COMMENT: MALARIA ERADICATION IN THE AMERICAS

David Roodman[1]

Bleakley (2007) and Bleakley (2010) both find that large-scale campaigns in the 20th century to eradicate a parasitic disease—hookworm and malaria—were followed by income gains for those native to historically endemic areas. Roodman (2017) reanalyzes and questions Bleakley (2007), arguing that no historical discontinuities coincide with hookworm eradication in the American South. The present paper applies the same methods, pre-registered, to Bleakley (2010), and returns more supportive results. Malaria eradication efforts indeed appear to have been followed by anomalous income gains for natives of historically malarial areas of Brazil, Colombia, Mexico, and perhaps the United States too. (JEL I18, O15; keywords: malaria, public health and economic development; replication)

Two important contributions to the literature on the long-term economic impacts of public health interventions are Bleakley (2007) and Bleakley (2010). Both find that large-scale campaigns in the 20th century to eradicate a parasitic disease—hookworm and malaria, respectively—were followed by income gains for those native to historically endemic areas. The first is set in the United States, the second in the United States, Brazil, Colombia, and Mexico adult earnings rose for people from more-malarial regions, relative to less-endemic regions. Roodman (2017) replicates and reanalyzes Bleakley (2007), and ultimately questions its conclusion, arguing instead that no historical discontinuities clearly coincide with the hookworm eradication campaign.

The present paper brings the same set of techniques to Bleakley (2010). As a replication, it returns to primary sources to reconstruct all the variables for the U.S. impact assessment. For Brazil, Colombia, and Mexico, it likewise reconstructs the *outcome* variables, but not the treatment proxies or controls. As a reanalysis, the paper introduces (pre-registered) innovations:

[1] Senior Advisor, GiveWell, 182 Howard St., #208, San Francisco, CA, 94105 (e-mail: david.roodman@givewell.org). The author thanks Christian Smith for help with data collection and Zachary Tausanovitch for help with data collection and database construction. GiveWell its sister organization the Open Philanthropy Project have no financial interest in the topic of this paper that constitutes a conflict of interest.

improving the outcome measures by incorporating the later and denser samples of census microdata now available; and applying formal and graphically informed inference to time series patterns. The paper uncovers some coding errors in the original, but these do not appear to greatly affect results.

Where these methods tend to challenge Bleakley (2007) they tend to corroborate Bleakley (2010). As Bleakley (2010) predicts, adult earnings as a function of birth year rose with anomalous speed in historically malaria-burdened regions about when the first babies were born who would spend at least part of their childhoods in the post-eradication regimes. And convergence decelerated as the last of these babies were born—that is, as the transition from pre- to- post-eradication regime completed. The finding is perhaps less certain for the United States than for the three Latin countries studied. It also holds less clearly for human capital accumulation, as measured by literacy in adulthood, and years of schooling completed. That result somewhat contradicts Bleakley (2010), which perceives indications of significant impacts on literacy, if not schooling.

This paper speaks not only to the impact of public health intervention on economic development. It also offers lessons on how journals archive data and code. The data availability policy of *AEJ: Policy*, which published Bleakley (2010), requires authors to provide "the data, programs, and other details of the computations sufficient to permit replication."[2] Hoyt Bleakley appears to have complied with this policy as it has normally been implemented, providing data and code to the journal's website.[3] Yet, in two important respects, the paper's results are impossible to exactly replicate. The figures, which are no less important than tables for inference, cannot be precisely replicated, because the public code does not generate them. Lack of public code for figures appears to be the norm for the American Economic Association journals. Also, neither the primary data nor the code that transforms it into analysis data are included—as again appears to be the norm—so one cannot fully reconstruction the chain from primary sources to final conclusions.[4] In these ways, the archive falls short of its purpose of making research transparent and replicable.

[2] web.archive.org/web/20171101092538/https://www.aeaweb.org/journals/policies/data-availability-policy.
[3] See aeaweb.org/aej/app/data/2008-0126_data.zip.
[4] See also Glandon (2011).

Section 1 of this paper describes the Bleakley (2010) research designs. Section 2 explores some cross-cutting themes in the replication and reanalysis. Section 3 reports on the (partial) reconstructions of the data sets. Section 4 replicates and reanalyzes the time series results. Section 5 concludes.

1 Designs

The Bleakley (2010) specifications combine up to three sorts of variables:

- Cross-sectional variables, observed once per geographic unit—e.g., per Brazilian state or Colombian *municipio*. These include indicators of pre-eradication malaria mortality or malaria ecology (M), as well as controls.
- Variables built from census microdata, including measures of schooling, literacy, and income. All microdata comes from the Integrated Public Use Microdata Series (IPUMS; Ruggles et al. 2015; Minnesota Population Center 2017).
- A pure time series indicator for exposure to the eradication campaign (Exp). Only the panel regressions, described shortly, include Exp explicitly. In an approach akin to difference-in-differences, these regressions interact Exp with M to form the treatment proxy, while effectively controlling for Exp and M individually.

Of the two components of the treatment proxy, $Exp \times M$, the second is a marker for geography and therefore potentially for economic history. While *external* to the causal pathways from malaria eradication to the outcomes of interest, it is not very credibly *exogenous*. The other component, Exp, is more plausibly exogenous in the short-term than the long-term. That is, it is not an accident of history that these campaigns occurred in the 20th century rather than the 19th or 21st. More accidental perhaps is that they took place when they did, rather than a few years earlier or later. Rather as in an interrupted time series design, the results that can most compellingly demonstrate causality will derive from changes in the time dimension over a few years.

All the Bleakley (2010) estimators begin by averaging an outcome Y within census year–birth year–birth place cells, with the dimensions indexed by c, t, j; this gives a set of values $\bar{Y}_{ctj}$. These are then demeaned nationally, within each census year–birth year group, yielding $\tilde{Y}_{ctj}$. The $\tilde{Y}_{ctj}$

are then modeled in regressions. A disadvantage of this preprocessing is that the imprecision of the initial demeaning step is not factored into the standard errors from the main estimation step.

