Response to JEEA editor's report

Johnson and Vollrath

Following this document are the editor's report from Imran Rasul at JEEA, along with the four referee reports. We have not prepared point-by-point responses to each of the four referees, as the paper was rejected at JEEA. This document explains our revisions to the paper based on the major issues highlighted by Imran, which relate to specific criticisms that can be found in the referee reports.

By far the greatest concern was with the empirical methodology. We do not have a natural experiment or clever instrument to leverage here. Our results are based on cross-sectional comparisons of districts within provinces of countries. As we explain in the first section here, this is plausible strategy under a mild set of assumptions.

Methodology

Simply put, we are trying to estimate a parameter of a production function: the elasticity of agricultural output with respect to land. This has the standard issues of attempting to identify the parameter from observable data, knowing there are unobervable things that enter into the decisions of producers.

The referees read our methodology as requiring several strict assumptions to eliminate these issues: perfect factor mobility across districts within a province, an absence of distortions to agricultural prices, and the homogeneity of non-agricultural and agricultural goods. Of course, it does not make sense to assert that those assumptions hold across all provinces in all countries of the world. Quite sensibly, referees either wanted some supporting evidence for these assumptions, or simply rejected that they could be true.

However, none of those assumptions are necessary for our methodology. But we did not make a clear explanation of this in the prior version of the paper (although we might grumble a little about referees overlooking some text where we did address these issues). In trying to lay out the intuition behind our method, we did use several simplifying assumptions that we later relaxed, but that did not translate well for the reader.

The substantial revision to the paper that occurred was to make the assumptions (or lack thereof) behind our methods more transparent in Section 2. The paper has the full explanation, but here we lay out the logic of the methodology and how it avoids making the unpalatable assumptions cited by the prior referees.

There are several districts within a province. Farmers within each district are price-takers for their output and factor inputs, and minimize costs. The prices of output and inputs facing farmers are each combination of a province-level average price and a district-specific "wedge". Using the first-order conditions of the cost-minimization problem we show that there is a simple log-linear relationship between agricultural productivity, rural density, the province-level average prices, and the district-specific wedges.

This relationship is our regression specification. The coefficient relating agricultural productivity and rural density is the land elasticity we want to estimate. Empirically, our problem is handling the presence of the province-level average prices and district-specific wedges, both of which are unobserved, in that regression.

The province-level average prices of output and inputs are handled with province fixed effects. This eliminates them without putting any assumptions on how they are set. Thus these province average prices may be arbitrarily distorted from a competitive equilibrium in any direction via any source: taxes, subsidies, the influence of the non-agricultural sector, international trade, barriers to movement, and so on. It is

irrelevant what determines the province level average prices because we only use comparisons of districts *within* provinces to make our estimates. Country, or even province, level differences in agricultural policy, urbanization, market access, institutions, etc.. are all absorbed by our province fixed effects.

District-level wedges to prices are the central issue for us empirically. As we do not have district-level measures of wedges, we use a series of control variables to proxy for them. The main controls include night lights, the urban percent of population, and total population of the district. We believe these will capture district-level differences in output or input prices within provinces that the wedges represent. As a robustness check, we also allow for third-degree polynomials of all of these to capture any non-linear relationships of them with district-level wedges.

For this version of the paper we also linked a subset of our districts to DHS survey data, and have controls for the characteristcs of the rural population in those districts: age, education level, household assets, and agricultural assets. Including those gives us results that are nearly identical to our baseline results (Table 5).

The credibility of our results hinges on whether one believes we have controlled for these district-level wedges or not. We believe the consistency of our results across specifications and robustness checks puts the weight of evidence in our favor. Others might disagree. But this version of the paper should make more clear that our results are not just valid under narrow assumptions about perfect factor markets or absences of distortions in agriculture.

Measurement issues

Taking the methodology as given, the second main area of concern raised regarded the data. We use a measure of agricultural productivity - the caloric suitability index (CSI) of Ozak and Galor - that is based on the FAO's GAEZ project. The CSI is built using GAEZ estimates of crop productivity based on the assumption of "low" input use, or basic subsistence farming. Referees raised the concern that this may not be the relevant level of input use to consider across the world. For this version of the paper, we included robustness checks where we re-derived the CSI using Ozak and Galor's methods but with underlying GAEZ data on "medium" or "high" input use, as well as for irrigated agriculture. That is in Table 4 of the paper. In all cases, the pattern of results is consistent with our baseline.

