA



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

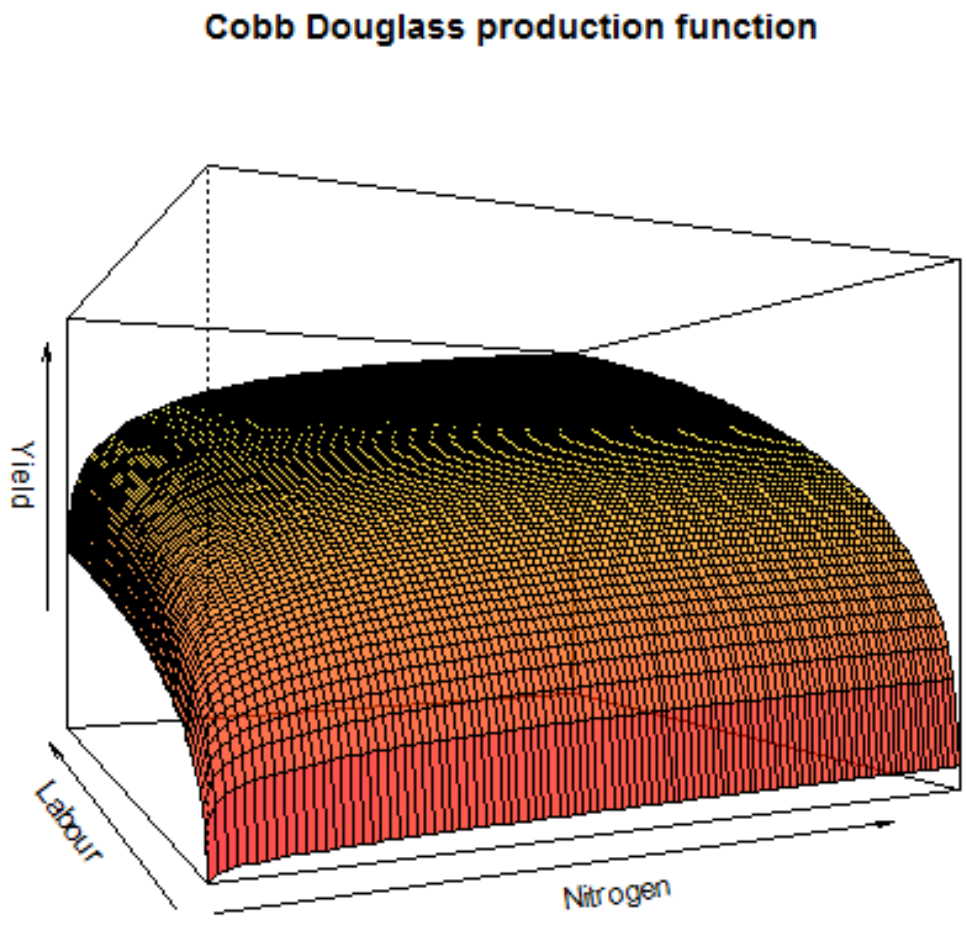
Table 1: Summary statistics

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
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

