Re-visiting the economic impact of the Guaraní Jesuit missions in South America: A replication of Valencia Caicedo (2019)

Wijdan Tariq
University of Toronto
wijdan.tariq@mail.utoronto.ca

Rohan Alexander
University of Toronto
rohan.alexander@utoronto.ca

February 21, 2021

Abstract

We replicate Valencia Caicedo (2019) and find several inconsistencies between the code provided by the author and the results published in the paper. We also identify various sensitivities to specific values used. We show that when these aspects are adjusted for, some of the main findings do not persist. In particular, we do not find a significant relationship between former Jesuit presence within the Guaraní area and modern literacy rates, median years of schooling, poverty levels, or income.

JEL Codes: I25; N36; O15; O43; Z12.

Keywords: Education and Economic Development; South America; Income Distribution; Institutions and Growth; Religion; Reproducibility; Replication

1 Introduction

In about 1609, the Jesuit order of the Catholic Church established 30 Christian missions in remote parts of what are today Brazil, Paraguay, and Argentina. The missions were established among the indigenous Guaraní people primarily to convert the locals to

Christianity. Jesuit missionaries also provided education for children, including 'how to read and write and do basic arithmetic,' and training for adults including 'masonry, wood carving, and embroidery' until being expelled in 1767 due to events in Europe (Valencia Caicedo 2019, 515).

Valencia Caicedo (2019) examines the Jesuit human capital intervention and finds that areas where the Jesuit order founded religious missions have higher educational attainment and incomes today. Valencia Caicedo (2019, 508) finds that '[i]n municipalities where Jesuits carried out their apostolic activities, literacy rates and median years of schooling are 10%–15% higher today, 250 years after their expulsion. The same locations are also ahead by 10% in terms of per capita income.' The particular independent variable that is most of interest in Valencia Caicedo (2019) is the distance between a given area and the site of the nearest Jesuit mission.

Using replication materials provided by Valencia Caicedo (2018) we document several inconsistencies between the code (Valencia Caicedo 2018) and the paper (Valencia Caicedo 2019). When we adjust the code to match what is described in the paper, we are unable to draw the same conclusions.[1].

We perform a narrow replication of Valencia Caicedo (2019) and do not attempt to place our paper within the broader literature that Valencia Caicedo (2019) contributes to. We find that Valencia Caicedo (2019) is reproducible because the data and code are available, and they can be used to reproduce the results in Valencia Caicedo (2019). But we find that the paper is not replicable because the code does not do what is described in the paper.

The replication materials in Valencia Caicedo (2018) were critical in allowing us to examine Valencia Caicedo (2019). Economics, as a discipline, is in the middle of a reproducibility transition. The provision of the code and data underlying a paper is something that is being increasingly adopted by journals. However, there are additional measures that would also be helpful as the discipline repositions.

We acknowledge and are grateful for the assistance provided by the author of Valencia Caicedo (2019) in guiding our replication. The remainder of this paper is as follows. Section 2 briefly details the data and model relevant to our replication, Section 3 compares our adjusted results with Valencia Caicedo (2019), and finally Section 4 provides context for our findings and recommendations for moving forward.

2 Data and model

Valencia Caicedo (2019) puts together a detailed area-based dataset drawing on archival sources such as 'the location, year of foundation, population, and general workings of

[1]The code for our paper is available at: https://github.com/RohanAlexander/replication-Valencia_Caicedo

the Guaraní Jesuit missions,' historical national censuses, as well as modern sources such as household surveys, modern censuses, and geographic and weather data (Valencia Caicedo 2019, 517). The areas of interest cover the modern provinces of Misiones and Corrientes in Argentina, the state of Rio Grande do Sul in Brazil, and the departments of Misiones and Itapúa in Paraguay. There were 30 Jesuit missions established in total, with '15 in Argentina, 8 in Paraguay, and 7 in Brazil' (Valencia Caicedo 2019, 510). These missions were large, for instance, '[a]t their peak, the Guaraní Jesuit missions contained more than 120,000 inhabitants... four times the population of Buenos Aires in 1779' (Valencia Caicedo 2019, 515).

The main comparison of interest is outcomes in these areas that had established Jesuit missions with outcomes in areas where either the missions were quickly abandoned or where Franciscan missions were established. The main explanatory variable of interest is the distance between an area and a Jesuit mission. Valencia Caicedo (2019, 510) adjusts for a large number of geographic and weather variables and finds that 'moving 100 km closer to a [Jesuit] mission brings a positive effect on educational attainment of around 0.7 years of schooling, and lower poverty rates... of 10%.' No effect was found from abandoned Jesuit missions or Guaraní Franciscan missions.

The key estimating equation is (Valencia Caicedo 2019, 519):

$$Y_{2000,i,j} = \alpha + \beta d(M_{ij}) + \gamma \text{GEO}_{ij} + \mu_j + \epsilon_{ij} \tag{1}$$

The dependent variable, Y_{2000}, is for municipality i in state j and refers variously to: the illiteracy percentage for people aged 15 years and older; median years of schooling; log income; or a poverty index. M refers to whether a mission was present in municipality i in state j, and d is variously a dummy variable or the distance to the nearest mission. GEO is a vector of control variables to account for geography and weather. These include 'area, altitude, longitude, temperature, rainfall, ruggedness, slope, distance to nearest river, distance to the nearest coast, and a landlocked dummy' (Valencia Caicedo 2019, 519). Finally, μ captures the state-fixed effects.

Valencia Caicedo (2019) estimates Equation 1 using OLS with Conley (spatial) standard errors. We use OLS with unadjusted errors. While Valencia Caicedo (2019) is correct to use adjusted errors, they should not have a substantial effect on the error distribution, and they should have no effect on the parameter estimate. The main findings of Valencia Caicedo (2019) are contained in 'Table II' (Valencia Caicedo 2019, 523), which is made available here as Figure 10 for ease of reference.

From Valencia Caicedo (2018) the data that underpin 'Table II' are contained in 'Tables/Literacy Argentina Brazil Paraguay.dta' and the code that analyze these data is contained in 'Figures/Tables.do,' specifically lines 33-61. The general pattern of Figure 10 is that for each country grouping first Equation 1 is estimated without geographic

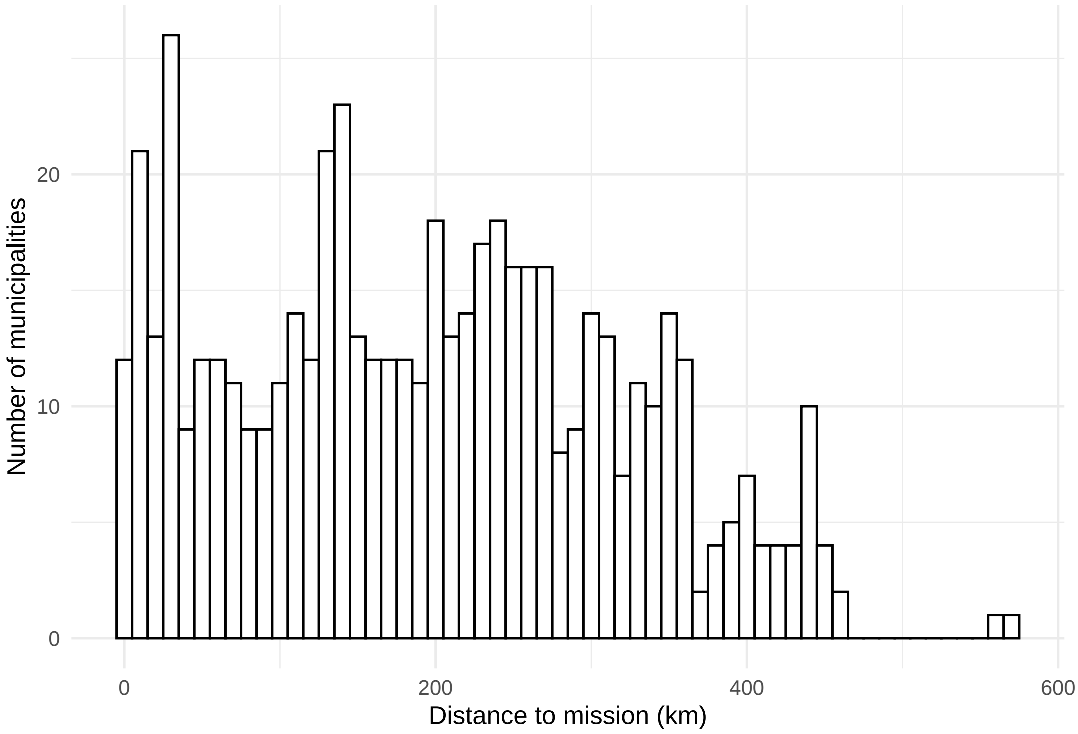

Figure 1: Distibution of average distance to closest mission for each municipality

controls and then with geographic controls. Geographic controls are described as including '...distance to the nearest coast, distance to the nearest river, altitude, ruggedness, temperature, area, rainfall, latitude, and longitude' (Valencia Caicedo 2019, 523). Valencia Caicedo (2018) uses Stata, whereas we use the open-source statistical programming language R (R Core Team 2020) and especially draw on the `tidyverse` (Wickham et al. 2019) package, as well as the `binsreg` (Cattaneo et al. 2019), `haven` (Wickham and Miller 2020), `kableExtra` (Zhu 2020), `knitr` (Xie 2021), and `stargazer` (Hlavac 2018) packages.

3 Results

3.1 Figure II

We are primarily interested in the relationship between the distance to the nearest mission and various dependent variables. In Figure 1 we show the distribution of distance to the nearest mission for various municipalities. There is a fairly even spread of municipalities through to around 300-350 km, after which it declines. There are only two municipalities that are further than 500km away from the nearest mission and both have high rates of literacy, and so if anything, including them would make the positive relationship even more so, but we will remove them as outliers.