Bleakley (2010) first fits cross-sectional long-difference regressions, with the model

$$\Delta\tilde{Y}_j = M_j\beta + \mathbf{x}_j'\boldsymbol{\gamma} + \epsilon_{ij} \quad (1)$$

j indexes geographic units and β is the parameter of interest.[5] $\mathbf{x}$ is a set of controls. ϵ_{ij} is the mean-zero random error. $\Delta\tilde{Y}_j$ is the change in the average value of $\tilde{Y}_{ctj}$ for area j, from the "before" to the "after" period. The "after" period begins when the eradication campaigns are taken to have commenced—1920 in the United States, 1957 in Brazil, Colombia, and Mexico. The "before" period ends in 1890 in the United States, and in 1940 in the Latin countries. The latter cut-offs are chosen to assure that all children born in the "before" period would have reached adulthood by the campaign, and so would have experienced no campaign-induced reduction in childhood malaria exposure. Individuals born in the gap between the two periods do not figure in these regressions.

The long-difference regressions, reported in Bleakley (2010) Tables 1–3, show that most outcomes tested improved faster in places with high pre-eradication malaria burden. These relative rises constitute circumstantial evidence that eradication delivered substantial benefits. However, as Bleakley (2010, p. 13) points out, the regressions do not speak to the historical distinctiveness of the rises. Perhaps, for example, these trends began too early or continued too long for the malaria eradication campaigns to naturally explain them.[6]

The Bleakley (2010) panel regressions look more sharply at timing. To do so, they define the exposure variable Exp as the fraction of childhood spent in the post-eradication regime, as a function of birth year. As a pure time series variable, Exp takes the same value regardless of the historical malaria burden of one's birth place. According to the Bleakley (2010) text, childhood is taken to last 21 years. This makes Exp a piecewise-linear "step" function with a 21-year rise.

[5] These can also be viewed as two-period panel regressions in which Exp is a dummy for the second period, $M \times Exp$ is the treatment proxy, and M and Exp are effectively controlled for through dummy sets for place and year of birth.

[6] Bleakley (2010, p. 13) suggests that because they apply to data aggregated over time, the long-difference regressions have the advantage of avoiding high-frequency serial correlation. However, the Bleakley (2010) panel regressions also address serial correlation, by clustering standard errors by place of birth.

In the Latin countries, for example, Exp is 0 through 1936, then rises linearly until it reaches 1 in 1957, and then goes flat again.

The panel regressions fit

$$\tilde{Y}_{ctj} = \left(Exp_t \times M_j\right)\beta + \mathbf{x}'_{tj}\boldsymbol{\gamma} + \delta_c + \delta_t + \delta_j + \epsilon_{ctj} \quad (2)$$

β remains the parameter of interest. The δ_c, δ_t and δ_j are the indicated dummy sets, with the δ_t and δ_j obviating the inclusion of Exp_t and M_j as controls. The controls $\mathbf{x}_{tj}$ are not true panel variables, in the sense of being observed in primary sources in multiple times in multiple places. Rather, all are products of pure cross-sectional and pure time series variables. For example, the Bleakley (2010) "full controls" panel regressions include interactions between geographic control variables and Exp.[7]

Regressions based on (2) can be viewed as testing whether the step function Exp is a strong explanator for the temporal evolution of the spatial association between baseline malaria burden M and the outcome Y. The model will fit well if the association takes a low (potentially negative) value among cohorts born well before the campaign, begins to rise steadily among those born late enough to still be children during the campaign, and then plateaus again among people born after the campaign.

However, fitting the model can generate a false positive if such convergence begins well before or extends well after the dates implied by the construction of Exp—and is in fact caused by other forces. Regressions in such cases could estimate β as being statistically different from zero, and create the misleading impression that Exp is a good explanator for long-term trends. In the language of time series analysis, regressing one $I(1)$ variable on another could generate spurious results.

[7] Equation (2) elides one nonstandard complication in the fitting procedure. However, it is nearly immaterial for the cohort-by-cohort regressions of interest here. Before estimation, Y_{cjt} is demeaned within each census year–birth year cell. In other words, the interacted dummies δ_{ct} are partialled out of the left-side variable, but not the right-side ones. Failure to partial these effects out of the right-side variables could cause some of their explanatory power load misleadingly onto those variables in an OLS regressions, causing omitted-variable bias. However, in the context of (3),discussed next, there is no problem. There, δ_c is controlled for separately in each t-indexed birth cohort, which is equivalent to first partialling the $\delta_c \times \delta_t$ out of all other regressors. Partialling the δ_{ct} out of the left-side variable before estimation has no effect if the estimation itself effectively partials the dummies out of all other variables.

Bleakley (2010) takes several steps to rule out such possibilities. All the Bleakley (2010) regressions include measures of initial conditions in order to control for mean reversion. Some introduce state- or *municipio*-specific time trends, linear or quadratic. These measures suffice *if* the augmented models largely capture the ambient time trends. But in general, we do not know the functional form for major extraneous trends. And it is hard to judge how close the models come only by viewing tabulated estimates of β.

Bleakley (2010)'s graphical time series approach can give more insight into ambient trends. It runs a version of (2) for each (t-indexed) birth cohort:

$$\tilde{Y}_{ctj} = M_j\beta_t + \mathbf{x}_j'\boldsymbol{\gamma}_t + \delta_c + \epsilon_{cjt} \tag{3}$$

The regressions yield a series of coefficients, β_t, which measure the cross-sectional association between Y and baseline malaria burden. The β_t can be graphed for visual inspection of long-term trends. And they can be subject to formal inference. Indeed, in studying hookworm eradication, Bleakley (2007, Table VI) uses time series regressions to perform inference on whether Exp is a determinant of the β_t. In contrast, Bleakley (2010) discusses the evolution of the β_t only informally. I resurrect and revise the Bleakley (2007) approach and apply it to malaria eradication, just as Roodman (2017) does for Bleakley (2007).