This version of the paper is also more explicit that our results are not sensitive to the specific crops that are used to calculate the CSI, and in fact our results do not depend on using Ozak and Galor's CSI index at all, as we show. Referees were also concerned that our findings were peculiar to our choice of how to denote districts as "temperate" or "tropical", but as we explain in Section 3.2 we can use several alternative ways of assigned districts and receive similar results (and see Table 2, Panel A).

There were also complaints about the use of the HYDE data for rural population by district. This complaint seemed unfounded to us, in the sense that Table 3 includes robustness checks where we used population data from the GRUMP project, as well as from IPUMS. The IPUMS data also allowed us to confirm that rural population was in fact a good proxy for agricultural population. Our results have never been sensitive to the use of HYDE as the baseline population data.

Epidemiological transition

Both in this JEEA submission and a prior one, referees were somewhat split on the analysis in Section 4.2 that uses the epidemiological transition to test whether in fact areas with a high land elasticity (temperate) are more sensitive to population shocks, a prediction of the simple model we use to illustrate the effect of the land elasticity. It uses the data and set-up from Acemoglu and Johnson's 2007 paper on the effects of life expectancy on living standards.

There are open questions about the Acemoglu and Johnson paper, of course, and so some referees question the inclusion of this in our paper. For our purposes, it is the distinction in the results between temperate and tropical areas that is of interest. Whether those results are strictly causal is less important. But the inclusion of this section in the final paper, if we were fortunate enough to have it accepted, is not central to our findings.

MANAGING EDITOR
Imran Rasul

CO-EDITORS
Paola Giuliano
Bård Harstad
Claudio Michelacci
Nicola Pavoni
Giovanni Peri

Thursday, 17 January 2019

Dear Dietrich,

I am writing about your manuscript, MS-9485 "The Role of Land in Temperate and Tropical Agriculture," that you submitted for publication in the *Journal of the European Economic Association* on November 27th 2018.

I have now heard back from four trusted and expert referees on your manuscript: both are well familiar with the related literature on the causes and consequences of agricultural productivity, and winder implications for structural change and economic development. One of them (R2) previously review your paper at the *Economic Journal*, but has prepared a new report based on this version of the paper (and they recognize the improvements made to the paper over submissions).

As you would expect, the referees are all highly sympathetic to your research agenda, and none of them needs convincing that you are addressing an important question on the origins of comparative development, that is of general interest. However, despite these positives, the news is not altogether good as all referees raise first order concerns with the analysis conducted, and the majority remain unconvinced that the core analysis reaches the threshold for *JEEA*, that aims to establish itself as a top-6 general interest journal.

These are clearly expressed in each report, but relate to the core of the paper in terms of the assumption of factor mobility that underpins identification (this goes against the body of evidence on the misallocation of factors in low-income settings, and the concern is not fully ameliorated by the results in Appendix A3 – see R1 and R4 on this point especially), the additional assumption on agricultural and non-agricultural goods being substitutes (which also goes against a body of work and so needs to be justified more fully), the lack of supportive evidence for a range of other important assumptions made (see R2's report), and on the determination of wages in the model. Beyond the issues related to production function estimation in the cross section (I agree with R4 that the panel estimates are somewhat more inconclusive given the limited sample), I share R2's concern on the lack of discussion of prices (taxes and subsidies on crops), and how that might lead to OVB.

Also, as you can imagine for this kind of exercise, there are going to be data issues concerns (although all of these could be addressed) – such as on the measurement of productivity from the FAO data, and population measurement from the HYDE data.

University College London
Drayton House
30 Gordon Street
London WC1H OAX
United Kingdom

i.rasul@ucl.ac.uk
http://www.jeea.org

Overall, R1 and R4 recommend rejection, R2 is on the fence (but would reject if JEEA aims to become a top-6 journal). R3 is the most positive referee suggesting a revise and resubmit, but you can see from his/her report, they remain unconvinced on the empirical findings, and in their cover letter to me suggest, "I'm not fully convinced by their approach, and the focus of my report is on the doubts I have regarding their methodology."