'Figure II' from Valencia Caicedo (2019) (made available here with slight re-arrangements

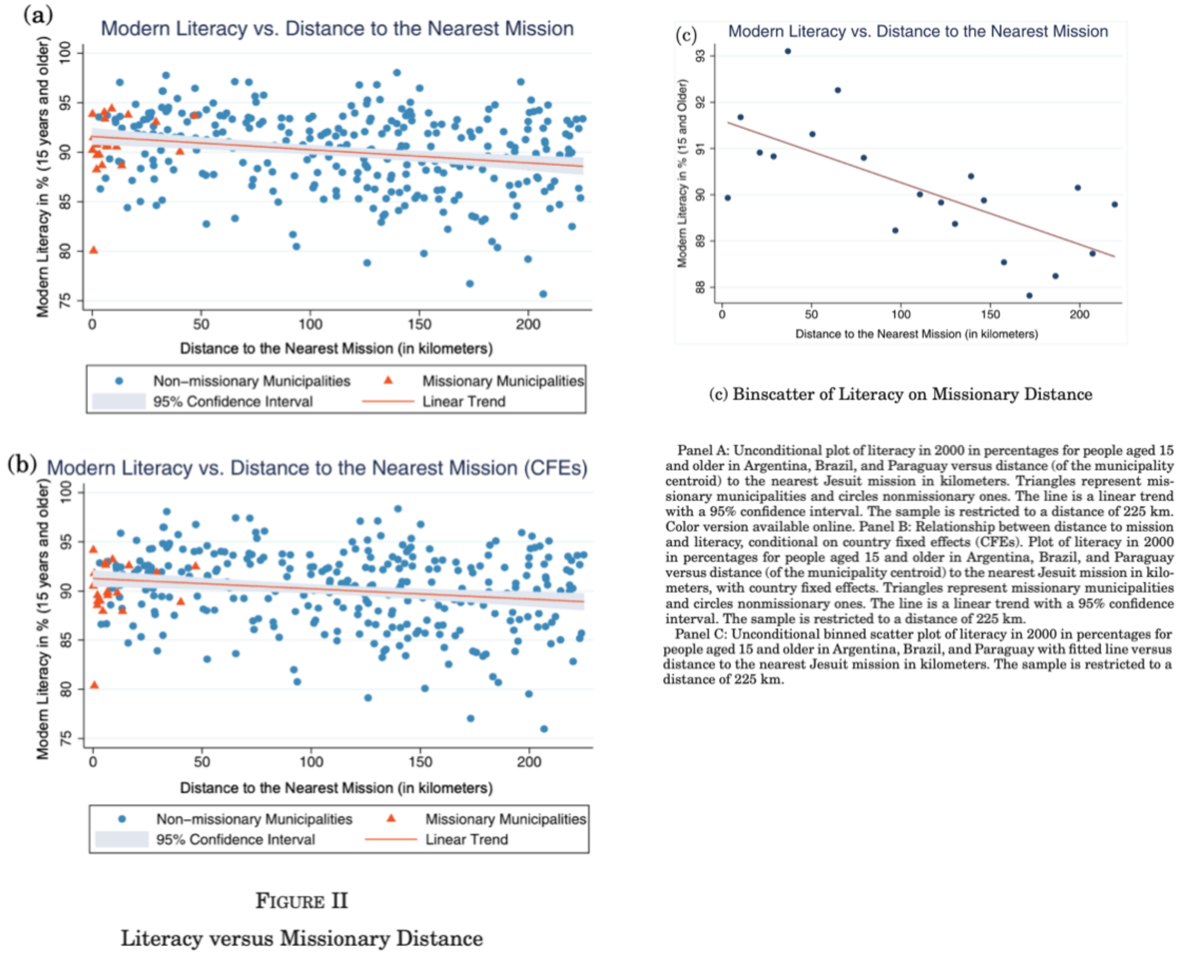

Figure 2: ‘Figure II’ from Valencia Caicedo (2019)

for presentation purposes as Figure 2 for ease of reference) shows a negative relationship between literacy and missionary distance for Argentina, Brazil, and Paraguay.

In Figure 3 we illustrate the relationship between the literacy rate in the years 2000-2002—as a percentage of people aged 15 and older in Brazil, and 10 and older in Argentina and Paraguay—compared with distance (of the municipality centroid) to the nearest Jesuit mission in kilometers.

The left panel of Figure 3 is a reproduction of ‘Panel (a)’ in ‘Figure II’ from Valencia Caicedo (2019). Here the sample is restricted to only municipalities that are within 225km of a mission, leaving 320 observations. This illustrates a negative relationship between these variables. In the right panel of Figure 3 we instead use almost the full sample of 578 observation and show that when a larger distance is considered the relationship is less clear. In that right panel we removed 29 missing values and two observations that appear to be highly influential leverage points—one at a distance of 556km and the other at a distance of 567km—noting that those two ‘outlier’ observations both have high rates of literacy. There are 547 observations in the right panel and the furthest observation has a distance of 461km. Figure 1 does not show a clear reason for the use of a cutoff

Figure 3: Relationship between literacy and distance to nearest mission

point at 225km.

In Figure 4 we illustrate the same relationship, but on a binned basis. The left panel of Figure 4 is analogous to that of 'Panel (c)' in 'Figure II' from Valencia Caicedo (2019) where the sample has been restricted to only municipalities that are within 225km of a mission. The right panel is the same almost-full sample, with the same observations removed as before. Again, the nature of the relationship turns on whether the sample is restricted or not.

To summarize, in Figure 3 and Figure 4 we reproduce 'Figure II' from Valencia Caicedo (2019). That figure illustrates the relationship between mission distance and modern literacy rates. Valencia Caicedo (2019) uses a cutoff point of 225km from the nearest mission to reduce the sample from 578 observations to 320 observations. There is no clear reason given for the cutoff point in Figure 1. In Figures 3 and 4 we show this cutoff influences the observed relationship between missionary distance and literacy rates. Moreover, Figure 5, which groups the distances, suggests the relationship between literacy and distance to the nearest missionary may differ at different distances. It may even be the case that literacy rates are higher for municipalities that are at least 200km away from the nearest the mission.