This revised time series approach begins by fitting the models (2) and (3) directly to census microdata, as in most of the Bleakley (2007) hookworm study, rather than to nationally demeaned, cell-aggregated outcomes as in Bleakley (2010). This change brings three benefits. First, moving to microdata sidesteps the debatable choice in Bleakley (2010) to weight the cell aggregates $\tilde{Y}_{ctj}$ by the square root of cell size instead of cell size.[8] Instead, one weights individuals by the IPUMS-provided sampling weights. Second, the move allows one to incorporate individual-level demographic controls. As the regressions are carried out here, this amounts to including a dummy for sex in the expanded-sample regressions, since they add

[8] Weighting by the square root of cell size is evidently meant to improve efficiency by reducing heteroskedasticity. But theory favors weighting simply by cell size. The variances of the cell-average values $\overline{Y}_{ctj}$ are inversely proportional to cell size. Assuming that this inverse law carries over to the $\tilde{Y}_{ctj}$ and ϵ_{ctj}, the heteroskedasticity is reversed by weighting by inverse variance, i.e., cell size. In symbols, if $\mathbf{Y}$ is a column vector holding the $\tilde{Y}_{ctj}$, $\mathbf{X}$ holds the right-side variables, and $\mathbf{W}$ is a diagonal matrix whose entries are cell sizes, then Aitken's efficient generalized least squares estimator is $(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{Y}$. The Bleakley (2010) code performs $\left(\mathbf{X}'\mathbf{W}^{1/2}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{W}^{1/2}\mathbf{Y}$.

women (see section 2.1); and likewise for race in the expanded U.S. regressions, which also add blacks. Bleakley (2007) uses both dummies too. (Within birth cohorts, controlling for fixed census round effects effectively control for age already.) Third, as done here, fitting to microdata merges the Bleakley (2010) preprocessing step—national demeaning—into the main estimation step, to assure that standard errors reflect imprecision in both steps.

Formally, I rewrite the panel model (2) and the cohort-specific cross-section model (3) as

$$Y_{itj} = \left(Exp_t \times M_j\right)\beta + \mathbf{z}'_{itj}\boldsymbol{\alpha} + \mathbf{x}'_{tj}\boldsymbol{\gamma} + \delta_{ct} + \delta_j + \epsilon_{itj} \tag{4}$$

$$Y_{itj} = M_j\beta_t + \mathbf{z}'_{itj}\boldsymbol{\alpha}_t + \mathbf{x}'_j\boldsymbol{\gamma}_t + \delta_c + \epsilon_{itj} \tag{5}$$

Here the new index i identifies individual census observations. The δ_{ct} are dummies for each census year–birth year combination and effect the Bleakley (2010) preprocessing. The variable set **z** holds individual-level traits.

The regressions (5) are implemented for all birth cohorts at once via a single, full-sample regression in which time dummies δ_t are interacted with all the right-side variables. This approach facilitates clustering the standard errors by birthplace, across birth cohorts, to mitigate serial correlation.

To formally test whether Exp helps predict the $\hat{\beta}_t$, I then estimate two versions of (4). The first version introduces three linear spline terms to generalize the step-like functional form of Exp. This loosens the restriction that Exp is flat before and after the transitional ramp-up period, and allows a formal test of whether relative progress in high-malaria regions accelerated and decelerated when expected. Since the Bleakley (2010) text ascribes a 21-year ramp-up phase to Exp, I give each spline segment 21 years of coverage. To be precise, the spline model regression replaces $Exp_t \times M_j$ in (4) with the three terms:

$$t \times M_j,\ \min(0, t - Campaign\ year - 21) \times M_j,\ \min(0, t - Campaign\ year) \times M_j \tag{6}$$

where $Campaign\ year$ is 1920 for the United States and 1957 for Brazil, Colombia, and Mexico; and $\min(\cdot)$ is the minimum function. The sample is restricted to those born between 21 years before the first kink and 21 years after the second, for a range of up to 63 years (data availability permitting).

Giving each segment a length of 21 years reflects an arbitrary choice, but one intended to be minimally so. In general, lengthening the outer segments would give more weight to long-term developments, in a context where the plausibly exogenous variation is short-term. For example, if in the United States, the β_t fell steadily between 1830 and 1865 and then symmetrically recovered between 1865 and 1900, extending the first spline segment from 1899 back to 1830 might give it a flat slope in the best fit, obscuring the steady rise that begins well before the first kink point. On the other hand, shortening the outer segments reduces statistical power. Giving the outer segments the same 21-years span as the inner one therefore seems reasonable.

The second version of (4) used to formally test the explanatory power of Exp retains $Exp_t \times M_j$ as a unitary term in the regression and instead echoes Bleakley (2007, Table VI) in introducing controls for polynomial trends in time. The terms of interest, inserted in **z** in (4), are:

$$\left\{M_j \times t^r\right\}_{r=0,\ldots,d} \tag{7}$$

d ranges up to 5 because Bleakley (2007, note 25), reports testing up to quintic order. Unlike the linear spline models, the polynomial models are fit to the full time span of available data. This gives more influence to longer-term developments, while attempting to compensate with the flexible controls in (7).

As tools for testing the explanatory value of Exp, the two models have advantages and disadvantages. The polynomial models carry some risk of generating spurious results: the true trend may not contain a component of Exp yet may correlate with it the context of these models. The birth year polynomials are inserted to combat this risk, but the higher ones may overparamaterize, imposing a tougher test than a noisy trend could be usually expected to pass, even when it contains an Exp component.[9] For its part, the step model provides a focused and intuitive test of whether relative gains in income and human capital broke from ambient trends in ways naturally explained by malaria eradication efforts. Yet the model is somewhat arbitrarily moored to specific kink dates—1920 or 1957 for the moments of eradication, and 21 years earlier. Perhaps a truer model would shift dates of the eradication campaigns, which did not actually take place within single years. Or it might extend the critical period of sensitivity to

[9] Bleakley (2010, p. 24) warns that "horse-racing the exposure with second-degree trends across cohorts is a more difficult test to pass" in the data sets from Latin America, with their shorter time spans.

malaria into early adulthood, so that even exposure in one's 20s mattered for human capital accumulation and income. Or the opposite: Victora et al. (2008) suggest that health in the first two years of life may matter especially for later outcomes. If the kink points are wrong, then the step model may be less able than the polynomial model to detect a true Exp component, since the model focuses so sharply on whether kinks at occur at predetermined times.