I have now read your paper with interest. Putting aside these bottom line recommendations, I do find myself in agreement with the substantive comments raised on identification. Although I like the question, I don't see sufficient grounds on which to overrule the majority of referees. I must therefore reject your manuscript for *JEEA*. My sense is that a suitably revised version of your paper could be a better match for a lower-tier and more macro-specialized journal, such as the *JEG* or *AEJ: Macro*.

We now receive close to 1000 annual submissions and can only publish 50 papers or so across all fields. You will thus appreciate that the acceptance rate at *JEEA* is less than 5%. For applied papers, I can only see a path for publication for papers that make significant contributions to existing empirical work, irrespective of how clean their identification is. Your paper falls short of what the *JEEA* threshold is, as we aim to become established a top-6 general interest journal.

Although I understand you will be disappointed with this rejection, I hope this particular negative outcome does not discourage you from submitting your work to *JEEA* in the future. I hope you appreciate the relatively quick turnaround and the many constructive comments on your work provided by all four referees, and although the outcome is not one that you would have wanted, you view the process as constructive and useful for your work in future.

Yours sincerely,

Imran Rasul.

Professor of Economics, University College London
Co-Director, ESRC Centre for the Microeconomic Analysis of Public Policy (CPP),
Institute for Fiscal Studies
http://www.homepages.ucl.ac.uk/~uctpimr/

University College London
Drayton House
30 Gordon Street
London WC1H OAX
United Kingdom

i.rasul@ucl.ac.uk
http://www.jeea.org

Referee Report for JEEA MS 9485

What the Paper Does

The paper under review addresses the following question. Consider the aggregate production function for the agricultural sector for each region:

$$Y_i = A_i X_i^{\beta} (K_{Ai}^{\alpha} L_{Ai}^{1-\alpha})^{1-\beta},$$

Parameter β governs the elasticity between regional agricultural output and land input. Does this elasticity differ between topical and non-tropical area?

The key identification strategy is that, assuming free factor mobility, it is straightforward to show that labor-to-land ratio should be proportional to land productivity (A_i). Intuitively, more people choose to work in a region if its land is more fertile. Hence the authors estimate the elasticity between land productivity $\log(A_i)$ and labor-to-land ratio $\log(L_{Ai}/X_i)$ across regions and the coefficient of the linear regression is exactly β.

The paper measures land productivity using GAEZ data and labor-to-land ratio from various sources. Then the authors estimate β separately for tropical and non-tropical regions, and find that it is substantially larger for non-tropical regions. They also show that (as expected) this elasticity is important to structural transformation using a two-sector model.

I think this paper makes an important contribution since this elasticity has never been clearly estimated in the literature. Earlier works, as the authors cited, typically use aggregate data and come to different conclusions. This paper under review uses much richer micro data which are prepared in a very careful way.

Major Concern

My biggest concern is on the assumption of free factor mobility, which is key to the identification. Consider the scenario where agricultural output is mobile across regions but factors of input are not. This is the case of, say, China (due to the *hokou* system that restricts labor mobility). In this case, the labor-to-land ratio is exogenous and is largely independent of the land productivity, and therefore our estimate of β can be biased downwards.

This can be a systematic problem for the following reason. We generally think that factor misallocation is worse in poor countries which makes the labor-to-land ratio less responsive to

land productivity in poor countries. Since there are more poor countries in tropical areas, the estimated β can be more biased downwards in tropical areas. That is exactly consistent with this paper's result that $\beta_{\text{tropical}} < \beta_{\text{temperate}}$.

The authors show in Appendix A.3.1 and A.3.2 that their results remain true even without assuming free factor mobility, but in the arguments they still assume that factors are mobile between agricultural and non-agricultural sectors, and/or agricultural output is immobile across areas, both of which are unlikely to be true. In the China case that I described above, labor is immobile while output is mobile.

One solution to address this problem is that one can look at the wage dispersion within regions (both among farmers and rural-urban comparison) and see if it is indeed larger among tropical areas. IPUMS-international has some of this information but it may not be complete.

Other Concerns

I have some other concerns that I document below.

- Could you please give readers a sense in how to map the land elasticity differences between climate types into elasticity differences between rich and poor countries? The latter seems to be important in any two-sector models. It will be nice to provide a table that includes coefficients for a list of countries. (Country-level coefficients come from aggregation from region-level coefficients and inevitably have some problems but they are still informative for macroeconomists who work on cross-country comparison.) I imagine that it would be well-cited and used.