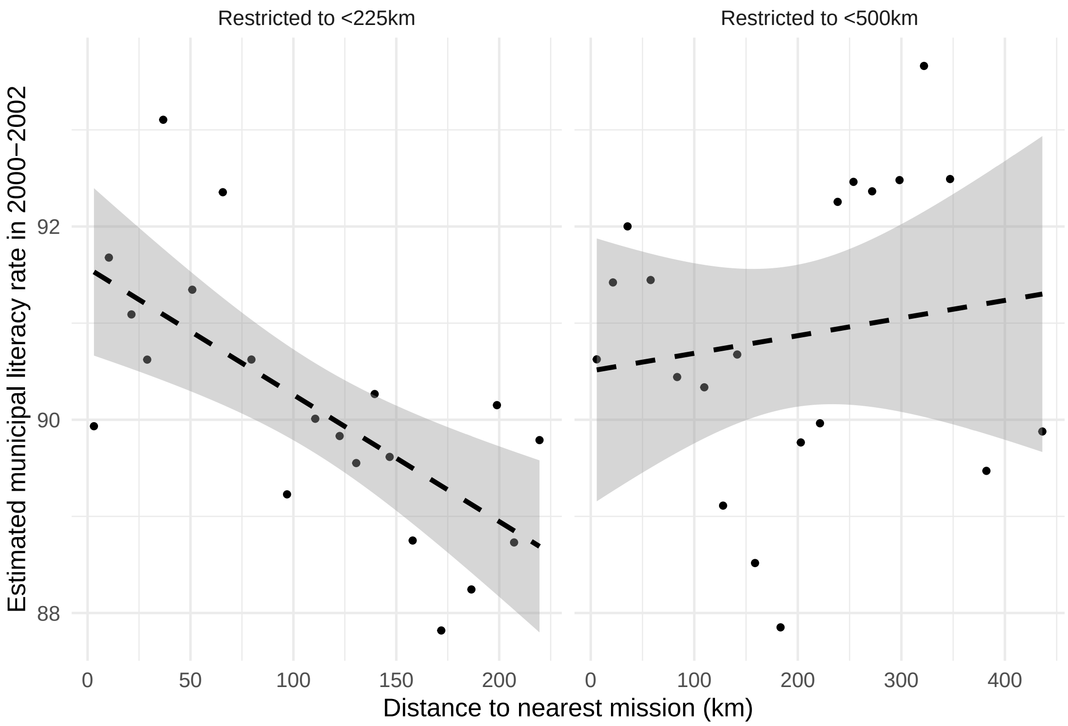

Figure 4: Binned relationship between literacy and distance to nearest mission

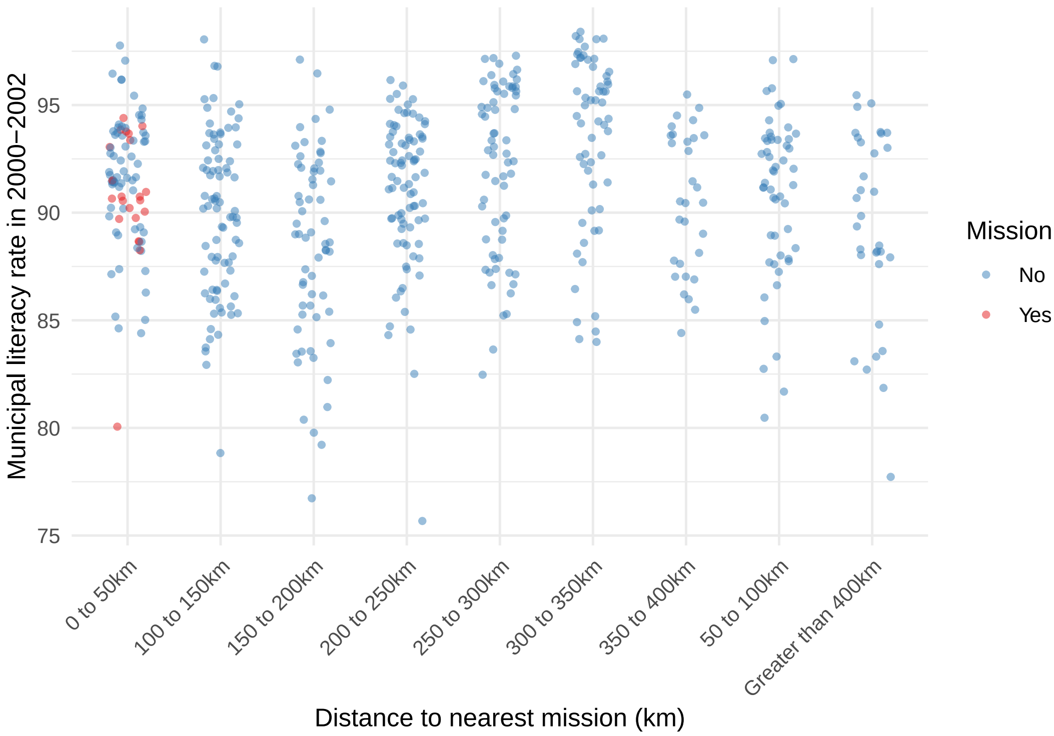

Figure 5: Relationship between literacy and binned distance to nearest mission

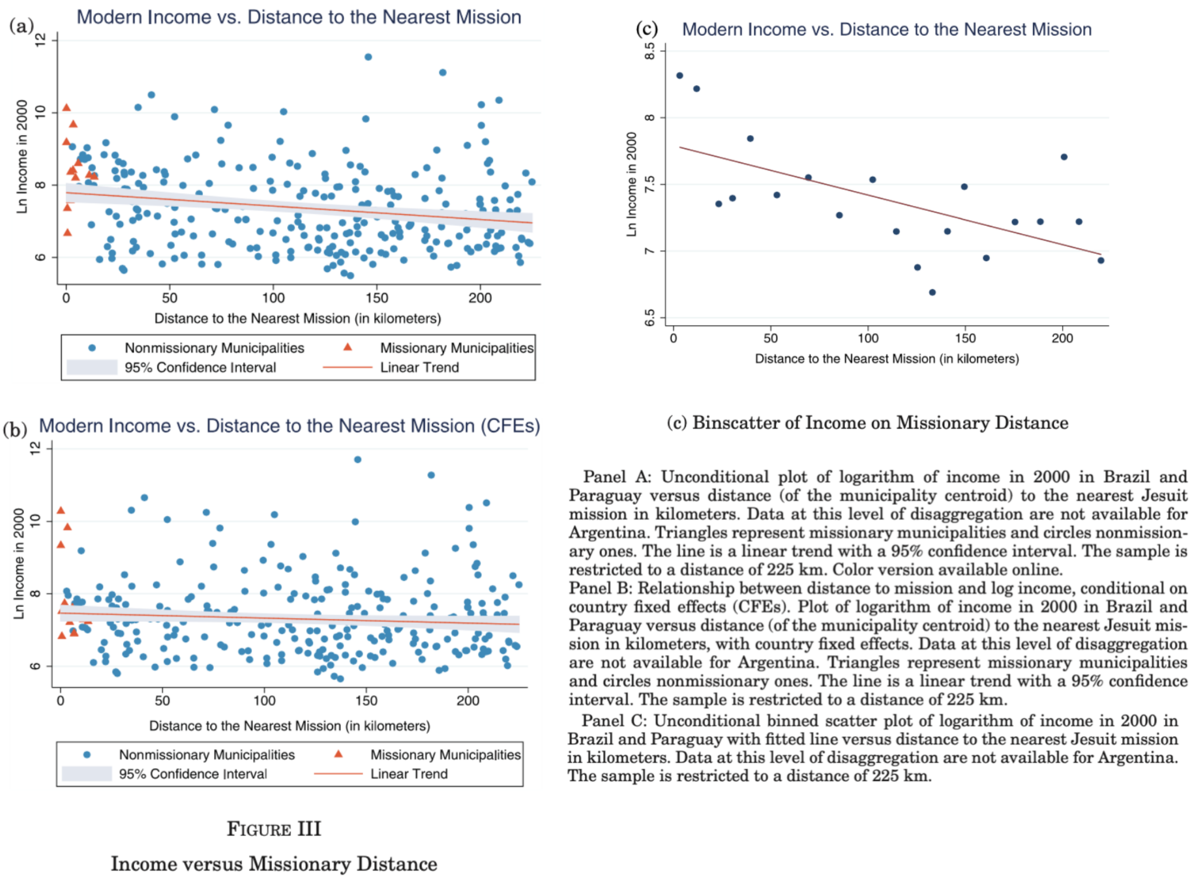

Figure 6: 'Figure III' from Valencia Caicedo (2019)

3.2 Figure III

'Figure III' from Valencia Caicedo (2019) (made available here with slight re-arrangements for presentation purposes as Figure 6 for ease of reference) shows a negative relationship between income and missionary distance for Brazil and Paraguay.

Valencia Caicedo (2018) provides the code and data underpinning this figure. Lines 27-40 from 'data/figures/Figures.do' identify the relevant data file as 'Figures/Figures IIII IV.dta' [sic]. The Brazilian incomes data in that file are not adjusted on a per capita basis, in contrast to the Brazilian incomes data in other datasets in Valencia Caicedo (2018) including those used in the regressions later such as 'Income Brazil Paraguay.dta.' However, the dataset also contains incomes data for Paraguay that do match the incomes in other datasets [Figure 7].

Our view is that the Paraguay data in the 'lnincome' variable in 'Income Brazil Paraguay.dta' is not on a per capita basis while the Brazil data in that same variable is. This variable is used as the dependent variable later in 'Table III.' Further, 'Figure III' combines data from Brazil and Paraguay. Given there are many differences between those countries that would influence modern incomes, our view is that it would be more appropriate to separate these.

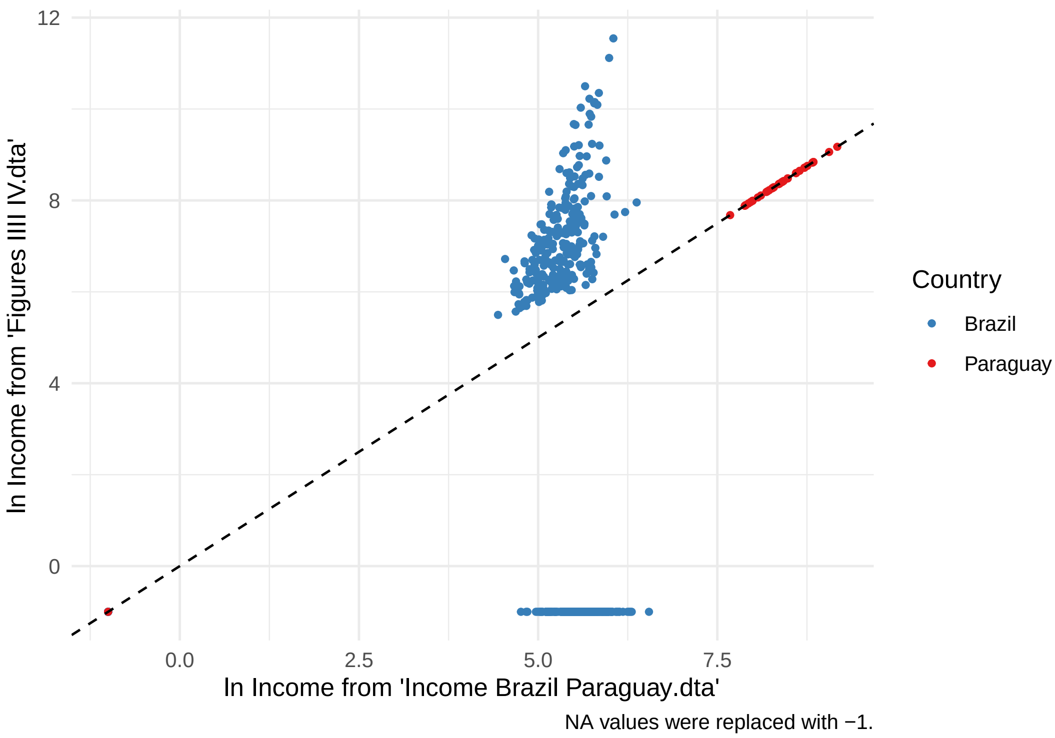

Figure 7: Comparison of ln income variables across two datasets

In Figure 8 we firstly examine the effect of separating Brazil from Paraguay on a ln income (not per capita) basis. The left panel reproduces 'Panel (a)' in 'Figure III' from Valencia Caicedo (2019). In the other two panels we separate Brazil and Paraguay. We show that the separated fits do not distinctly identify a negative relationship.

'Figure III' from Valencia Caicedo (2019) is focused on cross-municipal comparisons. Our view is that given this, using per capita income values is more appropriate, hence we need to use a different dataset 'Income Brazil Paraguay.dta,' which is also used later in the regressions. We are also interested in whether the 225km cutoff has an effect. This means that we use data from 'Income Brazil Paraguay.dta' because that has the full distance, not just the cut-off version. However, as discussed earlier, we believe that the Paraguay data in the lnincome variable in this dataset is not on per capita basis, while the Brazil data in that same variable is. For this reason, we show the plots with and without Paraguay, as well as with and without the 225 km cutoff. The result of making the above adjustments is presented in Figure 9. We find that there is a positive relationship between distance from a mission and ln income on a per capita basis for municipalities in Brazil regardless of the cut-off.