The upshot of these conceptual difficulties is that one should not take any one of the regression results as definitive, and instead exercise judgment in blending all.

2 Themes in the replication and reanalysis

2.1 Pre-analysis plan

I registered a pre-analysis plan for this paper with the Center for Open Science.[10] I did not allow the plan to limit the analysis. But I found little cause to deviate because I had nearly completed the replication and reanalysis of the closely related Bleakley (2007), and this strongly informed the plan for revisiting Bleakley (2010).

The plan sets out several steps, which are listed here with commentary:

- *"Searching the figures and tables for asymmetries, such as one set of regressions being conducted at the individual level and another at the geographic level, and, where appropriate and practical, testing robustness of the results to copying specification choices from one to the other."* Two (arguable) asymmetries are exploited. The U.S. regressions are for whites only while the Brazil, Colombia, and Mexico include all races, while most of the Bleakley (2007) U.S. regressions also include blacks. I also add blacks when expanding the census samples. Also, the Bleakley (2010) long-difference regressions apply to more outcomes than do the panel regressions. These include, for example, literacy and years of schooling in the Latin countries. The reanalysis treats all the outcomes symmetrically.[11]
- *"Formally testing whether the curve fits in figure 4 are statistically significant, and whether those results are robust to inclusion controls for linear or higher-order trends in*

[10] See osf.io/h98yf.
[11] The Bleakley (2006) working paper does also include panel regressions and graphs for these additional outcomes.

time (up to order 5).” The “curve fits” are the graphical fits of Exp to the β_t. The formal methods for assessing this fit were discussed just above.

- “*Testing robustness of the above to 1) a switch from data aggregated by census year, birth year, and birth state to individual-level data; 2) expansion to blacks and women; and 3) incorporation of controls for race, sex, census year, and all their interactions.*” The move to microdata was motivated just above. All these choices mimic the majority of the Bleakley (2007) specifications. Bleakley (2010) argues that restriction to men makes for a cleaner analysis since “their labor-force participation is higher and more consistent across the wide swath of years” (p. 11). Bleakley (2010, note 7) makes a similar argument for excluding blacks, but here the paper is not quite as internally consistent. The Latin American samples include all races, if only because “race was not measured consistently in the Latin America sample” (Bleakley 2010, note 7). However, the present reanalysis is premised on the view that the most plausibly exogenous identifying variation comes the specific timing of eradication, which argues for maximizing power to detect temporal developments over shorter spans, even if at the expense of longer-term comparability. Even if distinctive over the long run, trends for blacks and women could be expected to kink in the same ways as for white men.
- *“When working with aggregate data, testing robustness to weighting by cell size rather than the square-root thereof.”* Weighting by cell size—the number of primary observations behind each aggregated observation in the analysis data set—should better assure efficiency in the face of heteroskedasticity.[12] However, this point is largely moot since I work mainly with microdata.
- “*Testing robustness to the incorporation of newer and larger census samples from IPUMS.*” This is done, as discussed in the next subsection.
- “*In the case of the U.S., testing robustness to switching as much as possible to the data set recently reconstructed from primary sources [for Roodman (2017)] in order to replicate Bleakley (2007).*” This is done.

[12] See note 8.

2.2 Expanded census samples

The IPUMS census microdata collection has expanded steadily over the years: in countries and census rounds included and, at least for the United States, in the size or "density" of samples digitized. Bleakley (2010) largely does not specify the densities of the samples it uses. But they can be estimated from the reported download dates and the history of certain ipums.org pages at archive.org.[13] Table 1, column 1, shows my estimates.

I test robustness by switching to newer, larger IPUMS samples. For the United States, the expansion introduces data for 1860, 1870, and 1930. And it raises the density from 1 percent to 5 percent in 1900 and 1960, and to 100 percent for 1910–40. Column 2 of Table 1 provides more detail. As just noted, in expanding the samples, I add women and, in the U.S. case, blacks. The Latin American IPUMS samples have not become denser since Bleakley accessed them. But more have become available, and are incorporated here: Brazil 2010; Colombia 2005; and Mexico 1995, 2010, and 2015.[14]

All new regressions reported here incorporate person-level sampling weights provided by IPUMS. Most U.S. and Colombia IPUMS samples are "flat," meaning that this weighting is not needed to make them statistically representative. However, there are exceptions (Ruggles et al. 2015; usa.ipums.org/usa/intro.shtml#weights). And more of the Brazil and Mexico samples require weighting because of systematic under- and over-sampling of various subpopulations.[15] Bleakley (2010) does not mention using sampling weights. The paper appears to use them in aggregating the outcome variables into birth place–birth year–census year cells (to form the Y_{ctj}), for I obtain the best matches to the public data when also doing so. However, after aggregation, the Bleakley (2010) regressions are weighted only by the square root of cell size, which—again, going by what produces the best match—is based on the unweighted observation counts within

[13] Bleakley (2010) reports last obtaining U.S. data from IPUMS on November 14, 2005, and last accessing Brazil, Colombia, and Mexico data in April 2006. See the change log at usa.ipums.org/usa-action/revisions and the archive.org histories of ipums.org/usa/sampdesc.html, international.ipums.org/international/sample_designs/sample_designs_br.html, international.ipums.org/international/sample_designs/sample_designs_co.html, and international.ipums.org/international/sample_designs/sample_designs_mx.html.

[14] IPUMS also offers 2005 census records for Mexico, but these lack the birthplace variable BPLMX, which obstructs their use here.

[15] See international.ipums.org/international-action/sample_details.

cells. Thus, the Bleakley (2010) regressions do not fully correct for non-representative sampling within the IPUMS data sets.

My use of IPUMS weights is not pre-registered. However, it is implicitly preregistered in that Roodman (2017) does the same. And one reason it was not pre-registered is that only by examining the public Bleakley (2010) data did I determine that the original does not fully incorporate the weights.

3 Reconstruction of analysis data

From IPUMS microdata, I reconstruct all the Bleakley (2010) left-side variables. As for the right-side variables, I import reconstructed versions for the United States from the Roodman (2017) replication of Bleakley (2007). I do not attempt to reconstruct the right-side variables for the Latin countries, viewing the time cost as prohibitive.[16] In the regressions, I use reconstructed variables where available and take them from the public Bleakley (2010) data otherwise.

