Ethiopia Yield Gap Analysis

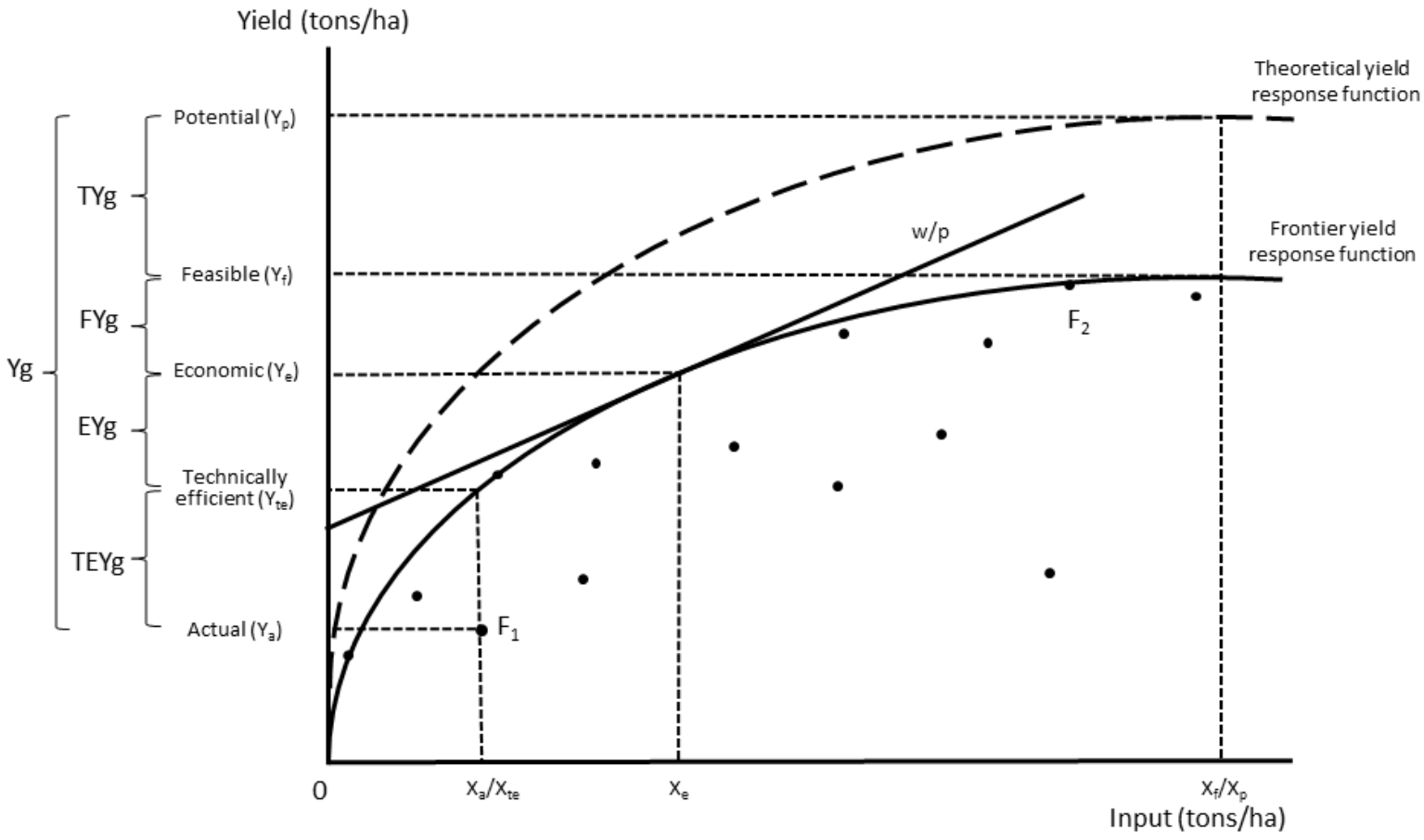
Michiel van Dijk & Tom Morley

27 February 2017

Outline of analysis

- Decomposition of maize yield gap in Ethiopia
- A flexible translog production allowing for interactions between inputs
- Numerical methods for finding optimal nitrogen use
- GYGA environmental control variables
- Heterogeneity not dealt with due to data limitations
- Endogeneity

Yield gap framework



LSMS-ISA and GYGA

Data

- Two (now three) LSMS-ISA surveys 2011 and 2013 (and as of a month ago, 2015 - 2016)
- 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

- Stochastic frontiers analysis
- Includes an asymmetric error component which accounts for plot level inefficiency.
- Estimated by maximum likelihood
- Altered density of error term to allow exogenous determinants of inefficiency such as extension services.
- Marginal effects of exogenous determinants of inefficiency

