How Tight are Malthusian Constraints?

Identification and Estimation

The intuition of what we are doing is easiest to see in the attached figure (please see the top of the next page). The labor demand curves for different districts are, as normal, downward sloping. We assume that each district faces an infinitely elastic labor supply curve, given mobility of labor between districts, so that the labor supply curve is horizontal.

The location of the labor demand curves differs across districts because of differences in agricultural TFP (A), which shifts the demand curves up or down relative to one another. But knowing A for different districts (and we'll come back below to whether we can measure this well) we can compare the difference in (log) A to the difference in (log) L/X to recover an estimate of the slope of the labor demand curve.

In math, let us back up prior to the original equation (5) in the text. The labor demand curve could be written as

$$w_{is} = \phi_L A_{is} \left(\frac{X_{is}}{L_{Ais}} \right)^{\beta}.$$

which states that the wage paid in district i of state/province s is equal to some fraction ϕ_L of the average product of labor in that district. We suppressed the notation regarding capital here to keep things cleaner.

The assumption that labor is mobile between districts means $w_{is} = w_s$ is the horizontal labor supply curve for each district, leading us to

$$w_s = \phi_L A_{is} \left(\frac{X_{is}}{L_{Ais}} \right)^{\beta}.$$

In logs,

$$\ln w_s = \ln \phi_L + \ln A_{is} - \beta \ln L_{Ais}/X_{is}.$$

The value of β can be identified from this equation because of the assumption of a horizontal supply curve. In what one might call a "textbook" case, you could assume that A is constant across units (e.g. firms or locations) and use variation (e.g. over time) in the w_s along with varation in L/X to estimate β. This would be allowing the horizontal supply curve to move up and down to trace out a single demand curve.

But given the above equation, that is not the only way to identify β. For us, we have many districts within a state/province that all face the same w_s, so the supply curve is common. Variation in A can be used along with variation in L/X to identify β, as shown in the figure. Given the structural relationship above, L/X and A are related through the parameter β. But they are related this way however one might rearrange this equation, with either L/X or A on the left or right.

In the original text, our equation (5) had the L/X term on the left, as a function of A. This facilitated solving the model further. The state/province specific term in the original equation (5) was the wage w_s solved out in terms of all the district-specific productivity terms.

Our specification equation (12) in the text - with A on the left - is just another way of rearraning the last equation, and relies on the same identification strategy. But this leap was not explained clearly, for sure, as

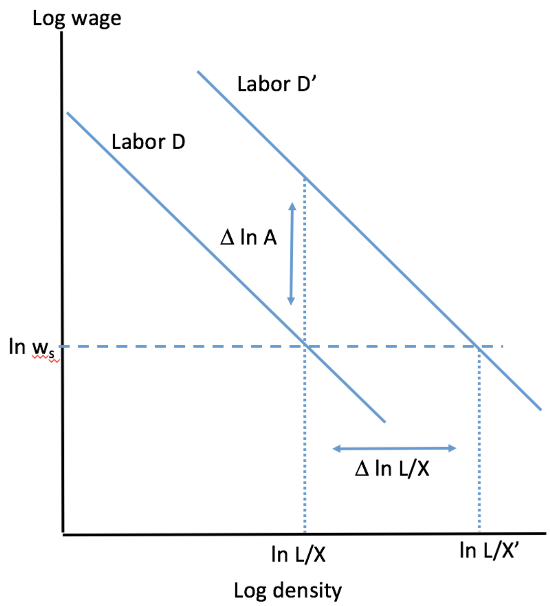

Figure 1: Labor demand and supply

it not only involved rearranging the equation, but also adding in the error term and some controls. We took too many steps at once in going from (5) to (12). In presenting the paper this had not been a sticking point, and so as the paper evolved a more complete explanation of this did not get incorporated into the text.

So let's establish how we see equation (12) coming about. Rearrange the last equation above as

$$\ln A_{is} = \beta \ln L_{Ais}/X_{is} + \gamma_s,$$

where $\gamma_s = \ln w_s - \ln \phi_L$ is the state/province specific term that captures the wage level. This equation relates the "true" level of productivity to the "true" labor density in the district. This is an equilibrium relationship, but does not imply that somehow a change in L/X would alter underlying agricultural productivity.

Our actual measure of district-level productivity is a proxy for true productivity,

$$\ln A_{is}^{GAEZ} = \ln A_{is} + \epsilon_{is}$$

where ϵ_{is} is the deviation of the GAEZ data from the truth. This gives us

$$\ln A_{is}^{GAEZ} = \beta \ln L_{Ais}/X_{is} + \gamma_s + \epsilon_{is}.$$

If we take this specification to the data, then whether $\hat{\beta}$ is a consistent estimate of β depends on whether ϵ_{is} is related to L/X.

One particular concern is that high productivity in agriculture may have, in the past, led to more urbanization in a district, and that urbanization may push down L/X because the urban area takes up space. That would create a negative correlation of L/X and ϵ_{is}, and mean our estimates of β are biased. To address that, we include control variables for urbanization and night light density, denoted with the vector Z_{is}.

$$\ln A_{is}^{GAEZ} = \beta \ln L_{Ais}/X_{is} + \gamma_s + \delta' Z_{is} + \epsilon_{is}.$$

This is our main estimating equation, derived from the labor demand curve and assumption of a horizontal labor supply curve. As noted, we did not do a good job in the existing paper in working from the original equation (5) to this specification.

Whether this specification is delivering consistent estimates depends on whether one still believes there is a correlation of ϵ_{is} and L/X, even after controlling for the state/province fixed effects, urbanization, and night lights. We think our baseline specification and our various robustness checks in the paper make a good case that we are getting an unbiased estimate, but as with any empirical work a reader may not find the case compelling.

And finally, we've got to consider measurement error in L/X as well. This would create attenuation bias in our estimate of β. We address this in section 3.6, where we also discuss sources of measurement error in the productivity term, A. In both cases we think that the measurement error is not driving our results, but again a reader may have an argument with that.

Our point is that our specification - equation (12) in the original paper - is a plausible one. We are fully aware of the possible issues in getting unbiased estimates from that equation, both from measurement error in the data and from the failure of the underlying assumption of horizontal supply curves. Within the paper, we take each issue seriously and provide what evidence and analysis we can to support our work.

We are not trying to do a reduced form policy evaluation, where one would expect to have a policy variable on the right, and an outcome variable on the left. Instead, we have a structural relationship between A and L/X that holds given our assumption about the horizontal labor supply curve. There is no such thing as a policy and an outcome variable here. There are two variables - L/X and A - that are related through a structural parameter β. We are not asking whether a change in L/X has a causal effect on agricultural productivity. We are only trying to estimate the relationship between L/X and A, which given our assumptions must inform us about the underlying parameter β. That said, whether L/X is mis-measured, or whether it remains correlated with ϵ_{is}, are legitmate concerns for whether the relationship we measure is accurate in picking up the true β or not.

Thank you for taking the time to consider this, we appreciate the attention you're giving it. Our apologies for the original paper not being clear in how we arrived at our estimation strategy.