To check for problems in the reconstructed variables—or the originals—I compare the two to the degree possible. The public Bleakley (2010) data observe the variables in two forms. Long-difference cross-sections contains one observation, in differences, for each geographic unit. Panel data sets aggregate more finely, within birth year–birth place–census year cells; but they only cover one outcome variable per country.

Table 2 presents means and standard deviations for all Bleakley (2010) outcomes, as well as their cross–data set correlations. All statistics incorporate IPUMS sampling weights. The matches are mostly good, especially in the data arrayed for panel analysis, which is the framework of exclusive interest here. By chance, the panel correlations round to 0.931 for the Unites States and Colombia; the correlation is 1.000 for Brazil and 0.998 for Mexico (right side of

[16] This paper began as an offshoot of a longer-term project to review the evidence of the long-term impact of deworming. Having fully reconstructed the U.S.-focused Bleakley (2007), and discovered the publicly available analysis data for Bleakley (2010) the choices made here amounted to picking low-hanging fruit. The only additional variable reconstruction carried was for the outcomes in the Latin countries, which was made practical by the accessibility of IPUMS International online data system.

Table 2). [17,18] In the long-difference data (left side of the table) the correlations are a bit lower for the U.S. outcomes, at around 0.9 and 0.8, and are much lower for earned income in Brazil, at 0.15.[19] Lacking full access to the original data and code, it is hard to know what causes these discrepancies.

In the case of the United States, I copy from Roodman (2017) the reconstructed right-side variables. Table 3 does for these variables what

Table 2 did for the left-side ones. The first three rows show nearly perfect matches for the indicator of regional malaria burden (M) and the two controls included in all Bleakley (2010) panel specifications, a state-level measure of agricultural wages in 1899 and a dummy for being in the South. The remaining rows turn to the variables introduced in Bleakley (2010)'s "full controls" specifications, which are the focus here.[20] The matches are close, except for the education variables. This is unsurprising given the ambiguity in dates given for the changes: "circa 1902–32". Most likely the reconstructed variables use different editions of the underlying federal government report. And possibly the negative correlation for log change in pupils per teacher owes to Bleakley (2010) inverting this variable, to teachers per pupil—which in itself would be harmless when taking log changes.

Indeed, the juxtaposition of original and reconstruction exposes discrepancies between the Bleakley (2010) text and the Bleakley (2010) data, some of which appear to be implementation errors. Since the publicly available data and code exactly replicate the published Bleakley (2010) tables, the published results reflect all these discrepancies. The cross-state control variables are

[17] Total income in the 1960 Brazil data is reported after censoring into an ordinal variable. Bleakley (2010) appears to "top-code" the 50,000-and-above category as 50,001, so I do the same. For lower categories, range midpoints are used, as documented in the original.

[18] The match with Colombia is most hard-won. After much trial and error, I determined that the "bplcol2" fields of the Columbia data sets, which index the geographic unit, the *municipio*, had been rearranged relative to other variables, as if the column had been sorted in Excel while leaving other columns untouched. Thus, the variable does not in fact obey the coding of the IPUMS International field from which it ultimately derives, BPLCO2. After consulting the primary source for the altitude and temperature variables (Banco de la Republica 1960), I estimate that the mapping to IPUMS codes can be recovered from the Bleakley (2010) public long-difference data using the following algorithm. Sort it by bplcol1 and bplcol2; then numbering the rows starting from 1, except skipping indexes 284 and 473. I cannot tell whether only bplcol2 was rearranged—which in itself would not affect the Bleakley (2010) results—or whether other variables were too, which would be an error.

[19] For Brazil, total income, as distinct from earned income, is of primary interest in the analysis, partly because more census rounds collected it.

[20] Bleakley (2010) Figure 4 is the sole figure in the original exploring the temporal evolution of the β_t in (5). Its specifications all include the full control set.

to have been multiplied by Exp before entering the regressions; they are multiplied by birth year instead. While the text defines Exp assuming childhood lasts 21 years, in the panel data, Exp in fact rises from 0 to 1 over 18 years. (Likewise for Brazil, Colombia, and Mexico.) The control "Doctors per Capita, 1998" is actually residents per doctor. The main text lists the log change in teacher salaries among the controls but Bleakley (2010) Appendix III and the code refer instead to the log change in school term length. The U.S. panel regressions include birth cohorts back to 1815, which is earlier than the 1825 starting point stated in text.

Table 4, below, checks whether these problems drive the Bleakley (2010) U.S. panel results. They do not. The table closely follows the format of Table 4, panel A, of Bleakley (2010), which presents all the U.S. panel estimates, except that it doubles the number of columns. The odd columns copy from the original. The even columns present results obtained from the public Bleakley (2010) data set after properly constructing the interaction terms with Exp, inverting residents per doctor to doctors per resident, and defining childhood as lasting 21 years. As well, observations are weighted by cell size rather than the square root thereof, as set forth in the pre-analysis plan. These fixes (largely not pre-registered) cause no substantive change in the Bleakley (2010) panel results.

4 Time series results

Having reconstructed all of the Bleakley (2010) variables except for the cross-section ones from Brazil, Colombia, and Mexico, I implement the revised designs defined in section 1. To start,

Figure 1, below, strives to imitate the Bleakley (2010) Figure 4, the sole presentation in the original of time series results. Each data point represents an estimate based on (3) of β_t, which, recall, is the cross-sectional association among people born in year t between historical malaria burden and average adult earnings. The graph uses only public Bleakley (2010) data, which aggregates from samples of (white) men. Like the public Bleakley (2010) code,

Figure 1 takes childhood to last 18 years, as in the Bleakley (2010) code. However, For the United States, the dependent variable is log occupational income score; for Brazil and Mexico, log total income; and for Colombia, the log of a Bleakley-constructed variable called the industrial income score. The figure departs substantively from the original only in drawing 95 percent confidence intervals for the point estimates. It departs cosmetically in not superimposing

a plot of the Exp step function. But vertical lines are drawn to mark the birth cohorts at which Exp kinks—the years the eradication campaigns began, and 18 years before.

