

Reply to editor MS-2019-3962

Francesco,

Thanks for the careful read and allowing us the chance to take another step forward. We’ve prepared responses to the new comments raised. Your comment is in italics in each section, followed by our response. We believe we’ve been able to provide a solution or answer to each of the points you raise.

Transport costs and distance

Question: *“I have become more aware of other equally sensitive identifying assumptions. One is the equalization of relative prices across districts within a state. Since this is largely a relative price between rural-produced and urban-produced goods, my intuition is that it could vary considerably with a rural district’s effective distance from a rural district, even within a state. Also, the previous intuition that the effect of distance may be more severe in tropical regions (bad roads, etc.) applies, so the omitted variable bias may be more severe there. Is there something that can be done/said to assuage this concern?”*

and

“On a very similar note, it seems that an important identifying assumption is that tau vary only at the state level. However, looking at your equations a very natural interpretation of tau is a transportation cost, and it seems unlikely that this will be the same across districts. Again, remote districts with bad roads may have different transportation costs from districts closer to the urban centers. and more so in the tropics.”

Response: The point here is well taken. Even if evidence suggests labor moves freely and that the wage is equalized across districts, that doesn’t imply that the relative price of agriculture (p_A in our setting) is equalized, in particular due to distances and transport costs.

To accommodate this, the first thing we’ve done is update section 3 of the paper to incorporate an explicit district-level wedge, τ_{Ais} . This shows explicitly that we must control for a district-level wedge in our regressions or we may end up with a biased estimate of β .

As you state, a natural interpretation of τ_{Ais} is as a transport cost and/or distance. Empirically, to handle this we have added data at the district level on road density and road types (e.g. highways versus two-lane roads). We’ve also added a control for the distance of the centroid of each district to the nearest major city (meaning with more than 100,000 people). Finally, we added a control for slope characteristics (i.e. ruggedness) that would capture transport cost differences conditional on road density or distance from a major city. Those transport cost variables are all now part of the baseline set of controls we include in all regressions in the paper.

What we see in the results is that there is still a significant difference between temperate and tropical land elasticities. The point estimates for both regions are higher than before. Our tropical estimate goes from around 0.09 in the old version to around 0.126, while the temperate goes from around 0.24 to 0.285. These differences survive all of our existing robustness checks. Your intuition was correct that distance/transport costs was biasing our results downward. However, it appears that this bias is common to both temperate and tropical areas, and so controlling for this does not erase the difference. The absolute gap remains about

0.15-0.16 regardless.

Transport costs are not the only thing that could be part of τ_{Ais} , given that this is a catch-all wedge. We don't believe that district-specific tax or subsidy policies would be very relevant, given the small size of the districts (around 26,000 people at the median). However, the handful of very large population districts (e.g. large cities) may well have distorted prices relative to other districts. We've eliminated those large urban districts from the baseline sample by excluding any district with more than 90% of population in an urban area, and our robustness checks in Table 3 (Panel B) show that the results hold up if we exclude any districts with more than 50,000 urban residents.

GAEZ productivity and TFP

Question: *"The second point was about the functional form in the relationship between TFP and your agro-climatic index. You still state this functional form without discussion. In the response letter you say that the coefficient of one is not critical (please explain) but what about greater deviations from the functional form assumption?"*

Response: Sorry for not responding to this point fully in the prior report. We mistook the question for a different one that we've had before regarding yields, and gave the short answer. This is the (much) longer answer to your actual question.

We're assuming $\ln A^{GAEZ} = \ln A$ (plus noise), or that the GAEZ based measure of caloric suitability is not only a good proxy for actual productivity (A), but in fact is related to it with elasticity of one. If that elasticity were not one, or $\ln A^{GAEZ} = \lambda \ln A$, then our estimation equation would really be based on $\ln A^{GAEZ} = \beta/\lambda \ln L/X$. We wouldn't necessarily be recovering an estimate of β . Worse, if the value of λ were itself different between tropical and temperate areas, then that could explain our results without any actual difference in β existing.

So why is it plausible to think that A^{GAEZ} is linearly related to A with an elasticity of one? We think the way the GAEZ measure is built ensures this is the case. A^{GAEZ} is reported in terms of a yield. Most important, it is calculated holding input use per hectare constant, which we'll be more explicit about in a moment. A one percent variation in A^{GAEZ} measures a one percent variation in yield for a given set of inputs per hectare.

Our theoretical object, A , is Hicks-neutral TFP. Hence a one percent variation in A is also associated with a one percent variation in yield, holding input use per hectare constant. As the elasticity of A^{GAEZ} and A with respect to yield are both one, their elasticity with respect to one another should be one as well. This isn't to say that A and A^{GAEZ} are identical in *level*, only that their elasticity with respect to yield is the same, and hence their elasticity with respect to one another should equal one as well.

That logic rests on the idea that A^{GAEZ} measures yield holding inputs constant. What does it mean that GAEZ holds inputs constant? To explain that, we need to explain a little more about how GAEZ calculates the potential yields that make up A^{GAEZ} . First, GAEZ calculates the maximum agro-climatic attainable yield for each crop, which is based on (among several things) raw energy (sunlight) and available water (evapotranspiration). This represents an upper bound for each pixel, regardless of technologies employed or inputs applied.

Second, GAEZ calculates a series of constraints for each pixel that keep yields from reaching those

maximums. Examples are high slopes, unworkable soils, or a lack of soil nutrients. Each of these constraints applies some kind of multiplier less than one to the maximum yield. So very high slopes might mean a multiplier of 0.25, meaning the pixel can only reach 25% of the theoretical maximum.

