

Questions from John Antle

1. IN Table IIA you say values are in real terms, but I cannot find an explanation of how you did that in your paper or the notes.

Sorry for missing this in the notes – I have now updated the notes to include this. All the monetary values (hired labor, land preparation costs and expenditure on fertilizer and for the robustness checks mentioned in the footnote, maize harvests in shillings and profits) used in the estimation were deflated using the GDP deflator. This deflated the 2000 data by 0.768 and the 1997 data by 0.622.

2. Pls let me know the # of observations you had in the “low-HIV sample”; I get 1061 after dropping the two high-HIV districts from the 1202 obs full sample.

Correct, I have 1061 as well.

3. Apparently you treated adult and child labor hours as equal?

Yes I did.

4. In your notes on the data you describe hired labor hour calculations but report hired labor in value terms in table IIA. So did you use hired labor quantity or value in the modeling? If value, I assume you used the reported hired wages to value hired labor, etc. Again, if value, how you convert to real terms?

Hired labor is used in value (as the notes say – see point 6 under final dataset in the notes) and the family labor is used in quantity/hours. The two enter the estimation separately (please see footnotes to the tables in the original paper).

5. For fertilizer, I followed procedures similar to yours, and obtain about the same quantities for dsp, map and can fert quantities; but the total fert value I get is quite different. Again, part of the issue may be how you converted to “real” terms. But if you could be more precise about which ferts you included, what prices you used for missing prices, etc, or provided your code to me, that would help. (I am also curious why you used values rather than quantities of active ingredient as is more typical practice in production function estimation).

I used values as I wanted one overall estimate of fertilizer input costs given the modeling was going to allow fertilizer to be endogenous. Allowing for many (well over 10) different types of fertilizer would have really complicated the estimation. I used all fertilizer types as the notes highlight and aggregated all the expenditures on all these types.

6. Did you include province dummies in the CRC models? They are not listed in the covariates identified in the table VIII B footnote.

The CRC models include household fixed effects. Those subsume the province fixed effects.

7. Can you pls provide the numbers of observations and R2 stats for Table IX? Also, in the footnote to table IX you describe these as “correlations” but I think they are regression coefficients, right?

Yes, these are indeed regression coefficients (though regression coefficients are just scaled correlations). I used the term correlations in this table as I wanted to be clear that these should not be interpreted as causal estimates of anything, but highlight correlation rather than causation.

For the columns the number of observations are as follows: 1058, 1058, 1058, 1057
For the columns the R-squareds are as follows: 0.1114, 0.1300, 0.1293, 0.1296

Also note that in this table I dropped two outliers in the distance to the closest fertilizer supplier (the households that were 70km or further from a supplier). I have added this to the data document.

8. Can you pls provide the summary stats (means and std devs) for the hybrid return distributions, i.e., corresponding to the values in your figure 5A and 5B.

| Hybrid Transition | Mean | Standard Deviation | Observations |
|--------------------|-------------|--------------------|--------------|
| Stay in Hybrid | -0.01025289 | 0.02815381 | 631 |
| Stay in Non-Hybrid | -0.25619588 | 0.13714393 | 210 |
| Leave Hybrid | 0.02908049 | 0.09260254 | 142 |
| Join Hybrid | 0.47484709 | 0.13648276 | 78 |

9. You mention in your notes about how you dealt with price outliers – but did you not do anything about quantity outliers? E.g., there are some very extreme values of fertilizer quantities and some of the labor variables.

I did not do anything with these. I assumed that the cleaning had dealt with some of these outliers (from my memory that was part of the cleaning that was conducted).