Figure 1 matches Bleakley (2010) Figure 4, but not perfectly. This is to be expected when original data is used, but original code is not. (The public Bleakley (2010) code only generates tables, not figures.) In all four countries, as in the original, β_t rises with time—generally from negative values toward zero, but in Colombia from approximately zero to positive values.

Figure 2 updates

Figure 1 by fitting to the expanded data set at the microdata level, according to (5). Now, census samples are added or increased in density. Women are included. For the United States, blacks are added too. Sex and race dummies enter the control set. Observations are weighted using IPUMS individual weights. In marking the first potential kink point, childhood is taken to last 21 years, as stated in the Bleakley (2010) text, rather than the 18 in the Bleakley (2010) code.

Except in Mexico, the expanded-sample results appear statistically compatible with the previous, smaller-sample results. In Mexico, an apparent rise *before* the predicted take-off year of 1936 now disappears.

Figure 2 confronts us with the paramount empirical question in this reanalysis: did the cross-sectional association between baseline malaria endemicity and future earnings rise at an historically anomalous rate among the cohorts born in the run-up to eradication, marked by the dashed, vertical grey lines? A glance at Figure 2 suggests that the answer is "yes" in all the countries save Mexico.

To formally test that interpretations, Figure 3 and Figure 4 fit the linear spline and polynomial models, defined in (6) and (7), to the expanded microdata. These figures retain the dots from Figure 2 but, for legibility, drop the confidence intervals. The linear spline fits, in Figure 3, largely support the Bleakley (2010) impact model, even in Mexico. The hypothesis of no acceleration at the first kink is comfortably rejected in Latin America (p = 0.00, 0.00, 0.07 for Brazil, Colombia, and Mexico, clustering standard errors by birth state). An upward bend in the United States appears to have begun earlier than predicted in the Bleakley (2010) impact model,

making the null of no upward bend at the *expected* time harder to reject (p = 0.39). Meanwhile, the null of no deceleration at campaign onset (second kink point) is strongly rejected for the United States, Brazil, and Mexico (p = 0.03, 0.00, 0.05) and more weakly so for Colombia (p = 0.23).

The polynomial fits, in Figure 4, tell a similar story. The fits for models of order 0 to 5 are shown in orange, green, blue, red, purple, and brown, respectively, while the six corresponding p values for the coefficient on $Exp \times M$ are listed beneath. Corresponding estimates of β in (4) are gathered in Table 5. Even with controls up to order 5 in time, the fits to U.S., Brazil, and Mexico data mostly assign a statistically strong positive value to β. The results are more mixed for Colombia, yet generally the p values on $Exp \times M$ stay low.

Last, Figure 5 and Figure 6 apply the methods of the previous two figures to the outcomes for which Bleakley (2010) reports long-difference but not panel results. These are Duncan's socioeconomic indicator (SEI) for the United States, earned income for Brazil, and literacy and years of schooling for all three Latin countries.

Somewhat like the Bleakley (2010) long-difference regressions, these new figures produce a more mixed bag for these outcomes. Turning first to the linear spline fits in Figure 5, in the United States, the trend on Duncan's SEI appears to bend at the first allowed kink, but not at all at the second, reversing the pattern for the closely related socioeconomic index (refer back to the upper-left of Figure 3). In Brazil, while relative progress on earned income (as distinct from total income) slows when expected, it does not appear to accelerate when expected, perhaps owing to low statistical power from small samples in the early years. In none of the Latin countries does relative progress on adult literacy or years of schooling slow at the expected time (second kink point). In all, it bends with statistical significance at the first kink point—but bends the "wrong" way in Mexico.

The polynomial models for these outcomes produce somewhat more encouraging results. Figure 6 depicts these and Table 5 displays the corresponding impact estimates and standard errors. Forced to fit to the full U.S. historical record, the polynomial models confidently endow the treatment term $Exp \times M$ with explanatory power. Polynomial controls also strengthen the fit for earned income in Brazil. For human capital variables, signs, magnitudes, and statistical

significance of the impact estimates vary substantially with the polynomial order, which is easier to see in Table 5. Signs of impact do not appear robust.

Overall, the new time series results mostly strongly support the proposition that reduced childhood malaria exposure increased adult earnings in Latin America. It may well have done so too in the United States too, but there the step model fits less consistently point to acceleration and deceleration with the expected timing (top left of Figure 3 and of Figure 5). Eradication did not so clearly increase literacy or schooling.

5 Conclusion

Bleakley (2010) identifies impacts from variation in the product of two factors: the geographic pattern of baseline malaria burden and the sudden onset of campaigns to relieve that burden. The first factor cannot credibly be viewed as exogenous since it is a marker for climate and geography, and thus economic history. The second can be taken as exogenous, but only in the short term. That malaria eradication campaigns took place between, say, 1900 and 2000, is of a piece with the economic and scientific development of the Americas. That the individual campaigns started in the years they did, rather than a few years before or after, is more an accident of history. Thus, given the informal priors I bring to this study, for it to produce strong evidence of impact, it must demonstrate certain distinctive changes in the outcomes of interest in the time dimension, and that with a precision measured in years, not decades.

In my view, only the time-series analysis performed here fully confronts this challenge. The Bleakley (2010) long-difference regressions speak to whether relative gains occurred in historically malarial areas but not their functional form. The Bleakley (2010) panel regressions get more at functional form, introducing birthplace-specific quadratic time controls. But as presented, it is hard to judge whether these results come from models that are flexibly enough specified to largely absorb ambient trends. If the models are overly parsimonious, they can generate spurious regressions. Graphing the time series patterns and performing formal inference on them provides a clearer view of the temporal variation that is the most credible source of causal identification.

Applying this paper's methods to the Bleakley (2007) assessment of the impact the *hookworm* eradication effort in the American South leads, in my view to a significant update: the suggestion

of historically anomalous convergence coincident with that effort disappears (Roodman 2017). But for malaria, the reanalysis does not trigger much update. Bleakley (2010) finds "that cohorts with less childhood exposure to malaria have higher literacy rates, but results are mixed for years of schooling." The new analysis tends to produce mixed results for both. Meanwhile, it broadly supports Bleakley's "main result" that the evidence indicates that eradication raised adult income.