3.3 Table II

We are able to reproduce a detailed version of Figure 10 ('Table II' in Valencia Caicedo (2019)) with equivalent R code. There are slight differences in a few of the standard

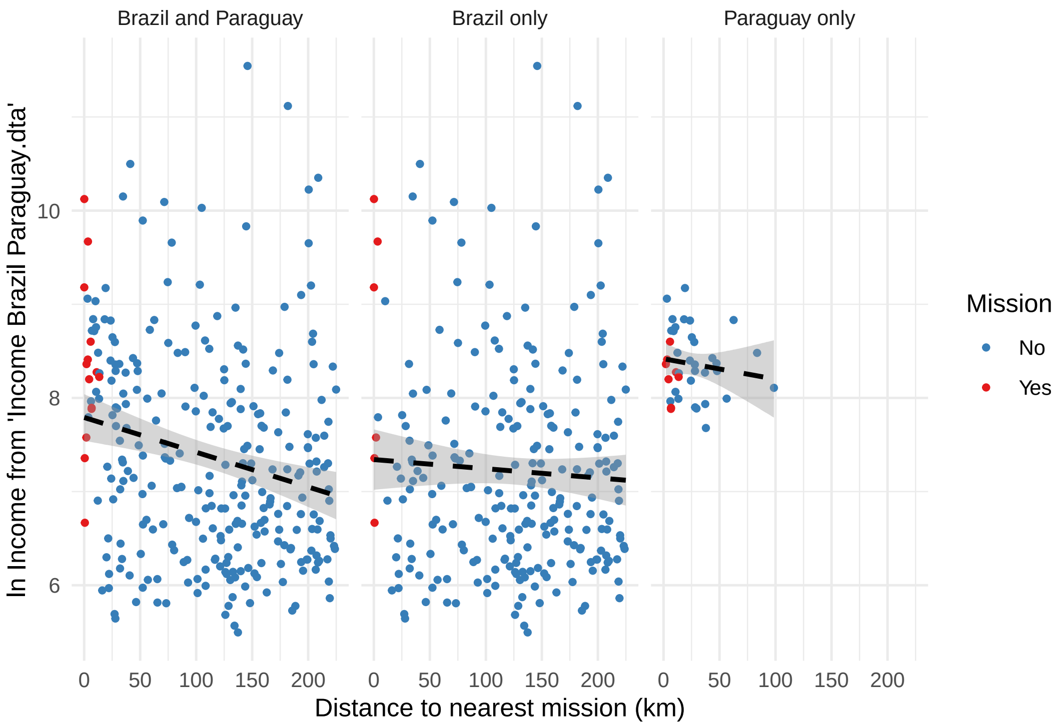

Figure 8: Comparison of ln income with distance to nearest mission

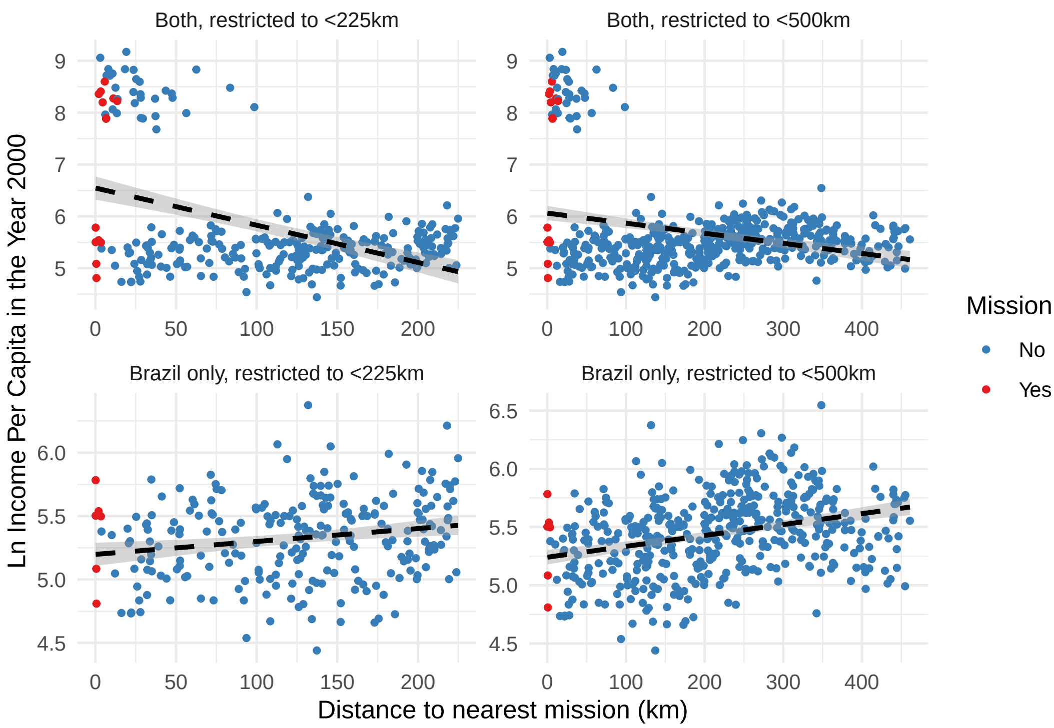

Figure 9: Comparison of ln income with distance to nearest mission

TABLE II

MISSIONARY EFFECT ON MODERN EDUCATION

	Illiteracy							
	Argentina, Brazil, and Paraguay		Brazil		Argentina		Paraguay	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mission distance	0.0105***	0.0112**	0.0200***	0.0313***	0.0157**	0.0669***	0.00451	0.0138
	(0.004)	(0.005)	(0.007)	(0.010)	(0.007)	(0.022)	(0.012)	(0.027)
	{0.004}	{0.005}	{0.007}	{0.010}	{0.008}	{0.023}	{0.016}	{0.026}
Geo controls	No	Yes	No	Yes	No	Yes	No	Yes
State fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	547	548	467	467	42	42	40	39
Within R^2	0.037	0.068	0.052	0.091	0.102	0.567	0.003	0.250
R^2	0.042	0.073	0.056	0.095	0.165	0.669	0.004	0.251

Notes. The table shows the coefficient of distance to the nearest Jesuit mission in kilometers (equation (1)). The dependent variable is illiteracy for people aged 15 years and older in 2000 in percentages for Argentina, Brazil, and Paraguay. Geographic controls include distance to the nearest coast, distance to the nearest river, altitude, ruggedness, temperature, area, rainfall, latitude, and longitude. Mesoregion fixed effects are included for Brazil. Please refer to Section I of the Online Appendix for units and additional details of these variables. Estimation is by OLS with state fixed effects. Robust standard errors are in parentheses and Conley standard errors are in curly brackets *** $p < .01$, ** $p < .05$.

Figure 10: 'Table II' from Valencia Caicedo (2019).

errors and p-values as we do not use Conley (spatial) standard errors; however, they do not affect the regression coefficients (Table 1). Note that this dataset uses all of the observations, not the cut-off sample illustrated earlier.

Although we are able to replicate Figure 10, we find a difference between the specifications for each country and how they are described in Figure 10. The particular issue is to do with geographic controls. From Valencia Caicedo (2019, 523) geographic controls include 'distance to the nearest coast, distance to the nearest river, altitude, ruggedness, temperature, area, rainfall, *latitude, and longitude*' [emphasis added].

In contrast we find that the code from Valencia Caicedo (2018) includes latitude and longitude for the regressions that are not meant to contain geographic controls for the specifications: (1) 'Argentina, Brazil, and Paraguay,' (3) 'Brazil,' and (5) 'Argentina'; but not for (7) 'Paraguay.' Moreover, latitude and longitude are not included for (8), in contrast to the footnote. When we remove latitude and longitude we find different estimates, in particular for the key effect of 'mission distance' (Table 2). And if we add latitude and longitude to (8) 'Paraguay,' we similarly find different estimates.

This revised version (Table 2) creates a table as described in Valencia Caicedo (2019), as opposed to what is implemented in the code that creates 'Table II.' Regressions (1) and (3), and Regression (8), in Table 2 now respectively include and exclude the geographic control variables longitude and latitude, in line with the description in Valencia Caicedo (2019). The effect on the coefficient of missionary distance is to change the sign in Regression (1), make it insignificant in Regression (3) and both change the sign and make it significant in Regression (8).

Table 1: Reproduction of an expanded version of Table II from Valencia Caicedo (2019) based on the code in Valencia Caicedo (2018).

	Dependent variable:							
	Illiteracy							
	Argentina, Brazil, and Paraguay		Brazil		Argentina		Paraguay	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mission dist.	0.0105***	0.0112**	0.0200***	0.0313***	0.0157*	0.0669***	0.0045	0.0138
	(0.0039)	(0.0046)	(0.0056)	(0.0077)	(0.0081)	(0.0232)	(0.0161)	(0.0264)
Latitude	0.5561**	0.0698	0.3592	3.2044**	0.0837	−9.3382**		
	(0.2509)	(0.7785)	(0.4028)	(1.4712)	(0.6277)	(3.6265)		
Longitude	−1.1083***	−1.0069	−1.7164***	−5.0435***	1.0950	7.1855***		
	(0.2694)	(0.6861)	(0.3688)	(1.5149)	(0.7782)	(2.4357)		
Corrientes	−5.3408***	−6.0319***			3.7714*	−3.0430		
	(1.5213)	(1.7922)			(2.2079)	(3.3376)		
Itapua	−3.1871***	−2.4093**					−0.2311	0.8290
	(0.9602)	(1.0567)					(0.8029)	(2.2116)
Misiones P	−4.3237***	−4.7341***						
	(1.5415)	(1.7395)						
Misiones A	−3.2794***	−2.2992*						
	(1.0817)	(1.2171)						
Coast		0.2086		−3.8518**		1.8894		0.8264
		(0.9886)		(1.8401)		(3.4242)		(4.2468)
River		1.4704**		1.6323**		9.7954***		0.9834
		(0.7283)		(0.7876)		(2.8368)		(5.3398)
Altitude		0.0057		0.0054		0.0654***		0.0160
		(0.0036)		(0.0045)		(0.0130)		(0.0151)
Ruggedness		−0.00000		−0.00000		−0.00005**		0.0001
		(0.00000)		(0.00000)		(0.00002)		(0.00004)
Temperature		0.0587		0.0550		0.9675***		0.3598*
		(0.0770)		(0.0974)		(0.2369)		(0.2102)
Rainfall		−0.0026		−0.0016		−0.0171**		0.0001
		(0.0021)		(0.0025)		(0.0081)		(0.0053)
Area		0.0001		−0.0004		−0.0001		0.0004
		(0.0002)		(0.0003)		(0.0003)		(0.0007)
Mesoregion			−0.4183*	−0.2100				
			(0.2165)	(0.2522)				
Constant	−35.3285***	−53.7412	1,724.6680*	731.8046	69.2628*	−41.0579	8.6735***	−80.7225*
	(12.0213)	(35.1185)	(919.8876)	(1,075.0750)	(39.0155)	(67.1521)	(0.7354)	(42.3090)
Observations	549	548	467	467	42	42	40	39
R^2	0.0418	0.0730	0.0562	0.0951	0.1651	0.6689	0.0036	0.2513
Adjusted R^2	0.0294	0.0486	0.0481	0.0732	0.0749	0.5475	−0.0503	0.0190
Note:	*p<0.1; **p<0.05; ***p<0.01							

Table 2: A revised version of Table II from Valencia Caicedo (2019) in line with the description.