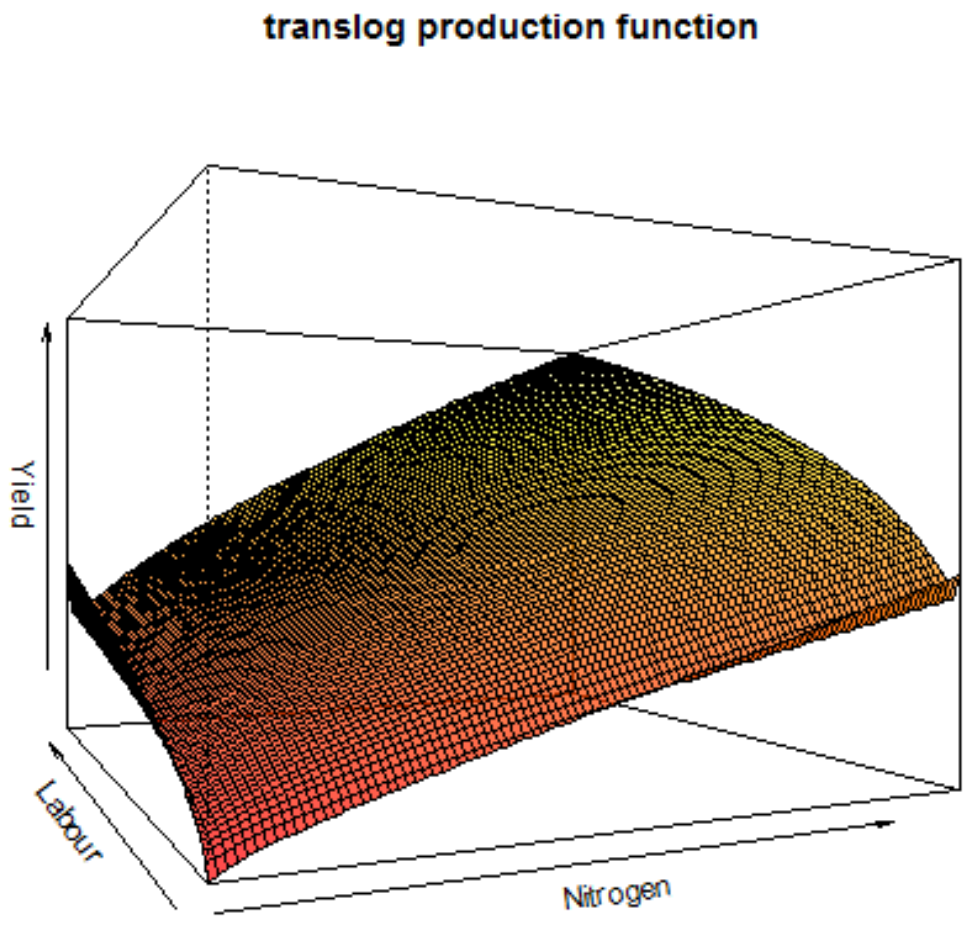
Methodology

- Three stages in the estimation of the production function
- First stage Tobit model to avoid feedback loop/confounding
- Second stage stochastic frontiers estimation
  - bootstrapping SEs due to nonlinear first stage function function
  - Numeric solution to find the economically optimal Nitrogen use
- Third stage policy variables

Cobb Douglass function



Translog production function



Endogeneity

- Idea of a feedback loop between the level of nitrogen applied and the yield
- In econometrics speak: the residuals from the estimation are correlated with the explanatory variable (Nitrogen)
- Agronomic interpretation: Higher yield in good soils, but Nitrogen also works better in good soil confounding
- Policy interpretation: Farmers listen to extension agents and see improvements spurring on more improvements
- Solution: Instrumental variables style approach, called a control function for a nonlinear first stage regression
    - This requires finding suitable instruments - distance, age, social capital, politics, soil quality??

Stochastic frontier analysis

parameter	Basic	GYGA	Basic + r	GYGA + r
(Intercept)	7.06 (1.29)**	6.84 (1.42)**	7.67 (1.28)**	7.47 (1.41)**
log(N)	0.16 (0.17)	0.15 (0.17)	0.08 (0.06)	0.08 (0.06)
log(lab)	0.16 (0.05)**	0.16 (0.05)**	0.15 (0.06)**	0.14 (0.06)**
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	0.04 (0.01)**	0.04 (0.01)**	0.04 (0.01)**	0.03 (0.01)**
log(rain_wq)	-0.10 (0.19)	-0.22 (0.21)	-0.15 (0.19)	-0.27 (0.21)
noN	0.27 (0.31)	0.24 (0.31)	NA	NA
impr	0.36 (0.06)**	0.37 (0.06)**	0.34 (0.06)**	0.35 (0.06)**
crop_count2	0.47 (0.04)**	0.48 (0.04)**	0.48 (0.04)**	0.49 (0.04)**
phdum_gt70	-0.24 (0.09)**	-0.26 (0.10)**	-0.26 (0.09)**	-0.27 (0.10)**
phdum55_2_70	-0.04 (0.07)	-0.06 (0.07)	-0.07 (0.07)	-0.08 (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	-0.09 (0.03)**	-0.09 (0.03)**
sigmaSq	1.95 (0.10)**	1.94 (0.10)**	1.91 (0.10)**	1.90 (0.10)**
gamma	0.81 (0.02)**	0.81 (0.02)**	0.80 (0.03)**	0.80 (0.03)**

Stochastic frontier analysis

- Residuals (rd) are significant implying endogeneity indeed exists
- GDD is significant and including GYGA variables controls for climate conditions
- Nitrogen coefficients: level is not significant, square is positive, interaction is negative

Optimal Nitrogen level and MPP values

ZONE	n	N plots	N	Nopt	MPP
AMHARA	450	254	92	132	13
BG	145	12	27	53	31
DIRE DAWA	39	3	26	85	14
GAMBELLA	18	0	NA	169	NA
HARARI	166	97	53	68	9
OROMIYA	747	235	68	63	15
SNNP	580	215	65	116	18
SOMALI	80	2	9	59	16
TIGRAY	135	53	59	383	14