Separately, this reanalysis points up limitations in the data and code archiving practices of the American Economic Association journals. One purpose of those archives is to increase confidence in published results by documenting precisely how they are obtained. Current archiving practices appear to undercut this purpose in two respects. First, they provide no access to the primary data, or at least to the code that transforms the primary data into the analysis data. The *American Economic Review's* own assessment of compliance with its data availability policy suggested as much in 2011. "Simply requiring authors to submit their data prior to publication may not be sufficient to improve accuracy….The broken link in the replication process usually lies in the procedures used to transform raw data into estimation data and to perform the statistical analysis, rather than in the data themselves" (Glandon 2011).[21] Second, code is provided for tables only, not figures. Yet figures too can play a central role in a study's conclusions. Like tables, they distill large amounts of data to inform inference. They ought to be replicable.

As a result of these two gaps, to the extent that Bleakley (2010) and this reanalysis disagree, it is impossible to be sure why they do so. And to the extent they agree when the reanalysis copies variables from the publicly archived data, one cannot know to what extent the shared conclusions are driven by bugs in the non-archived transformation code. These avoidable ambiguities mis-serve the researchers and decisionmakers that journal authors and publishers aspire to influence.

[21] All of the code and data for this reanalysis are posted online, with one exception. The IPUMS International confidentiality rules prevent the redistribution of the IPUMS extracts for Brazil, Colombia, and Mexico. For that, a precise description of the extracting query is posted.

References

Banco de la Republica. 1960. *Atlas de Economia Colombiana*. Bogota, Colombia: Departamento de Imprenta del Banco de la Republica. Available at ufdc.ufl.edu/uf00077026/00002.

Bleakley, Hoyt. 2006. "Malaria in the Americas: A retrospective analysis of childhood exposure." Working paper. 2006-35. Centro de Estudios sobre Desarrollo Económico, Universidad de los Andes. Bogotá. Available at sticerd.lse.ac.uk/seminarpapers/dg12032007.pdf.

Bleakley, Hoyt. 2007. "Disease and Development: Evidence from Hookworm Eradication in the American South." *The Quarterly Journal of Economics* 122 (1): 73–117. DOI: 10.1162/qjec.121.1.73.

Bleakley, Hoyt. 2010. "Malaria Eradication in the Americas: A Retrospective Analysis of Childhood Exposure." *American Economic Journal. Applied Economics* 2 (2): 1–45. DOI: 10.1257/app.2.2.1.

Duncan, Otis Dudley. 1961. "A Socioeconomic Index for All Occupations." *Class: Critical Concepts* 1: 388–426.

Glandon, Philip J. 2011, "Appendix to the Report of the Editor: Report on the American Economic Review Data Availability Compliance Project," *American Economic Review* 101(3), 696–699. DOI: 10.1257/aer.101.3.684.

Inter-university Consortium for Political and Social Research (ICPSR). *Historical, Demographic, Economic, and Social Data: The United States, 1790–1970,* Ann Arbor, MI. DOI: 10.3886/ICPSR02896.v3.

Minnesota Population Center. Integrated Public Use Microdata Series, International: Version 6.5 [dataset]. Minneapolis: University of Minnesota, 2017. DOI: 10.18128/D020.V6.5.

Roodman, David. 2017. Replication and reanalysis of Bleakley (2007): The impacts of hookworm Eradication in the American South. Working paper. Open Philanthropy Project. Available at files.givewell.org/files/DWDA%202009/Interventions/Deworming/2017-Roodman-replication-and-reanalysis-of-Bleakley-hookworm.pdf.

Ruggles, Steven, Katie Genadek, Ronald Goeken, Josiah Grover, and Matthew Sobek. 2015. Integrated Public Use Microdata Series: Version 6.0 [dataset]. Minneapolis: University of Minnesota. DOI: 10.18128/D010.V6.0.

Victora, Cesar G., Linda Adair, Caroline Fall, Pedro C. Hallal, Reynaldo Martorell, Linda Richter, Harshpal Singh Sachdev, and Maternal and Child Undernutrition Study Group. 2008. "Maternal and Child Undernutrition: Consequences for Adult Health and Human Capital." *The Lancet* 371 (9609): 340–57. DOI: 10.1016/S0140-6736(07)61692-4.

TABLE 1. U.S. IPUMS CENSUS SAMPLES IN ORIGINAL AND EXPANDED DATA SETS

Census year	Original (estimated; percent)	Expanded (percent)
1860	0	1.2[a]
1870	0	1.2
1880	100	100
1890	0	0
1900	1	5
1910	0.4	100
1920	1	100
1930	0	100
1940	1	100
1950	1	1
1960	1	5
1970	1	1
1980	5	5
1990	5	5
2000	5	5

[a] Excludes slaves.

Source: Authors' calculations.

TABLE 2. SUMMARY STATISTICS OF BLEAKLEY (2010) DEPENDENT VARIABLES

	Long-difference cross-section			Panel data		
	Original	New	Correlation	Original	New	Correlation
United States						
Log occupational income score	0.324	0.292	0.897	3.286	3.279	0.931
	(0.084)	(0.082)		(0.112)	(0.132)	
Log Duncan's SEI	0.560	0.504	0.806			
	(0.102)	(0.074)				
Observations	48	48		9604	9605	
Brazil						
Log total income	–0.012	–0.007	0.945	8.625	8.701	1.000
	(0.080)	(0.087)		(2.268)	(2.262)	
Log earned income	–0.012	–0.017	0.152			
	(0.075)	(0.223)				
Literacy	–0.002	–0.003	0.994			
	(0.053)	(0.052)				
Years of schooling	0.032	–0.001	0.949			
	(0.537)	(0.513)				
Observations	24	28		2231	2587	
Colombia						
Industrial income score	–0.049	–0.055	0.855	–0.112	–0.106	0.931
	(0.080)	(0.098)		(0.174)	(0.200)	
Literacy	–0.020	–0.018	0.973			
	(0.100)	(0.096)				
Years of schooling	–0.480	–0.487	0.973			
	(0.632)	(0.617)				
Observations	523	525		38070	39513	
Mexico						
Log earned income	–0.044	–0.110	0.925	9.659	9.391	0.998
	(0.173)	(0.328)		(2.914)	(2.902)	
Literacy	–0.017	–0.021	0.993			
	(0.072)	(0.082)				
Years of schooling	–0.229	–0.373	0.931			
	(0.497)	(0.529)				
Observations	32	32		2965	2965	

Variable means displayed with standard deviations in parentheses beneath. Third and sixth columns show cross–data set correlations. "Original" results computed from public Bleakley (2010) data. "New" results computed after reconstructing the data sets. All statistics weighted by cell-level sums of the IPUMS-provided individual weights in the reconstructed data set.
Source: Authors' calculations.