- On page 9 and 10: You use calories to compare crops, and the crop suitability also focuses on crops like wheat or rice. Hence your study is more on food crops. How about those cash crops in this case? In particular you constructed the suitability index by weighting crops according to their cultivated area. Food crops may have larger cultivated area than cash crops, and then drive the index. It will be great to know the results for cash crops only as they are more market-oriented and may contain some different story.

- Page 21: you assume that agricultural and non-agricultural goods to be substitutes, which seems odd. In particular, you write "...the expenditure share of non-agriculture rises while also having *lower* prices" in the footnote. This is not quite correct as you may

2

see from Figure 10 of Herrendorf et al. (2013). (Of course it depends on whether you assume value added or gross consumption approach, but I think your model maps into the value added approach better as there is no intermediate inputs.) Most papers in the structural transformation literature assume agricultural and non-agricultural goods to be *complements*, if not *Leontief*, rather than *substitutes*. See, for example, Gollin et al. (2002), Gollin et al. (2007), Herrendorf et al. (2013), Yang and Zhu (2013), among others. Assuming $\varepsilon < \gamma$ does not affect your main results, but some predictions of the model may go in the wrong direction, such as Proposition 1(b). I think one solution would be to use the Stone-Geary preferences instead, such as the one in Kongsamut et al. (2001), Restuccia et al. (2008), or Herrendorf et al. (2014). Let me make it clear that I do no mean that this Stone-Geary preferences setup is better than that of Boppart (2014). Rather, the Stone-Geary setup is more commonly used in the literature and hence it provides a better comparison that readers are more familiar with. In this case, you assume the elasticity is unity (Cobb-Douglas) to be simplicity. I understand that the Stone-Geary setup may not have a close-form solution, but the numerical solution should be very simple and some numerical experiments like Gollin et al. (2007) can be helpful for readers to see the impact of β.

- Page 21: Why do you assume an exogenous factor income shares? You implicitly create a wedge in the first order condition, whose role is not transparent.

References

Boppart, T. (2014). Structural change and the kaldor facts in a growth model with relative price effects and non-gorman preferences. *Econometrica*, 82(6):2167–2196.

Gollin, D., Parente, S. L., and Rogerson, R. (2002). The role of agriculture in development. *American Economic Review: Papers and Proceedings*, 92(2):160–164.

Gollin, D., Parente, S. L., and Rogerson, R. (2007). The food problem and the evolution of international income levels. *Journal of Monetary Economics*, 54(4):1230–1255.

Herrendorf, B., Rogerson, R., and Valentinyi, A. (2013). Two perspectives on preferences and structural transformation. *American Economic Review*, 103(7):2752–2789.

Herrendorf, B., Rogerson, R., and Valentinyi, A. (2014). Growth and structural transformation. *Handbook of Economic Growth*, 2:855–941.

Kongsamut, P., Rebelo, S., and Xie, D. (2001). Beyond balanced growth. *Review of Economic Studies*, 68(4):869–882.

Restuccia, D., Yang, D. T., and Zhu, X. (2008). Agriculture and aggregate productivity: A quantitative cross-country analysis. *Journal of Monetary Economics*, 55(2):234–250.

Yang, D. T. and Zhu, X. (2013). Modernization of agriculture and long-term growth. *Journal of Monetary Economics*, 60(3):367–382.

I refereed this paper when it was previously submitted to the EJ. Since then, the paper has been extensively revised. I have deleted below any comments the authors have addressed, and left intact any they have not. I have added clarifying notes on a few points, and have added a few minor points at the end.

1.   I have a general concern that what this paper is doing amounts to production function estimation without panel data, and without instruments there is a gigantic literature (e.g. Levinsohn and Petrin, 2003, RESTUD) on how it is hard to estimate TFP. And yet this paper tells us that by assuming Cobb–Douglas, profit maximization, and free movement of factors within provinces, we can solve for equation (5) and estimate it directly. It appears that there is a large gap between what the users of micro–data do and what macroeconomists do when confronted with the same problem. What the authors do is in the spirit of, say Chari, Kehoe, and McGrattan (2007) or Allen (1979). But can we lean so heavily on functional forms and theory to give us simple estimating equations that claim to tell us many things about the world?