	Dependent variable:							
	Illiteracy							
	Argentina, Brazil, and Paraguay		Brazil		Argentina		Paraguay	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mission dist.	−0.0031*	0.0112**	−0.0034	0.0313***	0.0097	0.0669***	0.0045	−0.0762**
	(0.0016)	(0.0046)	(0.0022)	(0.0077)	(0.0069)	(0.0232)	(0.0161)	(0.0340)
Latitude		0.0698		3.2044**		−9.3382**		−9.1836*
		(0.7785)		(1.4712)		(3.6265)		(5.1300)
Longitude		−1.0069		−5.0435***		7.1855***		24.3241***
		(0.6861)		(1.5149)		(2.4357)		(7.4086)
Corrientes	−0.0637	−6.0319***			0.7101	−3.0430		
	(0.8298)	(1.7922)			(1.0969)	(3.3376)		
Itapua	−1.1783	−2.4093**					−0.2311	2.6330
	(0.8106)	(1.0567)					(0.8029)	(1.9486)
Misiones P	−1.0258	−4.7341***						
	(1.3171)	(1.7395)						
Misiones A	−1.8572*	−2.2992*						
	(1.0192)	(1.2171)						
Coast		0.2086		−3.8518**		1.8894		26.1253***
		(0.9886)		(1.8401)		(3.4242)		(8.5459)
River		1.4704**		1.6323**		9.7954***		−6.2945
		(0.7283)		(0.7876)		(2.8368)		(5.3667)
Altitude		0.0057		0.0054		0.0654***		0.0311*
		(0.0036)		(0.0045)		(0.0130)		(0.0159)
Ruggedness		−0.00000		−0.00000		−0.00005**		0.00004
		(0.00000)		(0.00000)		(0.00002)		(0.00004)
Temperature		0.0587		0.0550		0.9675***		0.9127***
		(0.0770)		(0.0974)		(0.2369)		(0.2390)
Rainfall		−0.0026		−0.0016		−0.0171**		−0.0098*
		(0.0021)		(0.0025)		(0.0081)		(0.0057)
Area		0.0001		−0.0004		−0.0001		0.0025**
		(0.0002)		(0.0003)		(0.0003)		(0.0010)
Mesoregion			0.0001	−0.2100				
			(0.1318)	(0.2522)				
Constant	9.8306***	−53.7412	9.5721	731.8046	7.2233***	−41.0579	8.6735***	786.6667***
	(0.3886)	(35.1185)	(566.6376)	(1,075.0750)	(0.8222)	(67.1521)	(0.7354)	(243.1839)
Observations	549	548	467	467	42	42	40	39
R^2	0.0117	0.0730	0.0091	0.0951	0.1077	0.6689	0.0036	0.4973
Adjusted R^2	0.0026	0.0486	0.0048	0.0732	0.0619	0.5475	−0.0503	0.2925
Note:	*p<0.1; **p<0.05; ***p<0.01							

TABLE III

Missionary Effect on Development Proxies in Brazil, Argentina, and Paraguay

	Median years of schooling Brazil		Ln income Brazil and Paraguay		Individual poverty index Argentina and Paraguay	
	(1)	(2)	(3)	(4)	(5)	(6)
Mission distance	−0.00247**	−0.00679***	−0.00166***	−0.00204***	0.0409***	0.0938**
	(0.001)	(0.002)	(0.000)	(0.001)	(0.014)	(0.043)
	{0.001}	{0.002}	{0.000}	{0.001}	{0.018}	{0.046}
Geo controls	No	Yes	No	Yes	No	Yes
State fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	427	427	506	506	82	81
Within R^2	0.013	0.142	0.029	0.036	0.035	0.064
R^2	0.042	0.171	0.869	0.876	0.704	0.733

Notes. The table shows the coefficient of distance to the nearest Jesuit missions in kilometers (equation (1)). The dependent variables are median years of schooling in Brazil in columns (1) and (2), the logarithm of income per capita in 2000 in Brazil and Paraguay in columns (3) and (4), and the Unsatisfied Basic Needs (UBN) poverty index in Argentina and Paraguay at the individual level in columns (5) and (6). Median years of schooling is only available for Brazil at the municipal level and income is not available for Argentina. Mesoregion fixed effects are included for Brazil. Geographic controls include distance to the nearest coast, distance to the nearest river, altitude, ruggedness, temperature, area, rainfall, latitude, and longitude. Please refer to Section I of the Online Appendix for units and additional details of these variables. Estimation is by OLS with state fixed effects. Robust standard errors are in parentheses and Conley standard errors are in curly brackets *** $p < .01$, ** $p < .05$.

Figure 11: 'Table III' from Valencia Caicedo (2019).

3.4 Table III

In a similar manner to our reproduction of 'Table II' from Valencia Caicedo (2019) we are able to reproduce a detailed version of 'Table III' from Valencia Caicedo (2019) (included here for ease of reference as Figure 11)) with equivalent R code (Table 3). Again, there are slight differences in a few of the standard errors and p-values that are unimportant for our purposes.

We find that in contrast to how they are described in Valencia Caicedo (2019), regressions (1), (3) and (5) from 'Table III' (Valencia Caicedo 2019, 525) contain latitude and longitude in the estimating code (Valencia Caicedo 2018). The inclusion of latitude and longitude is necessary to obtain the headline results. Without the inclusion of these variables, the estimate of the effect of 'Mission distance' is different in many cases (Table 4). As with our earlier discussion about 'Table II,' this revised version of 'Table III' (Table 4) includes and excludes the latitude and longitude geographic controls in line with the description in Valencia Caicedo (2019). In Regression (1) it now becomes insignificant from zero, in Regression (3) it becomes significantly positive, in contrast to significantly negative, while in Regression (5) it remains positive and significant.

We have a few further points to raise with regard to 'Table III' from Valencia Caicedo (2019). Firstly, the code that underpins 'Table III' from Valencia Caicedo (2019) specifies these results require first filtering on rainfall to remove rows with missing values for rainfall. This is the case even for estimations that do not contain rainfall as a variable. Secondly, it requires considering additional variables to those mentioned in Valencia Caicedo (2019), such as 'slope' and a 'landlocked' binary variable. Thirdly, Valencia Caicedo (2019, 525) says that '[e]stimation is by OLS with state fixed effects.' However, again the code specifies that the 'mesoregion' variable was only used for Regressions (1)

Table 3: Reproduction of an expanded version of Table III from Valencia Caicedo (2019) based on the code in Valencia Caicedo (2018).

	Dependent variable:					
	Median years of schooling Brazil		Ln(income) Brazil and Paraguay		Individual poverty index Argentina and Paraguay	
	(1)	(2)	(3)	(4)	(5)	(6)
Mission dist.	−0.00247**	−0.00679***	−0.00202***	−0.00230***	0.04093**	0.09380*
	(0.00120)	(0.00166)	(0.00040)	(0.00053)	(0.01761)	(0.04944)
Latitude	−0.13909*	−0.48608*	−0.14709***	−0.07255	1.69757	−7.29712
	(0.08205)	(0.27515)	(0.02300)	(0.10439)	(1.45585)	(7.30583)
Longitude	0.17617**	0.99646***	0.20499***	0.16102	−0.34532	6.46119
	(0.07946)	(0.31176)	(0.02669)	(0.10966)	(0.99825)	(5.19658)
Mesoregion	0.06273	0.00535				
	(0.04504)	(0.05078)				
Coast		0.55119		−0.11587		2.85982
		(0.34781)		(0.13361)		(7.56710)
River		−0.20741		−0.09596		12.79641**
		(0.15752)		(0.05969)		(6.26056)
Altitude		−0.00035		−0.00007		0.05059*
		(0.00030)		(0.00029)		(0.02930)
Ruggedness		0.0000005		0.000001**		
		(0.000001)		(0.0000004)		
Temperature		0.00588		−0.00201		0.74223
		(0.00824)		(0.00619)		(0.49309)
Rainfall		−0.01278***		0.00020		−0.01286
		(0.00446)		(0.00017)		(0.01241)
Landlocked				0.02305		
				(0.08138)		
Area		0.00039***		0.00006***		0.00033
		(0.00006)		(0.00002)		(0.00059)
Slope		−0.00009**		−0.00002*		
		(0.00004)		(0.00001)		
Paraguay			3.53844***	3.63177***	23.95860***	26.31197***
			(0.07963)	(0.08692)	(2.44879)	(3.00289)
Constant	−259.00060	21.50930	12.40747***	12.58628***	55.97870	27.97562
	(191.21140)	(214.78030)	(1.12783)	(3.72727)	(42.48549)	(132.27520)
Observations	427	427	504	504	82	81
R^2	0.04190	0.17078	0.87182	0.87828	0.70362	0.73257
Adjusted R^2	0.03282	0.14675	0.87079	0.87505	0.68823	0.69436
Note:	*p<0.1; **p<0.05; ***p<0.01					

and (2), while a 'Paraguay' binary variable was used in Regression (3) - (6). Finally, we note that the data file 'Tables/Median Years of Schooling Brazil.dta' contains two variables for each of rainfall, temperature, and altitude. Figure 12 does not clearly identify a straight-forward transformation, and further, there are cases where the variable that is used contains NAs where the variable that is not used does not. It may be that the estimation results are sensitive to the choice of geographic variables chosen, but we did not test this sensitivity and instead followed Valencia Caicedo (2018).

Table 4: A revised version of Table III from Valencia Caicedo (2019) that follows the description.