Stochastic frontier analysis

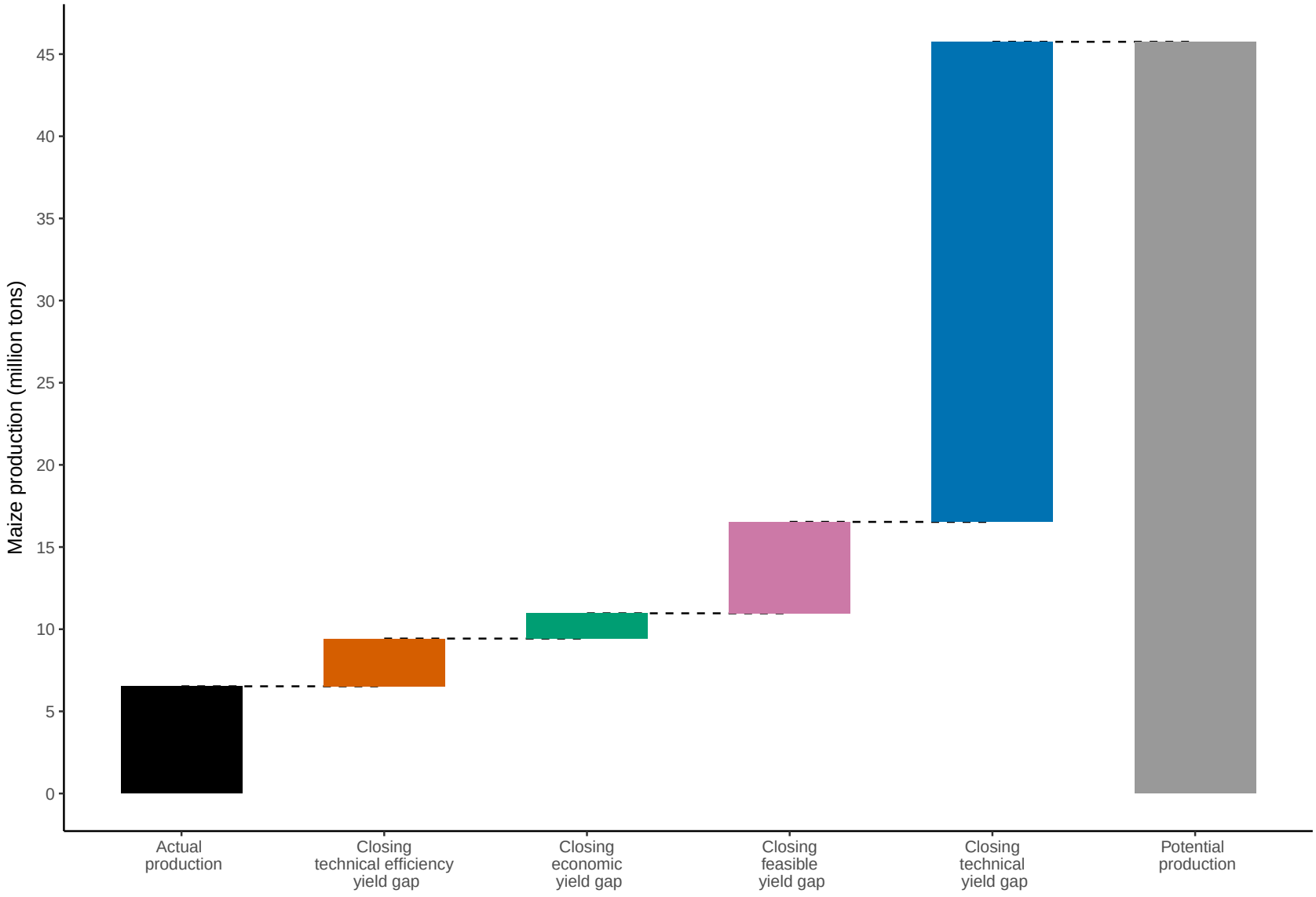
Variable	model 1	model 2	model 3	model 4
(Intercept)	4.761**	4.221**	4.828**	4.135**
logN	0.224**	0.223**	0.188**	0.149*
loglab	0.329**	0.286**	0.324**	0.281**
logseed	0.569**	0.456**	0.507**	0.392**
logNsq	0.046**	0.027**	0.046**	0.029**
loglabsq	0.018**	0.020**	0.018**	0.019**
logseedsq	0.011	0.009	0.017	0.014
logN:loglab	-0.022**	-0.023**	-0.023**	-0.023**
logN:logseed	-0.059**	-0.036**	-0.053**	-0.026
loglab:logseed	-0.063**	-0.058**	-0.060**	-0.053**
logarea	NA	-0.099**	NA	-0.094**
phdum55_2_70	NA	0.118**	NA	0.106*
crop_count2	NA	0.421**	NA	0.426**
dumoxen	NA	0.065	NA	0.067
SOC2	NA	0.015**	NA	0.020**
logslope	NA	-0.216**	NA	-0.216**
elevation	NA	-0.000**	NA	-0.000**
GGD	NA	0.136**	NA	0.167**
AI	NA	0.035**	NA	0.038**
TS	NA	-0.000**	NA	-0.000**
Z_age	NA	NA	0.002	-0.002
Z_sex	NA	NA	-0.209	-0.047
Z_ed_any	NA	NA	0.030	-0.010
Z_title	NA	NA	0.139	0.112
Z_extension	NA	NA	-0.681**	-0.752**
Z_credit	NA	NA	-0.495*	-0.524**
Z_dist_market	NA	NA	-0.003*	0.002
Z_popEA	NA	NA	-0.000*	-0.000*
Z_logarea_tot	NA	NA	-0.123*	-0.044
sigmaSq	2.254**	1.885**	2.716**	2.241**
gamma	0.822**	0.786**	0.840**	0.809**

