Another look at the Precision of IV estimates of the Development Effects of Access to Electricity in Brazil *

Molly Lipscomb
University of Virginia

Ahmed Mushfiq Mobarak
Yale University

Dimitri Szerman
Universität Mannheim

September 15, 2021

*We thank Gunther Bensch and Jorg Peters for bringing to our attention the discrepancy in the definition of the Amazon in Lipscomb et al. (2013).

Abstract

Lipscomb et al. (2013) uses two different definitions of the Amazon region in different stages of analysis when estimating the development effects of hydropower in Brazil. In this corrigendum, we present results with alternative definitions of the Amazon region. Using the revised data, LMB's preferred specification leads to a weak instrumental variable. We show that first stage power is retained under reasonable alternative specifications, and the second stage results in those cases are similar to the those originally reported. We also report corrected standard errors according to the Lee et al. (2020) *tF* procedure.

The Instrument To estimate the development effects of electricity, Lipscomb et al. (2013) (henceforth, LMB) construct an instrument using variation in the cost of producing hydropower based on local geographic characteristics, together with variation in the budget for hydropower across decades. They generate an instrument by matching existing hydro-electric power plants to geographic variables in a probit regression to estimate the relative importance of each geographic factor (Table 1 in LMB). This creates a "suitability index" that ranks every location in Brazil in terms of how sensible it is to place a hydropower dam at that location, if one were solely focusing on geographic factors. The underlying parameters for this index were water flow accumulation, average and maximum slope in the river, and an indicator for whether the point was in the Amazon. The Amazon coefficient was large and negative in this probit regression because building infrastructure in the rainforest is costly. Generation plants were therefore relatively unlikely to be placed there by the model. The time variation in the instrument comes from variation in the number of hydropower plants built in each decade. These are combined in an algorithm that matches the number of plants built in a decade to the actual number of new hydropower plants built by Brazil in that decade. The most "suitable" locations that had not yet been electrified receive those plants.

The Specification LMB add county fixed-effects in their estimation so as to not use any cross-sectional variation for identification. This isolates the identifying variation in the instrument to be the cutoff values for the last locations that received plants until the budget was exhausted for that decade, and the next most suitable areas that barely missed the threshold and had to wait for another decade for a plant. To mitigate concerns of common, spurious trends between the predicted grid expansion and development indicators, LMB directly control for any differential trends by adding geographic-specific time fixed-effects in their 2SLS estimation. Several specifications are possible, each using a different geographic characteristic. In their preferred specification, LMB include interactions between the Amazon dummy and time-fixed effects in the first and second-stage regressions, which accounts for differential dynamics in Amazon counties. That is the specification that LMB use when reporting their main results (Tables 3, 6, and 7) and mechanisms (tables 11–14). In Table 10, LMB show alternative specifications that include different geographic-specific time fixed-effects.

The Mistake The mistake in LMB stems from an inconsistency between the Amazon dummy used in constructing the instrument and the one included in the 2SLS estimation. When constructing the instrument, LMB equated the Amazon to Brazil's North Region, as defined the Brazil's official statistical agency, IBGE. When running the 2SLS analysis, LMB used an extended definition of the Amazon that included the Pantanal–another humid tropical forest.

The Correction We correct the inconsistency in the definition of the Amazon region. Table 1 reports the first-stage results consistently using one of three potential definitions of the Amazon. Each row of Table 1 corresponds to a different specification controlling for geographic-specific time trends, in the spirit of Table 10 in LMB. To facilitate the comparison with the original study, we add the first-stage results published in LMB's Table 10 in columns 1 and 2. In the remaining columns, we use three alternative definitions of the Amazon. Columns 3 and 4 use an ecological definition of the Amazon and Pantanal. That definition is, conceptually, the closest in spirit to what