	Dependent variable:					
	Median years of schooling Brazil		Ln(income) Brazil and Paraguay		Individual poverty index Argentina and Paraguay	
	(1)	(2)	(3)	(4)	(5)	(6)
Mission dist.	0.00006	−0.00679***	0.00093***	−0.00230***	0.03908***	0.09380*
	(0.00046)	(0.00166)	(0.00013)	(0.00053)	(0.01447)	(0.04944)
Latitude		−0.48608*		−0.07255		−7.29712
		(0.27515)		(0.10439)		(7.30583)
Longitude		0.99646***		0.16102		6.46119
		(0.31176)		(0.10966)		(5.19658)
Mesoregion	0.06806**	0.00535				
	(0.02661)	(0.05078)				
Coast		0.55119		−0.11587		2.85982
		(0.34781)		(0.13361)		(7.56710)
River		−0.20741		−0.09596		12.79641**
		(0.15752)		(0.05969)		(6.26056)
Altitude		−0.00035		−0.00007		0.05059*
		(0.00030)		(0.00029)		(0.02930)
Ruggedness		0.0000005		0.000001**		
		(0.000001)		(0.0000004)		
Temperature		0.00588		−0.00201		0.74223
		(0.00824)		(0.00619)		(0.49309)
Rainfall		−0.01278***		0.00020		−0.01286
		(0.00446)		(0.00017)		(0.01241)
Landlocked				0.02305		
				(0.08138)		
Area		0.00039***		0.00006***		0.00033
		(0.00006)		(0.00002)		(0.00059)
Slope		−0.00009**		−0.00002*		
		(0.00004)		(0.00001)		
Paraguay			3.09762***	3.63177***	25.33101***	26.31197***
			(0.05951)	(0.08692)	(2.06538)	(3.00289)
Constant	−287.71580**	21.50930	5.24284***	12.58628***	28.33134***	27.97562
	(114.43900)	(214.78030)	(0.03209)	(3.72727)	(1.96495)	(132.27520)
Observations	427	427	504	504	82	81
R^2	0.02888	0.17078	0.85603	0.87828	0.69729	0.73257
Adjusted R^2	0.02430	0.14675	0.85546	0.87505	0.68963	0.69436
Note:	*p<0.1; **p<0.05; ***p<0.01					

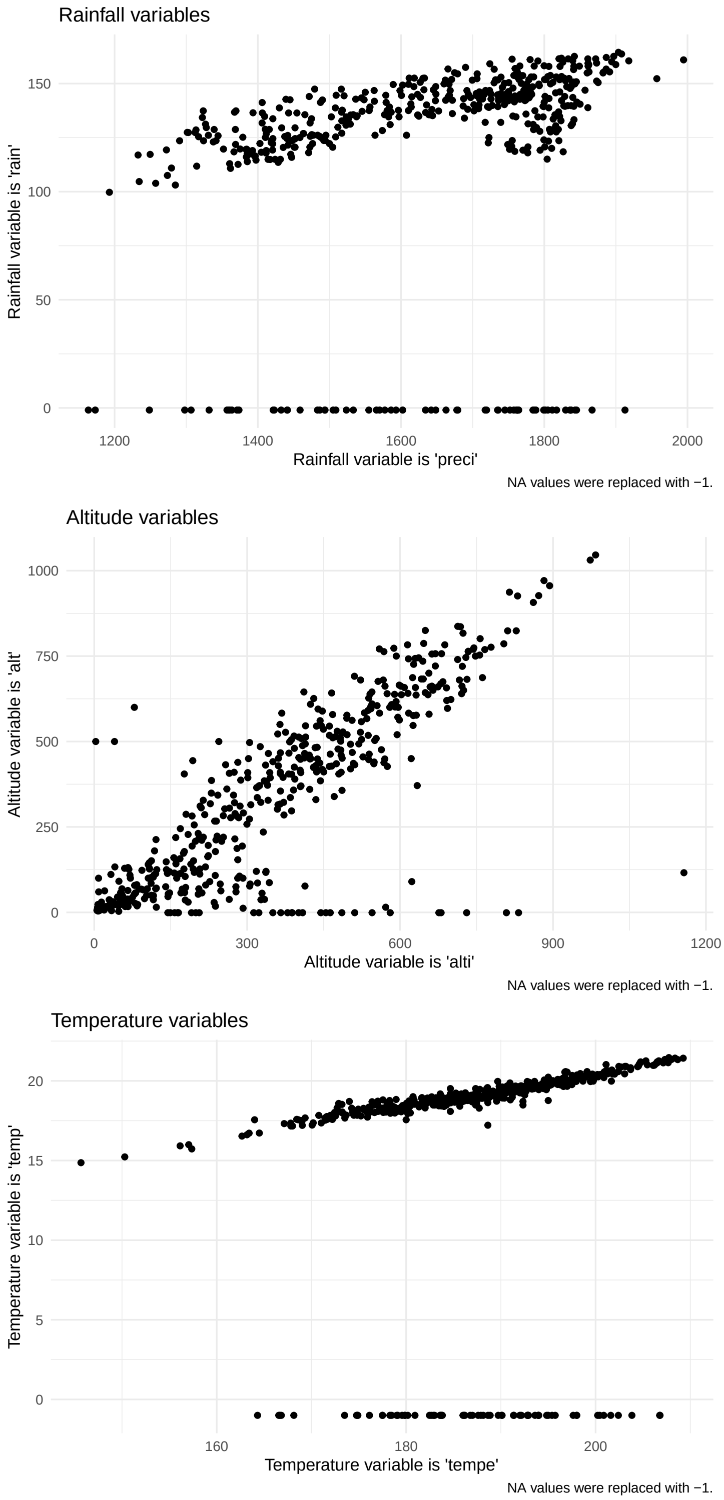

Figure 12: Comparison of different variables for rainfall, temperature, and altitude in 'Median Years of Schooling Brazil.dta'

3.5 Region fixed effects

The use of '[m]esoregion fixed effects' underpins the Brazil-specific regressions in 'Table II 'and 'Table III' of Valencia Caicedo (2019) (included here for ease of reference as Figure 11 and Figure 10) (Valencia Caicedo 2019, 525 and p. 523). In our reproductions of these tables as they appear in the text (Tables 1 and Table 3), and as they are described in the text (Tables 2 and 4), we follow Valencia Caicedo (2018) and Valencia Caicedo (2019). This means that the mesoregion variable enters as a numerical variable. However, as it is a fixed effect it should enter as a factor variable.

In Table 5 we focus on Regressions (3) and (4) of Table 10 (which is a reproduction of 'Table II' in Valencia Caicedo (2019)) and Regressions (3) and (4) of Table 11, (which is a reproduction of 'Table III' in Valencia Caicedo (2019)). These are the estimations that are focused on Brazil. We show that when the mesoregion variable enters as a factor rather than numerical variable, the magnitude of the key coefficient can change, and in one estimation, the statistical significance is lost.

4 Discussion

We find that Valencia Caicedo (2019) is reproducible but not replicable. That is to say, the code and data in Valencia Caicedo (2018) do reproduce the tables and figures in the paper. But we documented examples where the description of these tables and figures in the paper do not correspond to what occurs in the code. When we adjust the code to match what is described in the paper, we are not able to replicate the top-level findings of Valencia Caicedo (2019). While our results are specific to Valencia Caicedo (2019), our paper has many broader implications.

Our results suggest that having code and data alongside journal articles may represent the minimum requirement for publication. A useful summary of current practice in economics is provided by Christensen and Miguel (2018) however this issue extends to many social sciences, which are increasingly reliant on similar applied statistics methods. In response to requirements that authors submit code and data, increasingly researchers are requesting exemptions due to proprietary data. For instance, over the period 2005 to 2016, '...the proportion of papers using data that received exemptions from the data-sharing policy has risen rapidly, from roughly 10 percent to more than 40 percent over time' (Christensen and Miguel 2018, 937). This makes replication impossible. Possibly different requirements are needed for such papers if they are to be published, such as the need for their authors to arrange for access for a independent third-party as part of the peer-review process.

It may be that one reason the inconsistent inclusion of latitude and longitude in Valencia Caicedo (2019) was not identified was the formatting of the tables. For instance,

Table 5: Regressions (3) and (4) from Table II and Regression (1) and (2) from Table III from Valencia Caicedo (2019) with a comparison to the regions entering as factors not levels.

	Dependent variable:							
	Illiteracy				Median years of schooling			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mission dist.	0.01999***	0.01639**	0.03127***	0.02969***	−0.00247**	−0.00236	−0.00679***	−0.00569***
	(0.00560)	(0.00698)	(0.00766)	(0.00882)	(0.00120)	(0.00147)	(0.00166)	(0.00178)
Latitude	0.35916	0.40778	3.20443**	4.57468***	−0.13909*	−0.21966**	−0.48608*	−0.39074
	(0.40283)	(0.45106)	(1.47118)	(1.49880)	(0.08205)	(0.08965)	(0.27515)	(0.27702)
Longitude	−1.71637***	−1.02235*	−5.04351***	−5.69401***	0.17617**	0.14728	0.99646***	0.81132**
	(0.36876)	(0.55129)	(1.51491)	(1.54321)	(0.07946)	(0.11709)	(0.31176)	(0.32019)
Mesoregion	−0.41827*		−0.21002		0.06273		0.00535	
	(0.21648)		(0.25219)		(0.04504)		(0.05078)	
Meso 4302		−2.72006***		−2.54323***		0.25899		0.26255
		(0.84258)		(0.94371)		(0.16707)		(0.18008)
Meso 4303		−0.48291		−0.38298		0.04856		0.28901
		(1.04869)		(1.13722)		(0.21874)		(0.21594)
Meso 4304		−0.77105		0.19605		−0.18668		−0.27929
		(0.91778)		(1.06541)		(0.18902)		(0.21729)
Meso 4305		−3.02330***		−1.29000		0.29331		−0.01428
		(1.05072)		(1.32822)		(0.21018)		(0.25820)
Meso 4306		−1.72385		−3.42084*		0.67515**		0.60084*
		(1.65007)		(1.83335)		(0.34341)		(0.35703)
Meso 4307		−0.43677		0.32668		−0.38264		−0.41583
		(1.99507)		(2.08091)		(0.40785)		(0.40561)
Coast			−3.85176**	−4.97592***			0.55119	0.34158
			(1.84014)	(1.85436)			(0.34781)	(0.34829)
River			1.63231**	1.72297**			−0.20741	−0.31208*
			(0.78759)	(0.86914)			(0.15752)	(0.17060)
Altitude			0.00540	0.00063			−0.00035	−0.00019
			(0.00453)	(0.00497)			(0.00030)	(0.00031)
Ruggedness			−0.000005	−0.000003			0.0000005	0.0000002
			(0.000005)	(0.000005)			(0.000001)	(0.000001)
Temperature			0.05502	−0.06247			0.00588	0.01291
			(0.09740)	(0.10701)			(0.00824)	(0.00876)
Rainfall			−0.00157	0.00099			−0.01278***	−0.01318***
			(0.00246)	(0.00311)			(0.00446)	(0.00462)
Landlocked			−0.00042	−0.00016			0.00039***	0.00031***
			(0.00029)	(0.00032)			(0.00006)	(0.00007)
Area							−0.00009**	−0.00010**
							(0.00004)	(0.00004)
Slope	1,724.66800*	−35.27364	731.80460	−143.86890**	−259.00060	6.96724	21.50930	36.52843***
	(919.88760)	(28.03363)	(1,075.07500)	(59.94667)	(191.21140)	(6.05386)	(214.78030)	(10.52392)
Observations	467	467	467	467	427	427	427	427
R^2	0.05622	0.09402	0.09509	0.13481	0.04190	0.10282	0.17078	0.21588
Adjusted R^2	0.04805	0.07618	0.07321	0.10405	0.03282	0.08346	0.14675	0.18328
Note:	*p<0.1; **p<0.05; ***p<0.01							