Yield gap estimations

Table 3: Relative yield gap

Zone	TEYG	EYG	FYG	TYG	YG
AFAR	7	9	11	72	100
AMHARA	9	3	14	74	100
BG	7	7	15	71	100
DIRE DAWA	7	6	15	72	100
GAMBELLA	16	23	11	50	100
HARARI	10	2	23	65	100
OROMIYA	7	3	14	76	100
SNNP	8	6	15	71	100
SOMALI	6	5	9	80	100
TIGRAY	8	11	9	73	100
Total	8	5	14	73	100

Closing the yield gap



Explaining the TEYG

- marginal effects of exogenous determinants of inefficiency.

	model 3	model 4
age	0.000	-0.000
sex	-0.022	-0.006
ed_any	0.003	-0.001
title	0.014	0.013
extension	-0.070	-0.089
credit	-0.051	-0.062
dist_market	-0.000	0.000
popEA	0.000	0.000
logarea_tot	-0.013	-0.005

Remaining issues

- Take cap values for nitrogen from nutrient gap project? Also useful for other countries.
- Missing yield potential data - currently taking the national maximum, average values are too low, maximum is too high!
- Reasonable increases in other variables such as labour and seed rates and dummy variables. Plus anything else we might expect?

Endogeneity

- Idea of a feedback loop between the level of nitrogen applied and the yield.
- Agronomic interpretation: Higher yield in good soils, but Nitrogen also works better in good soil resulting in confounding
- Policy interpretation: Farmers listen to extension agents and see improvements spurring on more improvements.
- Bottom line: failing to account for endogeneity in survey data results in biased estimates.
- For nitrogen this is most likely biased up implying that we overstate the impact of nitrogen on yields. And there is a large literature identifying the endogeneity of nitrogen so we should include this.
- Solution in stochastic frontiers analysis is to use a 2 stage limited information ML technique.

first stage

	Estimate	Std..Error	t.value	liml
(Intercept)	4.364	0.107	40.68	4.364
Pn	-0.042	0.002	-23.85	-0.042
cost2large_town	-0.004	0.001	-3.987	-0.004
dist_market	-0.005	0.001	-7.002	-0.005

- R-squared on first stage is only 0.238, but this is OK because LIML performs well in the presence of weak instruments. Including other variables that affect the demand for nitrogen will also increase this.

Second stage

	Estimate	Std. Error	t value	CF Est	CF SE	CF p.val
(Intercept)	3.24	0.255	12.73	3.355	0.3196	10.5
logN	0.248	0.061	4.087	0.163	0.0603	2.71
loglab	0.387	0.082	4.731	0.369	0.0952	3.876
logseed	0.736	0.119	6.18	0.759	0.144	5.273
logNsq	0.054	0.011	4.933	0.058	0.01109	5.21
loglabsq	0.014	0.007	1.992	0.016	0.008115	1.915
logseedsq	-0.006	0.019	-0.312	-0.012	0.02361	-0.5268
logNlab	-0.022	0.009	-2.471	-0.025	0.008739	-2.817
logNseed	-0.073	0.014	-5.129	-0.077	0.01394	-5.526
loglabseed	-0.071	0.017	-4.064	-0.066	0.02049	-3.222
v	NA	NA	NA	0.213	0.01986	10.72

- Bootstrapped second stage SEs because v term is an estimate of the true error.

SF ML vs LIML with endogeneity

	Estimate	Std..Error	z.value	Pr. . . z..	LIML
(Intercept)	4.766	0.246	19.37	0	4.856
logN	0.222	0.058	3.854	0	0.143
loglab	0.336	0.076	4.422	0	0.321
logseed	0.56	0.109	5.117	0	0.587
logNsq	0.046	0.01	4.449	0	0.049
loglabsq	0.019	0.007	2.867	0.004	0.019
logseedsq	0.014	0.018	0.796	0.426	0.008
logNlab	-0.023	0.009	-2.638	0.008	-0.025
logNseed	-0.057	0.013	-4.27	0	-0.062
loglabseed	-0.068	0.016	-4.136	0	-0.063
sigmaSq	2.263	0.101	22.33	0	2.286
gamma	0.821	0.021	39.73	0	0.807

Wrap up

- New and better translog function means more flexibility of inputs
- Optimal nitrogen levels calculated per farmer using their actual values of other inputs
- Marginal effects of exogenous variables calculated giving a policy interpretation
- Identified bias introduced due to endogeneity but this needs to be extended somewhat.