If the exercise here really does hinge on some modeling assumptions, I would like to see evidence in their favor.

i.   [In the original paper, I worried about the Cobb–Douglas functional form, but this has been addressed]

ii.   If the model claims that wages and the return to capital are constant within a province, I would like to see evidence of that from micro–data, at least for some selected countries for which these data are easily available.

iii.   Similarly, there are datasets in which you can measure the assumption [page reference removed] that the share of non–agricultural output paid to labor is equal to the share in agriculture.

iv.   The assumption [page reference removed] that the components of TFP are log–linear with coefficients of one: is this without loss of generality? Is it testable in micro data where TFP estimation can be done more "correctly" or "completely"? Relatedly, how crucial is the assumption of a coefficient of 1 in equation 14?

v.   The paper assumes that land quality is the only component of TFP that varies within a province. It isn't immediately clear what this means in "the real world". That one district can't have better

government than its neighbor? That it can't have older tractors? This too is an assumption that can be tested in micro-data for a handful of countries for which such data exist.

Similarly, the data are a cross-section, not a panel. In a generic sense, this reduces the credibility of the estimation (a paper that can do "within" estimation is generally more credible than one that relies only on "between" estimation). In the specific context of this paper, the outcome variable can only be measured at one point in time, but the main independent variable (effectively log population density) could be measured at any point in history. Are the results different if the HYDE data from 1900 or 1500 are used? If so, do we worry? If not, do we doubt the HYDE data?

There is not much concern about identification here. Fixed effects remove province-level unobservables that correlate with both land quality and population density, but there is little done (e.g. coefficient stability) to show that district-level unobservables are unlikely to explain the results. A simple example: being bisected by a river may improve land quality via water availability while also increasing population because of reduced transportation costs. Of course, this is but one example — there must be hundreds of variables that vary within province and correlate with both the outcome and main independent variables.

2.   [In the original paper, I was worried about the relevance of the paper to Malthusian concepts. This has been addressed.]

3.   [In the original paper, I was worried if micro data would give similar answers. This has been addressed.]

4.   If I take a big step back and think about the paper, the exercise here seems strange. The authors assume that $\ln(Y)=\ln A + b \ln X + (1-b)a \ln K + (1-b)(1-a) \ln L$. Given this equation, I would try to find data on Y, data on X, and an instrument for X. Here, the authors estimate b by regressing $\ln(A)$ on $\ln(L/X)$ and some fixed effects. This follows from the model assumptions, but it seems to be a very roundabout way of doing something straightforward. Should I find it bizarre that a paper estimating the effect of land on output doesn't require data on output?

Minor worries:

•   [In the original paper, I worried about districts as the unit of observation. This has been addressed.]

•   Why use the HYDE data at all? These are a rather opaque source of mostly back-projected population data that are used for historical periods because of their availability, not because of the transparent construction, reliability, or accuracy. Modern development and growth economists use other sources (especially the GPW) so this is a conspicuously odd choice. When these are used for 1900 and 1950, aren't the HYDE data mostly created using back projection? Regressing log density in 2000 on log density in 1900 gives a too-high-to-believe R-squared. Allusions are made to censuses as the basis of the HYDE data, but it isn't clear to me that's how they are constructed. An appendix describing how these are made would help give credibility, presuming the HYDE data are credible.

•   Timing is not clear in this paper. Do we just presume that everything is measured in 2000? The FAO data are somewhat "timeless", the HYDE data appear to be for a specific year, and the lights data appear to be for that year as well. But the year 2000 is a single year… is it just selected for convenience? Why not 2010? 2000 is almost 20 years ago now.

•   [In the original submission, I worried about agricultural land per person. This has been addressed.]

•   Technically, the model predictions refer to agricultural labor, not to population. Is it fair to treat one as an empirical proxy for the other? Is a sufficient assumption that the share of the rural population engaged in agricultural labor is constant within a province?

•   [In the original submission, I wanted the share of labor in agriculture to be validated with IPUMS data. This has been addressed.]

•   It seems strange to make nighttime lights a control. That appears to be an outcome of both land quality and population, among other variables. Why is it on the right hand side and not the left hand side? The same comment applies to urbanization.