Here’s where inputs come in. GAEZ considers inputs to be “constraint eliminators”. You can build terraces, for example, to deal with slopes. Or you can use a plow to break up unworkable soil or fertilizer to make up for a lack of nutrients. For GAEZ, input use corresponds to higher multipliers, and they consider three different levels of input use: low, medium, and high. Low input use might mean that the slope multiplier is 0.5 rather than 0.25, medium inputs might correspond to a multiplier of 0.75, and high inputs might mean the slope multiplier is 1.0 (or high inputs completely offset the slope constraint). Pixels (and districts) with identical constraints, and under the same input assumption (low, medium, high), would hypothetically be using the same amount of inputs per hectare. In all of our regressions we use values for A^{GAEZ} calculated for the same input assumption (low is our baseline, and Table 5 shows results under the alternatives).

But pixels (and therefore districts) vary in the constraints they face. If they vary, then even using the same level of inputs (low, medium, high) may involve different actual amounts of inputs. A rugged pixel might need a lot of actual capital to relieve the slope constraints present, while a neighboring flat pixel might need very little capital given the lack of slope constraints.

So the claim that GAEZ is “holding inputs constant” when they calculate A^{GAEZ} holds true in the strict sense when pixels face identical constraints. For our argument about A^{GAEZ} having an elasticity of one with respect to A , we need districts to have similar constraints, so that the GAEZ input assumption implies similar actual input use.

Intuitively, our argument is that districts are very small, and within a given state they are relatively homogenous in geographic characteristics (and thus in the constraints they face). More concretely, GAEZ provides all the data on the constraints that hold at pixel (and hence at district) level. We can ensure that we are comparing districts with similar constraints most directly by controlling for this in our regressions. The specific constraints we identify from the GAEZ methodology as central are: dummies for soil nutrient availability classes, dummies for soil nutrient retention classes, dummies for soil workability classes, dummies for precipitation/evapotranspiration classes, dummies for the slope classes, and dummies for the length of growing period.

If we include controls for these constraints in our regressions, the results are the same, with a very small decline in the estimated β for temperate areas and an almost identical estimate for tropical areas. This holds for all of our standard robustness checks as well. We’ve added a table in the Appendix showing these results. We’ve also added a section in the Appendix explaining the logic about how the GAEZ measure is derived and why our specification for A^{GAEZ} makes sense.

Thus we believe that the specified relationship of A^{GAEZ} to A in the paper is appropriate, and does not create any inherent bias.

Specification

Question: *“The first point was that your estimating equation has the theoretically exogenous variable on the left, and the theoretically endogenous variable on the right. Why, and what happens if you switched them around?”*

Response: In this case we’ve simply made a judicious choice in how to arrange the regression to recover the value of β directly. Our specification equation is for an equilibrium relationship, and we’re trying to find the value of β that best rationalizes this relationship in different settings.

If we switched them around, our estimates would be wrong in expectation. To see this, write our baseline specification as $A = bL + e$, where we’ve dropped log notation for convenience, and you can imagine that A (TFP) and L (labor/land) are residuals left over after controlling for state fixed effects and our other control variables. This would give us some estimate \hat{b} . If we were right that our controls and assumptions ensure that L and e are uncorelated, then $E[\hat{b}] = b$.

Reversing this, we could run $L = cA + u$ as a regression, and we’d get \hat{c} as an estimate. Let’s assume again that our assumptions and controls work as expected so that in expectation $E[\hat{c}] = 1/b$. The problem arises when we go to calculate $1/\hat{c}$ in order to recover an estimate of the land elasticity. By Jensen’s Inequality, $E[1/\hat{c}] \geq 1/E[\hat{c}] = b$. In other words, in expectation looking at $1/\hat{c}$ will give us an estimate that is *too large*, even though there is nothing wrong, per se, with the reverse regression.

The only way for $E[1/\hat{c}] = 1/E[\hat{c}] = b$ to hold is if the variance of \hat{c} were exactly zero, which is clearly not the case given noise in the data. Hence regressing TFP on labor/land, as we’ve done, is the only feasible way to recover an estimate of

Additional modeling assumptions

Question: “Can you also discuss what happens to identification if there are state-level capital wedges (relative to labour), instead of just revenue wedges?”

Response: The short answer here is: not much. A state-specific wedge common to both ag and non-ag would alter each districts choice of K/L , but given that it is state-specific, it would get washed out with state fixed effects. Intuitively, if the state were closed the wedge would change the relative size of w/R in equilibrium for all districts, but it wouldn’t affect our empirics.

If there are sector-specific wedges (e.g. a different state-level wedge for agriculture than for non-agriculture) then the K/L ratios would differ between ag and non-ag, although they would still be proportional to one another. We show this in a new section in the Appendix. If this is the case, then the problem is essentially identical to your following point about the possibility of different ϕ values for capital across sectors. So let’s explain that first, and then provide our response to both points.

Question: “Another key assumption seems to be that the ϕ is the same for agriculture and non-agriculture. What happens to identification if this is not the case?”

Response: In this case the K/L ratios in ag and non-ag no longer equate, but they are still going to be proportional to one another. We show that in a new section in the Appendix.

Conceptually, both sector-specific wedges and sector-specific ϕ values mean we should be controlling for the agricultural capital/labor ratio in our specifications. In our DHS robustness checks, we have specific measures of agricultural capital, and our results are consistent.

Further, even if one believes that night lights are a measure of non-agricultural capital/labor ratios, this still provides a decent proxy for *agricultural* capital/labor ratios given that the two are proportional when there are either different wedges or different ϕ values. Variation in non-agricultural capital/labor ratios is still informative about variation in agricultural capital/labor ratios, even if the level of capital/labor ratios

is no longer identical across sectors.