the primary results are communicated in 'Table II' of Valencia Caicedo (2019) (included here for convenience as Figure 10) include only six rows, despite there being over ten variables in the regressions. If complete tables were included in the paper, then the inconsistent inclusion of latitude, longitude, other geographic variables, and state fixed effects, would have been evident. There is a clear need for full tables to be included in published work, especially in online appendices.

The replication exercise that we conducted was made possible due to the provision of code and data in Valencia Caicedo (2018). This reflects current practice in academic economics in terms of reproducibility. Economics has a robust tradition in motivating analysis through models and theory; however, on the applied side more can be done when these are translated into code. In particular we see the need for the adoption of model cards (Mitchell et al. 2019) and datasheets (Gebru et al. 2020), both of which are of increasing importance in computer science and applied statistics. Model cards are deliberately straight-forward one- or two-page documents that report aspects such as: model details; intended use; metrics; training data; ethical considerations; as well as caveats and recommendations (Mitchell et al. 2019). Datasheets accompany datasets and document 'motivation, composition, collection process, recommended uses,' among other aspects. Again, these should be deliberately straight-forward and relatively short documents.

The inclusion of latitude and longitude as control variables in regression models is common in economics. However, to the best of our knowledge, none of those prior papers suffer, in any obvious way, from the problem of multicollinearity between the key independent variables and latitude or longitude. For example, Nunn (2008) looks at the relationship between incomes per capita and the number of slaves exported per capita for a given country. Donaldson and Hornbeck (2016) focus on the relationship between the value of agricultural land and a measure of 'market access' which depends on factors such as availability of railroads. Yanagizawa-Drott (2014) investigates the relationship between violence and radio transmission of propaganda. Acemoglu, Hassan, and Robinson (2011) mainly looks at the impact of the Holocaust on outcomes such as population, and political and economic development. Becker and Woessmann (2009) investigates the relationship between literacy rate in a county with the percentage of Protestants living in the county. Acemoglu et al. (2011) looks at the relationship between urbanization and exposure to the French Revolution of 1789. In all these papers, the key independent variable of interest does not seem to suffer from the problem of multicollinearity with respect to latitude or longitude.

The issue with the inclusion of longitude and latitude in Valencia Caicedo (2019) is that the key independent variable of interest—missionary distance—depends in a very direct way on the longitude and latitude because of how it is constructed. While it may be true that multicollinearity among control variables usually does not pose a problem for applied econometricians, most would agree that multicollinearity involving a key variable of interest would be a cause for concern. For instance, in a passage largely downplaying the problem of multicollinearity, Wooldridge (2013, 96) acknowledges that: 'Although the problem of multicollinearity cannot be clearly defined, one thing is clear: everything else being equal, for estimating β_j, it is better to have less correlation between x_j and the other independent variables.' Greene (2012, 89) discusses some symptoms of multicollinearity one of which is that "coefficients may have the 'wrong' sign or implausible magnitudes."

In Table 6, we show the sensitivity of the coefficient of mission distance to the inclusion of longitude and latitude. If the problem of multicollinearity is accepted, our revised results—after removing the longitude and latitude variables and after the revisions we discussed in our paper—suggest that 'Table II' from Valencia Caicedo (2019) may have overstated the positive effects of mission proximity on illiteracy rates by a magnitude of at least four times for Brazil and approximately seven times for Argentina. Our results also suggest that while the sign of the main effects might be in the same direction as in Valencia Caicedo (2019), the statistical significance of the result disappears in most cases. Similarly, in Table 7, we find again that magnitudes may have been overstated and weaker levels of statistical significance.

To put this into context, the results in Valencia Caicedo (2019) suggest that moving 100 km closer to a mission increases median years of schooling by 0.67 years (8 months), whereas our best estimates suggest that this number is around 0.14 years (or less than 2 months). Similarly, we find that in our revised version of Valencia Caicedo (2019)'s 'Table III' taking into account our earlier discussion in Section 3, the magnitude and statistical significance of the relationship between mission distance and income or poverty is reduced in most cases if latitude and longitude are removed from the models (see Table 7 and compare with Figure 11).

Our results stress the importance of publishing regression diagnostics, such as tests of multi-collinearity, especially taking advantage of space provided by online appendices. It may be that in the same way that journals have requirements for certain types of formatting and citation styles, that there is a certain common set of quantitative results that all articles are expected to display. While there will always be appropriate exceptions, there should be an expectation that these requirements would be met.

Whether or not the types of issues that we identify with regard to Valencia Caicedo (2019) are systematic in the literature, there is a clear need for changes in the way that economics journals assess submissions. Vilhuber (2019) provides a helpful overview of the state of the art, and while some journals have implemented the idea of a data editor, our findings identify the need for journal editors to specifically ask journal reviewers to comment on the quality of the associated code and their ability to replicate findings. It is also clear that there are insufficient incentives for replication, and one aspect that may be helpful here would be for economics journals to set up the equivalent of a bug-bounty

Table 6: A revised version of Table II from Valencia Caicedo (2019) based on all our discussions above, highlighting the sensitivity of the key results to including latitude and longitude.

	Dependent variable:							
	Illiteracy							
	Argentina, Brazil, and Paraguay		Brazil		Argentina		Paraguay	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mission dist.	0.0087	0.0052	0.0297***	0.0068	0.0669***	0.0099	−0.0762**	0.0138
	(0.0053)	(0.0043)	(0.0088)	(0.0054)	(0.0232)	(0.0111)	(0.0340)	(0.0264)
Latitude	1.0455		4.5747***		−9.3382**		−9.1836*	
	(0.8731)		(1.4988)		(3.6265)		(5.1300)	
Longitude	−1.2346*		−5.6940***		7.1855***		24.3241***	
	(0.6838)		(1.5432)		(2.4357)		(7.4086)	
Corrientes	−7.0039***	−5.3103**			−3.0430	2.7154		
	(2.4959)	(2.2831)			(3.3376)	(2.7320)		
Itapua	−4.7657**	−4.5503**					2.6330	0.8290
	(2.2439)	(1.9643)					(1.9486)	(2.2116)
Misiones P	−5.7133**	−5.7234**						
	(2.6988)	(2.6170)						
Misiones A	−5.3668**	−5.0757***						
	(2.2812)	(1.9019)						
Meso 4301	−2.9024	−2.7038*						
	(1.8009)	(1.5253)						
Meso 4302	−5.2190***	−5.0913***	−2.5432***	−2.2000**				
	(1.6334)	(1.4088)	(0.9437)	(0.9443)				
Meso 4303	−2.5933	−2.3879*	−0.3830	0.2954				
	(1.5901)	(1.4242)	(1.1372)	(1.1115)				
Meso 4304	−2.3042	−2.3626*	0.1960	0.2998				
	(1.5541)	(1.2841)	(1.0654)	(1.0760)				
Meso 4305	−3.5223**	−3.4617***	−1.2900	−0.5801				
	(1.4151)	(1.1418)	(1.3282)	(1.3197)				
Meso 4306	−3.7854**	−3.5168**	−3.4208*	−1.4545				
	(1.7402)	(1.6758)	(1.8333)	(1.7477)				
Meso 4307			0.3267	2.4059				
			(2.0809)	(1.7333)				
Coast	−0.6651	0.9022*	−4.9759***	1.3023*	1.8894	−4.6216**	26.1253***	0.8264
	(1.0649)	(0.5456)	(1.8544)	(0.6841)	(3.4242)	(1.8227)	(8.5459)	(4.2468)
River	1.5641**	1.4185*	1.7230**	1.2738	9.7954***	10.9849***	−6.2945	0.9834
	(0.7937)	(0.7723)	(0.8691)	(0.8488)	(2.8368)	(3.0879)	(5.3667)	(5.3398)
Altitude	0.0031	0.0031	0.0006	0.0027	0.0654***	0.0477***	0.0311*	0.0160
	(0.0039)	(0.0032)	(0.0050)	(0.0038)	(0.0130)	(0.0121)	(0.0159)	(0.0151)
Ruggedness	−0.00000	−0.00000	−0.00000	−0.00000	−0.00005**	−0.00002	0.00004	0.0001
	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00002)	(0.00002)	(0.00004)	(0.00004)
Temperature	−0.0030	−0.0100	−0.0625	−0.0251	0.9675***	0.5850**	0.9127***	0.3598*
	(0.0819)	(0.0704)	(0.1070)	(0.0845)	(0.2369)	(0.2160)	(0.2390)	(0.2102)
Rainfall	−0.0005	−0.0023	0.0010	−0.0036	−0.0171**	−0.0034	−0.0098*	0.0001
	(0.0025)	(0.0023)	(0.0031)	(0.0028)	(0.0081)	(0.0064)	(0.0057)	(0.0053)
Area	0.0001	0.0002	−0.0002	0.0003	−0.0001	−0.0004*	0.0025**	0.0004
	(0.0002)	(0.0002)	(0.0003)	(0.0003)	(0.0003)	(0.0002)	(0.0010)	(0.0007)
Constant	−22.9710	13.3967	−143.8689**	14.6784	−41.0579	−91.9993**	786.6667***	−80.7225*
	(37.5845)	(13.4597)	(59.9467)	(17.0032)	(67.1521)	(36.5212)	(243.1839)	(42.3090)
Observations	548	548	467	467	42	42	39	39
R^2	0.1150	0.1095	0.1348	0.1086	0.6689	0.5727	0.4973	0.2513
Adjusted R^2	0.0814	0.0792	0.1041	0.0810	0.5475	0.4526	0.2925	0.0190

Note: *p<0.1; **p<0.05; ***p<0.01

Table 7: A revised version of Table III from Valencia Caicedo (2019) based on all our earlier discussions, highlighting the sensitivity of the key results to including latitude and longitude.