•   Why do the estimations not require data on non-land inputs? Is this for the same reason capital is netted out by the province fixed effect?

•   In equation (5), measurement error in population density, which must exist, will systematically bias downwards estimates of (5). Instrumental variables using an alternative measure would be a

standard solution.

•   [I have modified this slightly] The specific form of the indirect utility function should be defended, since it isn't clear to me whether it can be derived from any standard utility function.

•   Proposition 1c (and generally): should "population" be universally replaced as "the labor force"? These concepts seem to be used interchangeably in the paper and it is not clear if the reader should be misled into believing that the labor force in the model and the population in the data are truly the same.

•   [In the initial submission I worried about low quality land. This has been addressed.]

•   Should national parks and forest reserves be removed from the data?

•   Estimating results separately by region is one way to give us an idea of the variation across the world in beta estimates. Is there any way to plot these on a map? I don't think there's one "right" way to do this, but I have two suggestions. First, estimate a random coefficients model and plot the coefficient estimated for each district. Second, for each district, estimate results keeping districts only within a certain distance of it, and plot that number as the estimate on the map. Researcher-imposed sample splits cannot be the only valid way to examine heterogeneity in the data, nor can it be the best way to depict it visually.

•   [In the initial submission I worried about factor shares. This has been addressed.]

•   There doesn't seem to be much consideration of prices or cash crops in this paper. If my output is not consumed locally, would that change the model and invalidate the estimating equations? If my government subsidizes or taxes cash crops, is that an omitted variable that we can simply ignore in the main regressions? In section 3.6, it seems bizarre to drop districts below the 25th percentile in staple crop production. This just drops small districts, whether or not they are dependent on cash crops. This dropping should be based on the importance of cash crops (or their relative productivity) within the district, e.g. 25% of this district's output is staple crops, or e.g. the most high productivity crop in this district is a stable crop. Also: how are cash crops defined?

•   [I have modified this slightly] The section that replicates Acemoglu should be cut. It is somewhat unrelated to the rest of the paper, builds on a paper that has been heavily criticized, involves arbitrary cuts at the median, seems to forget that the main point of the Acemoglu paper was a zero effect on GDP.

New Points:

•   The authors have made it clear that they are estimating an aggregate production function that need not be approximated by any individual or farm-level production functions that can recovered from micro data. That is sensible, though they should comment on the issues involved in aggregating production functions. The Cambridge capital controversy may be relevant here.

•   On page 23 there is a space missing after the beta.

The paper estimates the elasticity of agricultural output with respect to land. It finds that the elasticity in temperate regions is significantly higher than that in tropical regions. As the effect of shocks in population or technology on living standards is increasing in this elasticity, the findings have implications to our understanding of the differences in economic development between temperate and tropical regions. The paper presents some evidence that supports this prediction and are consistent with the estimated differences in elasticities across regions of the world.

I believe this is an interesting and important paper that is using a cleaver approach to estimate the elasticity of agricultural output with respect to land, without simply using factor shares. Using factor shares requires strong assumption, in particular that factors are paid their marginal cost and information on land shares are available and distinct from other infrastructure attached to the land.

However, I have several concerns regarding the estimation method presented in section 2, and the estimation in section 3:

You write before equation (2) "In each district those wages and returns are determined by [equation 2]" But (2) seems to me like an identity. Are you claiming that the share of labor and the share of capital determine wages? Doesn't it go in the other direction?

Moreover, don't you assume in (2) that the shares of factors are identical across all districts within a province? Isn't this justified by the combination of assuming that the elasticities are identical across districts and that factors are paid their marginal product? If so, the claim you do not rely on prices equal marginal products is wrong. If not, how can you justify the claim that all districts share the same factor ratio? And if firms do not employ according to the marginal product, how do markets clear?

All this leads me to conclude that, unless I'm missing something, you should provide an explanation for how wages are determined, if not according to marginal products, and explain the optimization of firms and market clearing. But if I'm missing something, just clarify it.