Table 1: First Stage Results: Various Specifications and Amazon Definitions

Amazon Definition	AEJ Printed		Biome		Legal		Extended	
Specification	(1) First-Stage	(2) F-Stat	(3) First-Stage	(4) F-Stat	(5) First-Stage	(6) F-Stat	(7) First-Stage	(8) F-Stat
1. Water Flow × budget	0.32*** (0.05)	45.8	0.26*** [0.04]	37.05	0.20*** [0.04]	22.19	0.26*** [0.05]	30.67
2. River Gradient × budget	0.32*** (0.05)	47.8	0.26*** [0.04]	36.51	0.19*** [0.04]	21.04	0.26*** [0.05]	30.08
3. Amazon dummy × budget	0.22*** (0.04)	25.4	0.05 [0.04]	1.94	0.01 [0.04]	0.05	0.06 [0.04]	2.14
4. Water Flow × budget, Amazon dummy × budget	0.22*** (0.04)	24.8	0.05 [0.04]	2.03	0.01 [0.04]	0.07	0.06 [0.04]	2.19
5. River Gradient × budget, Amazon dummy × budget	0.22*** (0.04)	25.3	0.05 [0.04]	1.86	0.01 [0.04]	0.04	0.07 [0.04]	2.26
6. River Gradient × budget, Water Flow × budget	0.32*** (0.05)	48.6	0.25*** [0.04]	36.47	0.19*** [0.04]	20.87	0.26*** [0.05]	29.90
7. River Gradient × budget, Water Flow × budget, Amazon dummy × budget	0.22*** (0.05)	24.4	0.05 [0.04]	1.89	0.01 [0.04]	0.05	0.06 [0.04]	2.20
8. Water Flow × year dummies	0.32*** (0.05)	45.6	0.26*** [0.04]	37.13	0.20*** [0.04]	22.21	0.26*** [0.05]	30.89
9. Amazon dummy × year dummies	0.22*** (0.04)	24.6	0.06 [0.04]	2.05	0.01 [0.04]	0.05	0.07* [0.04]	2.92
10. River Gradient × year dummies	0.22*** (0.05)	47.8	0.26*** [0.04]	36.29	0.19*** [0.04]	20.90	0.26*** [0.05]	29.86
11. Water Flow × year dummies, Amazon dummy × year dummies	0.22*** (0.05)	23.9	0.06 [0.04]	2.17	0.01 [0.04]	0.07	0.08* [0.04]	3.03
12. River Gradient × year dummies, Amazon dummy × year dummies	0.22*** (0.04)	24.9	0.05 [0.04]	1.95	0.01 [0.04]	0.02	0.08* [0.04]	3.11
13. Water Flow × year dummies, river Gradient × year dummies	0.32*** (0.05)	48.3	0.26*** [0.04]	36.44	0.19*** [0.04]	20.81	0.26*** [0.05]	29.96
14. River Gradient × year dummies, water Flow × year dummies, Amazon dummy × year dummies	0.22*** (0.05)	23.9	0.05 [0.04]	2.00	0.01 [0.04]	0.04	0.08* [0.04]	3.04
15. Quartic suitability rank × year dummies	0.24*** (0.04)	29.1	0.22*** [0.05]	20.97	0.14*** [0.05]	8.55	0.24*** [0.05]	21.01

Notes:

LMB had in mind: a proxy for a geographic factor—dense forests—that increase construction costs of hydroelectric power plants. Columns 5 and 6 use an administrative definition of the Amazon, known as Legal Amazon, which includes savannah-like areas, and some policies are designed to target those boundaries. Finally, columns 7 and 8 use the "extended" definition originally used by LMB.

Two conclusions can be drawn from Table 1. First, for any given specification, the exact definition of the Amazon hardly matters. Second, for any given Amazon definition, the instrument becomes weak and therefore invalid, in LMB's preferred specification (row 9). In fact, the table shows that the instrument becomes weak whenever the specification directly controls for Amazon-specific time fixed-effects (see rows 3,4,5,7,11,12,14), and retains first-stage power when it does not (rows 2,6,8,10,13,15). The Amazon-specific trend controls simply absorb too much variation from the instrument.

Based on these considerations, we re-estimate the 2SLS specification with the following changes. First, we use the Biome definition of the Amazon in all regressions and apply it consistently across all stage of analysis, because that is, conceptually, the closest in spirit to what LMB wanted. Second, we replace LMB's preferred specification (shown in row 9 in Table 1) since the first-stage is now weak, with a specification that controls directly for the levels and differential trends in the index of cost factors used in construction of the instrument, reported in row 15. This alternative specification controls directly for time variation related to the estimated geographic suitability of a location for electricity generation. The first stage F-statistic in our new preferred specification is 21, lower than the 24.6 reported in LMB, but above the commonly used cutoff of 10 suggested by Staiger and Stock (1997). Because the F-statistic is below the 104 suggested by Lee et al. (2020) we also provide standard errors that have been corrected at the 95% level using their tF procedure.

Results Compared to Lipscomb et al, 2013 In table 2 we compare the results from the new preferred specification using the quartic of the suitability rank interacted with decade fixed effects as controls to those published in LMB. Of 15 key development indicators used as dependent variables in LMB, 6 are statistically significant (with 95% confidence) in the new specification while 12 were statistically significant in the original paper. Using the Lee et al. (2020) tF standard errors correction procedure, only 3 remain statistically significant at the 95% level.

LMB focused on two central dependent variables: housing values and the human development index (HDI). Rows 1 and 2 of Table 2 shows that the estimated effect of a 10% increase in electrification on housing values changes from $+881$ reais to $+421$ reais (a 52% decrease). For the human development index, it changes from $+0.01$ point increase to a $+0.02$ point increase for a 10% increase in electrification.

The impact of electricity access on HDI and other poverty measures was shown in table 11 of LMB, they are now shown in rows 3-8 of table 2. The estimated effect on HDI education decreases by 35% (from 0.19 to 0.12). The impact on HDI-health was neither statistically significant in the original paper, nor here. The estimated effect on the income component of HDI is smaller in the new specification and not statistically significant. Also shown are effects on infant mortality, income per capita and poverty.