Data issues

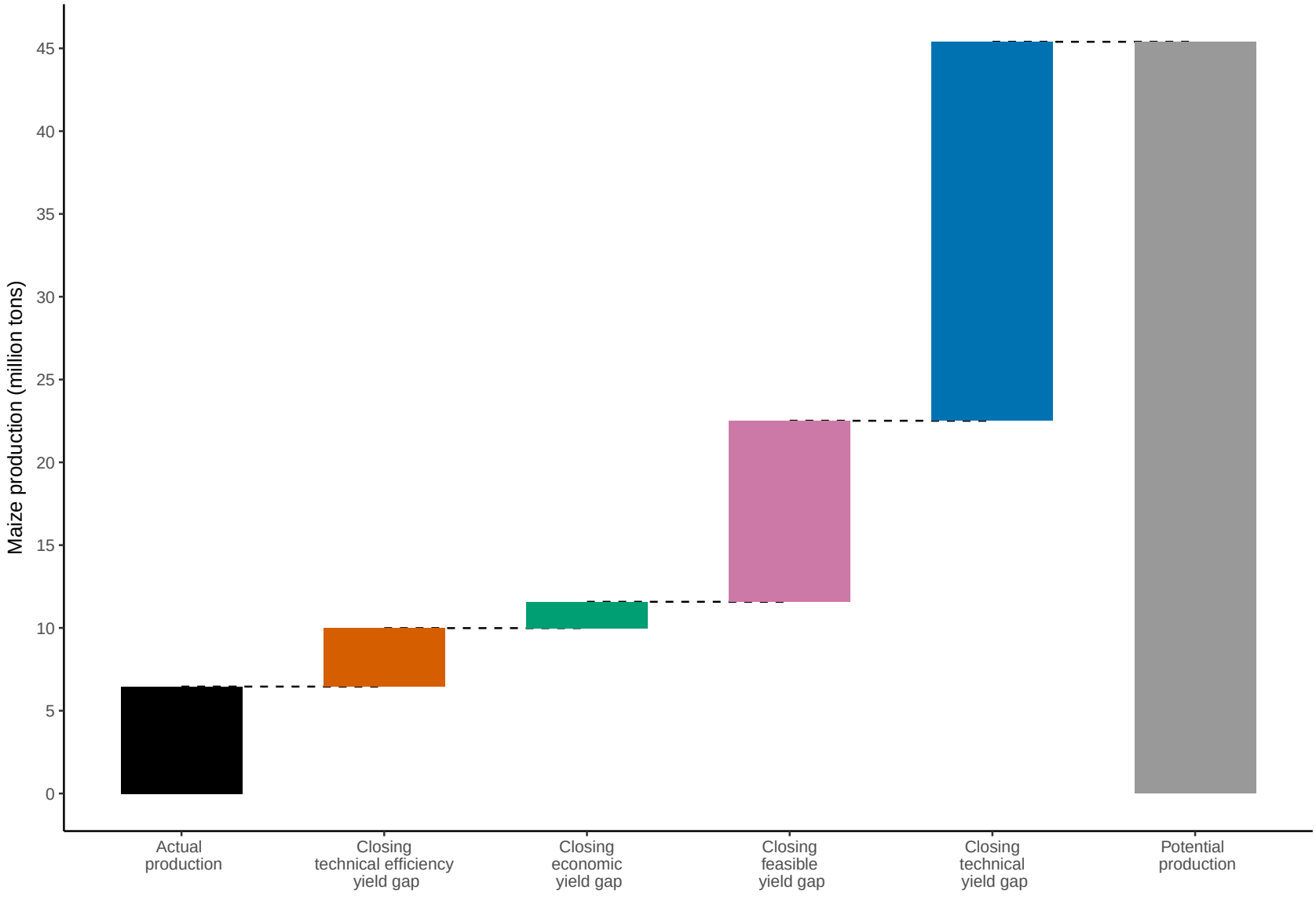
- How to solve missing yield potential data?
  - Average values are too low, maximum is too high.
- To calculate feasible yield gap information is needed on the use of inputs when costs are zero:
  - Level of nitrogen when yield diminishes (per region)
  - Use of labour and capital at this point
- Information on nitrogen/fertilizer use can be taken from experimental plot data => input from agronomy colleagues required.
- Use of labour and capital demand assumptions (e.g. 50% increase in labour)

Yield gap estimations

Table 4: Relative yield gap

Zone	TEYG	EYG	FYG	TYG	YG
AMHARA	13	5	25	56	100
BG	7	4	31	59	100
DIRE DAWA	10	6	51	33	100
GAMBELLA	9	10	41	40	100
HARARI	13	3	71	13	100
OROMIYA	8	3	29	61	100
SNNP	8	5	30	57	100
SOMALI	5	3	34	58	100
TIGRAY	11	19	26	44	100
Total	9	5	30	57	100

Closing the yield gap



Explaining the TEYG

Table 5: Third stage analysis

	Estimate	Std. Error	z value	Pr(>|z|)
Z_extension	-1.10	0.32	-3.38	0.00
Z_age	-0.01	0.02	-0.35	0.73
Z_I(age^2)	0.00	0.00	0.22	0.82
Z_ed_any	0.04	0.15	0.30	0.77
Z_sex	0.01	0.19	0.04	0.97
Z_title	-0.34	0.16	-2.17	0.03
Z_log(area_tot)	-0.07	0.06	-1.08	0.28

Wrap up

- New and better translog function means more flexibility of inputs
- Optimal nitrogen levels calcualted per farmer using their actual values of other inputs
- Identified and dealt with bias introduced due to endogeneity and feedback loops

Next steps

- Calculate elasticities of second stage variables to get a better interpretation and add more policy variables (e.g. cost of transportation)
- Slight experimentation with additonal control variables (e.g. use of animal traction)
- Think about how to link policies to yield gaps
- Finish writing paper for ETH meeting.
- Third wave of data???? but ..... time constraints