You use the FAO data for measures of productivity. But, as far as I understand, FAO provides estimates of expected output per unit of land. Are these estimates based on identical quantities of labor and capital per unit of land within all regions of the world? If yes, then you have a perfect estimate of productivity. But I'm not sure that's the case. If FAO provides estimates for potential output based on some "required" levels of factors that can perform the tasks of plowing, sowing, harvesting, maintaining irrigation systems, storage of grain, and so on, the quantities of factors could vary with productivity. In the extreme case that factors are proportional to output then FAO doesn't provide any relevant data on productivity. In the more reasonable case, the data is somewhere in between the two options of identical factors in all regions and factors proportional to output. That is, factors increase with output, but not proportionally. Some of the required labor, one could assume, isn't affected by land productivity (perhaps when plowing and sowing), but for others tasks (such as harvesting and storing), one can reasonably assume that the required labor is increasing with land productivity. In that case, I'm concerned, there will be some systematic bias in the estimates. This potential bias should be addressed.

My major concern is that there could be systematic differences in the FAO predicted output and required factors ratio between temperate zones and tropical zones. The suitable grain or other crop perhaps isn't the same in temperate and tropical regions, and different crops could require different capital labor or capital land ratios. Moreover, it could be the case, that the most suitable crop (maximum potential calories per unit of land) according to FAO isn't the crop that is used in that area.

I guess that these differences between temperate and tropical zones could bias the estimates and might show a difference in elasticity when such a difference doesn't exist, but perhaps one can still claim that the data and methodology support the authors claims. These concerns should be addressed.

Minor comments:

1.   The subscript A, denoting factors employed in agriculture, seems redundant, and is used inconsistently.

2.   It seems (5 – 6) should be $\ln(L/X)$ instead of $\ln L/X$.

3.   Footnote 8 repeats a statement in the main text on page 5.

4.   You show in an extension to the model that as the elasticity gets higher, the economy gets more sensitive to population and technological change. This seems rather obvious and doesn't really require a model, in particular since, as you mention, this is a standard result in the literature. Perhaps you can consider moving the formal analysis to an appendix and just keep the intuitive explanation, which is straightforward.

Report on "The role of land in temperate and tropical agriculture", MS#9485

Summary of Paper

This paper neatly blends simple theory and empirics to provide a new approach to estimating the elasticity of agricultural output with respect to land. The authors show that, provided the agricultural production function is fixed within a province, capital and labour mobile, and the aggregate agricultural production function has constant returns to scale in capital, labour and land, the elasticity of agricultural output can be estimated on within-"province" data by regressing the ratio of labour to land on agricultural TFP.

To obtain estimates of agricultural TFP, the authors use GAEZ data on crop productivity to estimate the calorific potential of the land in each district following Galor and Ozak (2016), albeit using a different subset of crops. Estimates of population are obtained from a variety of sources. The authors estimate separate elasticities of agricultural output with respect to the supply of land for tropical and temperate areas (as defined by the type of crops grown, or whether an area is frost free), finding that the elasticity of agricultural output is higher in temperate areas than tropical areas. They go on to use a two-sector model to show that an implication of this is that output per capita should respond more strongly to changes in population or technology in areas with lower elasticity. Lastly the authors estimate country specific elasticities of agricultural output with respect to land and use them to show that - as predicted by their model - the negative impact of the epidemiological transition on the living standards of survivors is larger in countries with a stronger elasticity of agricultural output to land.

Comments

1. Are returns to capital and labour really equalised across space? Because the analysis in the paper is primarily cross-sectional - based on the correlation between population density and agricultural productivity - it relies heavily on theory to give meaning to the results. As discussed above, this theory relies on the assumption that wages and returns to capital are equalised across space, however, a body of recent literature strongly indicates that this may not be the case either for labour markets (e.g. Bustos et al., 2016; Gollin et al., 2013; Munshi & Rosenzweig, 2016) or for capital markets (e.g. Bustos et al., 2017; Marden, 2017) in developing countries. It is unclear to me what the implications of allowing the real wage and cost of capital to vary over space for the results of the paper. It seems plausible that we would still expect a stronger correlation between population density and agricultural productivity in areas with a higher elasticity of output with respect to land – and this is consistent with the authors panel results - but it seems unlikely that this elasticity is would still be the estimated coefficient. It's possible that a richer model could bound and or sign the degree of bias from this type of misallocation, which would help generalise the results.

An empirical analysis restricted to countries and regions where capital and labour markets worked effectively could allay these concerns. Unfortunately, the lack of a material number of tropical regions in developed countries limits the scope for such an exercise. Moreover, the

elasticity of agricultural output with respect to land is surely of greater interest in developing countries.