The employment measures of economically active and formal employment remain statistically significant, but the point estimates are smaller. Effects on less than 4 years in school and number of years in school are no longer statistically significant. Human capital accumulation, on the other hand, has a larger coefficient and is now statistically significant while it was not under the original specification. This changes the interpretation of how electricity affects educational choices. Human capital is measured as the residual from a labor productivity equation, so it appears that electricity access improves labor productivity more than it changes school completion rates.

Rows 14-16 of Table 2 show the effects on a few outcomes that LMB was trying to rule out as mechanisms. Effects on life expectancy and population density were statistically insignificant in both the original paper and here. Effects on urbanization also becomes statistically insignificant. LMB ran these regressions to examine whether the documented effects of electrification on development outcomes were the result of simple population movements, and we can now say with a bit more confidence that they were not.

Other coding notes When preparing this corrigendum, we noticed that the original Matlab code failed to set a random seed for the grid simulations, making exact replication of the original instrument impossible. We amended the code to make it replicable.

Table 2: Corrected Estimates based on new specification

	AEJ Printed Results		Corrected Results	
	(1) OLS	(2) IV	(3) OLS	(4) IV
Results from Table 6 in LMB				
Housing Values	0.80***	8.81***	0.20*	4.21*
	[0.27]	[3.03]	[0.12]	[2.16]
				(2.83)
Results from Table 7 in LMB				
Human Development Index	0.01	0.11***	−0.00	0.20***
	[0.01]	[0.04]	[0.00]	[0.06]
				(0.08)**
Results from Table 11 in LMB				
HDI: Education	0.03***	0.19***	0.01***	0.12**
	[0.01]	[0.06]	[0.00]	[0.06]
				(0.08)
HDI: Longevity	0	−0.01	0.01***	0.04
	[0.01]	[0.05]	[0.00]	[0.06]
				(0.08)
HDI: Income	−0.03*	0.45**	−0.08***	0.15
	[0.02]	[0.15]	[0.01]	[0.14]
				(0.18)
Infant Mortality	−7.99***	−11.97	−14.62***	−77.03***
	[2.42]	[18.08]	[0.895]	[19.761]
				(25.89)**
Gross Income PC	−0.01	0.11**	−0.02***	−0.04
	[0.01]	[0.05]	[0.00]	[0.05]
				(0.07)
Poverty	−0.76	−42.17***	4.80***	−21.19
	[1.39]	[13.84]	[0.43]	[13.54]
				(17.73)
Results from Table 12 in LMB				
Economically Active	0.011*	0.173***	0.00	0.06*
	[0.01]	[0.05]	[0.00]	[0.04]
				(0.05)
Formal Employment	0.010*	0.184***	0.00	0.09**
	[0.01]	[0.05]	[0.00]	[0.04]
				(0.05)
Urban Employment	0.01*	0.18***	0.00	0.05
	[0.00]	[0.05]	[0.00]	[0.04]
				(0.05)
Rural Employment	0.01	0.17***	0.01***	0.11***
	[0.01]	[0.05]	[0.00]	[0.04]
				(0.05)**
Results from Table 13 in LMB				
Less than four years education	−0.36	−21.25***	2.26***	0.65
	[0.90]	[7.75]	[0.25]	[6.70]
				(8.77)
Years in School	0.06	2.02***	−0.16***	0.28
	[0.08]	[0.67]	[0.02]	[0.61]
				(0.79)
Human Capital	2.09	11.54	0.90***	19.93**
	[0.41]	[7.30]	[0.11]	[9.21]
				(12.07)
Results from Table 14 in LMB				
Life Expectancy	−0.44	−1.03	−0.52***	−2.20
	[0.32]	[2.39]	[0.10]	[2.24]
				(2.93)
Population Density	−1.11	−23.62	−8.86	3.79
	[3.74]	[19.20]	[6.09]	[13.47]
				(17.64)
Urban percent of pop	0.01	0.24**	−0.02***	0.03
	[0.01]	[0.11]	[0.00]	[0.12]
				(0.16)

Notes: Standard errors clustered by county reported in brackets, standard errors corrected using the Lee et al. (2020) *tF* procedure at the 95% level reported in parentheses. All regressions have county size weights.

References

Lee, D. S., McCrary, J., Moreira, M. J., and Porter, J. (2020). Valid t-ratio inference for iv. *working paper*.

Lipscomb, M., Mobarak, A. M., and Barham, T. (2013). Development Effects of Electrification: Evidence from the Topographic Placement of Hydropower Plants in Brazil. *American Economic Journal: Applied Economics*, 5(2):200–213.

Staiger, D. and Stock, J. (1997). Instrumental Variables Regression with Weak Instruments. *Econometrica*, 65:557–586.