	Dependent variable:					
	Median years of schooling Brazil		Ln(income) Brazil		Individual poverty index Argentina and Paraguay	
	(1)	(2)	(3)	(4)	(5)	(6)
Mission dist.	−0.0057***	−0.0014	−0.0017**	−0.0010**	0.0830*	0.0253
	(0.0018)	(0.0010)	(0.0006)	(0.0004)	(0.0478)	(0.0248)
Latitude	−0.3907		−0.1893*		−7.4942	
	(0.2770)		(0.1109)		(7.7296)	
Longitude	0.8113**		0.2024*		7.7444	
	(0.3202)		(0.1146)		(4.9946)	
Meso 4302	0.2625	0.2277	0.1881***	0.1804***		
	(0.1801)	(0.1803)	(0.0703)	(0.0697)		
Meso 4303	0.2890	0.0222	−0.0399	−0.0523		
	(0.2159)	(0.1844)	(0.0850)	(0.0827)		
Meso 4304	−0.2793	−0.4282**	−0.1602*	−0.1607*		
	(0.2173)	(0.2132)	(0.0831)	(0.0831)		
Meso 4305	−0.0143	−0.0998	−0.1634	−0.1797*		
	(0.2582)	(0.2578)	(0.1027)	(0.1014)		
Meso 4306	0.6008*	0.0716	0.1301	0.0769		
	(0.3570)	(0.3075)	(0.1363)	(0.1297)		
Meso 4307	−0.4158	−1.1145***	−0.3192**	−0.3537***		
	(0.4056)	(0.3069)	(0.1541)	(0.1285)		
Coast	0.3416	−0.2983***	−0.0190	−0.2514***	2.9442	−1.7947
	(0.3483)	(0.1132)	(0.1382)	(0.0508)	(8.1400)	(3.6208)
River	−0.3121*	−0.3755**	−0.1170*	−0.0980	19.7917***	16.6395***
	(0.1706)	(0.1657)	(0.0658)	(0.0638)	(6.2646)	(5.9433)
Altitude	−0.0002	−0.0001	0.00005	−0.0001	0.0648**	0.0605**
	(0.0003)	(0.0003)	(0.0004)	(0.0003)	(0.0292)	(0.0245)
Ruggedness	0.00000	0.00000	0.00000*	0.00000*		
	(0.00000)	(0.00000)	(0.00000)	(0.00000)		
Temperature	0.0129	0.0112	0.0045	0.0013	0.8920*	0.5341
	(0.0088)	(0.0087)	(0.0080)	(0.0063)	(0.4668)	(0.4165)
Rainfall	−0.0132***	−0.0078*	−0.00000	0.0001	−0.0152	−0.0064
	(0.0046)	(0.0043)	(0.0002)	(0.0002)	(0.0129)	(0.0119)
Landlocked			0.0115	−0.0081		
			(0.0803)	(0.0796)		
Corrientes					7.9445	6.7917
					(6.6660)	(5.3929)
Itapua					29.0260***	26.7154***
					(2.9867)	(2.6451)
Misiones (P)					21.9496***	19.6440***
					(5.3551)	(4.8997)
Area	0.0003***	0.0002***	0.00003	0.00002	−0.00001	−0.00002
	(0.0001)	(0.0001)	(0.00002)	(0.00002)	(0.0007)	(0.0006)
Slope	−0.0001**	−0.0001***	−0.00002	−0.00002		
	(0.00004)	(0.00004)	(0.00001)	(0.00001)		
Constant	36.5284***	5.2519***	10.1854**	5.8258***	62.4557	−75.8118
	(10.5239)	(1.9396)	(4.4353)	(1.2547)	(139.0287)	(69.6889)
Observations	427	427	467	467	81	81
R^2	0.2159	0.1979	0.2857	0.2804	0.7696	0.7599
Adjusted R^2	0.1833	0.1687	0.2570	0.2548	0.7290	0.7256

Note: *p<0.1; **p<0.05; ***p<0.01

program. Technology companies use these to encourage the responsible disclosure of undesirable results from their code. As economics, and the social sciences more generally, themselves become more reliant on code, implementing such a system becomes of greater importance. This could easily be done at a department level, but we would see the adoption by journals or organizations such as NBER as likely to have the largest effect. In the case of technology companies, the rewards for finding bugs are generally monetary; however, in the case of their adoption in economics and social sciences, mere official credit by that journal or organization would likely initially be sufficient. For instance, in the same way that some journals acknowledge prolific referees by listing them, see for instance "The Quarterly Journal of Economics gratefully acknowledges the help of the people listed below who refereed four or more papers in 2020" (2021), a list of all the people that found 'bugs' could be published annually. Finally, while bug bounties would be appropriate in the case of minor replication issues, we see the clear need for journals to set-up procedures for third parties to follow in the event of a more substantial failure of replication. The establishment of clear responsibility for such cases is no longer something that can be left to chance.

Although our paper is critical of Valencia Caicedo (2019), we close with admiration for the way in which the author provided their code and data in Valencia Caicedo (2018) and appreciation for the professional way in which the author assisted with our work. If they had not done this then our paper would not have been possible, and we thank and acknowledge this.

References

Acemoglu, Daron, Davide Cantoni, Simon Johnson, and James A Robinson. 2011. "The Consequences of Radical Reform: The French Revolution." *American Economic Review* 101 (7): 3286–3307.

Acemoglu, Daron, Tarek A Hassan, and James A Robinson. 2011. "Social Structure and Development: A Legacy of the Holocaust in Russia." *The Quarterly Journal of Economics* 126 (2): 895–946.

Becker, Sascha O, and Ludger Woessmann. 2009. "Was Weber Wrong? A Human Capital Theory of Protestant Economic History." *The Quarterly Journal of Economics* 124 (2): 531–96.

Cattaneo, Matias D., Richard K. Crump, Max H. Farrell, and Yingjie Feng. 2019. *Binsreg: Binscatter Estimation and Inference*. https://CRAN.R-project.org/package=binsreg.

Christensen, Garret, and Edward Miguel. 2018. "Transparency, Reproducibility, and the Credibility of Economics Research." *Journal of Economic Literature* 56 (3): 920–80.

Donaldson, Dave, and Richard Hornbeck. 2016. "Railroads and American Economic

Growth: A 'Market Access' Approach." *The Quarterly Journal of Economics* 131 (2): 799–858.

Gebru, Timnit, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III au2, and Kate Crawford. 2020. "Datasheets for Datasets." http://arxiv.org/abs/1803.09010.

Greene, William H. 2012. *Econometric Analysis*. 7th ed. Pearson Education.

Hlavac, Marek. 2018. *Stargazer: Well-Formatted Regression and Summary Statistics Tables*. Bratislava, Slovakia: Central European Labour Studies Institute (CELSI). https://CRAN.R-project.org/package=stargazer.

Mitchell, Margaret, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. 2019. "Model Cards for Model Reporting." *Proceedings of the Conference on Fairness, Accountability, and Transparency*, January. https://doi.org/10.1145/3287560.3287596.

Nunn, Nathan. 2008. "The Long-Term Effects of Africa's Slave Trades." *The Quarterly Journal of Economics* 123 (1): 139–76.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

"The Quarterly Journal of Economics gratefully acknowledges the help of the people listed below who refereed four or more papers in 2020." 2021. *The Quarterly Journal of Economics*, January. https://doi.org/10.1093/qje/qjab001.

Valencia Caicedo, Felipe. 2018. "Replication Data for: '"The Mission: Human Capital Transmission, Economic Persistence, and Culture in South America'." *Harvard Dataverse*. https://doi.org/10.7910/DVN/ML1155.

———. 2019. "The Mission: Human Capital Transmission, Economic Persistence, and Culture in South America." *The Quarterly Journal of Economics* 134 (1): 507–56.

Vilhuber, Lars. 2019. "Report by the AEA Data Editor." *AEA Papers and Proceedings* 109 (May): 718–29. https://doi.org/10.1257/pandp.109.718.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, and Evan Miller. 2020. *Haven: Import and Export 'SPSS', 'Stata' and 'SAS' Files*. https://CRAN.R-project.org/package=haven.

Wooldridge, Jeffrey M. 2013. *Introductory Econometrics: A Modern Approach*. 7th ed. South-Western, Cengage Learning.

Xie, Yihui. 2021. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. https://yihui.org/knitr/.

Yanagizawa-Drott, David. 2014. "Propaganda and Conflict: Evidence from the Rwandan Genocide." *The Quarterly Journal of Economics* 129 (4): 1947–94.

Zhu, Hao. 2020. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. https://CRAN.R-project.org/package=kableExtra.

Acknowledgements

Tariq acknowledges financial support from the Canadian Statistical Sciences Institute (CANSSI) Ontario and Monica Alexander. Alexander is grateful for the advice of Monica Alexander and anonymous colleagues.