2. Population data. The authors main source of population data is HYDE data which interpolates population data to provide a global gridded population data set. The authors use this data to create a global dataset of "district" level population using 2nd level administrative boundaries map of administrative boundaries.

As the authors make clear, the interpolation algorithm used to construct the HYDE data uses land-productivity as an input. To avoid estimating the interpolation algorithm, the authors conduct their analysis not at the grid-cell level but at the district level. This is only sufficient, if the data population data underlying HYDE is also at the district level, but this doesn't appear to be the case. Popustat, the input to HYDE cited in Goldwijk et al. (2011), often provides population data for higher level administrative regions than those used in the paper, so at the level of aggregation used in the paper, the authors may still be "estimating the algorithm" – exactly as they were attempting to avoid doing. In practice, it may be that this issue is mitigated for more recent periods by HYDE's use of Landscan population density patterns (which I am not familiar with), but this issue deserves further consideration.

The authors also present results result based on two other population data sources. Gridded "GRUMP" data and non-gridded IPUMS census data. I don't know enough about GRUMP to comment on its suitability for the authors purpose and the extent to which similar issues of "estimating the algorithm" are likely to arise. However, the use of IPUMS data seems clearly preferable to the use of interpolated data where possible, and unless the case for the use of interpolated can be made much more convincingly, using the IPUMS data as the principle data-set would make the results more convincing. The drawback would be to limit geographic coverage, but GRUMP and HYDE used to validate the results for the rest of the world.

When the authors use the IPUMS data, they obtain quite different results – there is no relationship between agricultural population density and productivity in the tropics if the IPUMS data is used (although the differential with temperate regions remains). The sample of districts in IPUMS is quite different so it is unclear whether this difference is due to the sample or the data source. The authors could reassure the reader as to the validity, or not, of the interpolated data by using the same sample for all data sources.

4. Panel Estimation. Given the necessarily small sample of countries, the evidence from the panel analysis, which interacts the elasticity of land to agricultural output and shows, is interesting but hardly determinative. The set of countries, and their associated elasticities, should be provided so as to help the reader assess the extent, or not, that other factors could plausibly be driving this relationship.

4.Appendix. The paper makes references to an Appendix that was not supplied with the paper and I have not been able to review as it was not obviously available on the authors websites. This means I have not, for example, reviewed the algebra feeding in to section 4.1. It may be that this was lost in the editorial system.

5.Standard errors. As a robustness check, standard errors should be clustered by country as an alternative to Conley errors as, in my experience, clustered errors tend to be a bit more conservative when specified at an appropriate geographic unit. This seems unlikely to affect the results, so could simply be flagged on page 8 (rather than providing an additional set of tables).

6. Visual Aids. I would have found it helpful to have more visual aids to help understand where the variation in productivity and population density that exist were to be found, and where in the world high elasticities of agricultural output with respect to land would be found.

7. Ruggedness. Slope and elevation are inputs to the GAEZ data, with more rugged areas tending to have lower agricultural productivity. Because ruggedness also impacts trade costs, it has the scope to affect the population-to-land ratio independently of its affect through land productivity (less rugged areas may be more intensively farmed holding productivity constant). The relationship may vary by crop and by region, so the authors should include an additional robustness check controlling for ruggedness.

8.Typos. The equation number identifying the wage definition at the bottom of page 5 should say (2) not (3). There is a "see: that should read "seen" at the top of page 24. (This is probably not a complete list.)

Bibliography.

Bustos, P., Caprettini, B., & Ponticelli, J. (2016). Agricultural productivity and structural transformation: Evidence from Brazil. *American Economic Review*, *106*(6), 1320-65.

Bustos, Paula, Gabriel Garber, and Jacopo Ponticelli. (2017). "Capital Accumulation and Structural Transformation."

Marden, S. (2017). The agricultural roots of industrial development: rural savings and industrialisation in reform era China.

Gollin, D., Lagakos, D., & Waugh, M. E. (2013). The agricultural productivity gap. *The Quarterly Journal of Economics*, *129*(2), 939-993.

Munshi, K., & Rosenzweig, M. (2016). Networks and misallocation: Insurance, migration, and the rural-urban wage gap. *American Economic Review*, *106*(1), 46-98.