TABLE 3. SUMMARY STATISTICS OF U.S. CROSS-STATE VARIABLES

Variable	Original	New	Correlation
Malaria share of mortality, 1889 (M)	0.318 (0.326)	0.295 (0.302)	0.994
Agricultural wage, 1899 ($/month)	16.938 (6.393)	17.415 (6.396)	0.999
South	0.271 (0.449)	0.271 (0.449)	1.000
Residents per doctor, 1898	743.333 (244.706)	743.361 (244.719)	1.000
Board of health spending, 1898 ($/1,000 residents)	6.333 (13.253)	6.779 (13.321)	0.976
Infant mortality rate, 1890 (per 1,000 births)	162.797 (68.310)	105.358 (51.474)	0.983
Hookworm prevalence among army recruits, 1917–19	0.069 (0.097)	0.069 (0.097)	1.000
Log change in teacher salaries, circa 1902–32	1.444 (0.175)	3.216 (0.199)	0.775
Log change in school term length, circa 1902–32	0.114 (0.122)	0.169 (0.149)	0.631
Log change in pupils/teacher, circa 1902–32	0.118 (0.275)	–0.043 (0.172)	–0.362
Adult literacy rate, 1910	0.907 (0.074)	0.907 (0.074)	1.000
Population urban, 1910	0.340 (0.231)	0.392 (0.225)	0.982
Population black, 1910	0.107 (0.164)	0.107 (0.163)	1.000
Male unemployment, 1930	0.043 (0.018)	0.079 (0.026)	0.913
Observations	48	48	

Variable means displayed with standard deviations in parentheses beneath. Final column shows cross–data set correlations. All statistics are unweighted. "Original" results computed from public Bleakley (2010) data. "New" results computed after reconstructing the data set from primary sources. Sample excludes Alaska, Hawaii, and the District of Columbia.
Source: Authors' calculations.

TABLE 4. REPLICATION OF BLEAKLEY (2010) PANEL ESTIMATES OF THE EFFECT OF CHILDHOOD EXPOSURE ON LOG OCCUPATIONAL INCOME SCORE IN THE UNITED STATES

Degree of polynomial trend for year of birth	Mean reversion and region controls						Additional controls					
	0		1		2		0		1		2	
	Original	New	Original	New	Original	New	Original	New	Original	New	Original	New
Baseline	0.131	0.183	0.115	0.196	0.131	0.093	0.130	0.174	0.099	0.199	0.111	0.055
	(0.030)	(0.038)	(0.031)	(0.038)	(0.025)	(0.024)	(0.032)	(0.036)	(0.035)	(0.044)	(0.020)	(0.012)
Post-1920 break in birthplace time trend	0.082	0.103	0.094	0.139	0.105	0.073	0.076	0.100	0.080	0.140	0.082	0.056
	(0.015)	(0.016)	(0.020)	(0.023)	(0.024)	(0.017)	(0.019)	(0.016)	(0.023)	(0.027)	(0.018)	(0.013)
Allow for birthplace x time effects	0.103	0.108	0.110	0.138	0.123	0.079	0.086	0.106	0.094	0.138	0.110	0.066
	(0.026)	(0.016)	(0.030)	(0.021)	(0.023)	(0.017)	(0.027)	(0.017)	(0.033)	(0.025)	(0.021)	(0.014)
Drop early census years (<1930)	0.106	0.107	0.105	0.084	0.032	0.014	0.098	0.107	0.108	0.068	0.030	0.014
	(0.021)	(0.016)	(0.017)	(0.018)	(0.015)	(0.014)	(0.022)	(0.016)	(0.020)	(0.015)	(0.018)	(0.014)
Add region x year x YOB effects	0.131	0.175	0.116	0.194	0.131	0.090	0.127	0.166	0.098	0.197	0.108	0.050
	(0.030)	(0.038)	(0.029)	(0.037)	(0.024)	(0.025)	(0.032)	(0.036)	(0.034)	(0.043)	(0.019)	(0.013)

"Original" results generated with Bleakley (2010) public data and code. "New" results use same data and address coding issues described in text.
Standard errors in parentheses, clustered by state.
Source: Bleakley (2010), Table 4; authors' calculations.

TABLE 5. IMPACT ESTIMATES ON ALL BLEAKLEY (2010) OUTCOMES, CONTROLLING FOR POLYNOMIAL TIME TREND UP TO ORDER 5

Country	Outcome	Coefficient on $M \times Exp$					
Order of Polynomial Trend		0	1	2	3	4	5
U.S.	Log occupational income score	0.087	0.070	0.064	0.041	0.040	0.015
		(0.010)	(0.012)	(0.014)	(0.015)	(0.014)	(0.016)
U.S.	Log Duncan's SEI	0.096	0.068	0.056	0.042	0.032	0.031
		(0.023)	(0.035)	(0.033)	(0.032)	(0.033)	(0.027)
Brazil	Log total income	0.439	0.477	0.486	0.504	0.433	0.253
		(0.071)	(0.132)	(0.120)	(0.082)	(0.097)	(0.122)
Brazil	Log earned income	0.276	0.285	0.340	0.323	1.000	0.692
		(0.060)	(0.134)	(0.113)	(0.103)	(0.180)	(0.212)
Brazil	Literacy	0.121	0.009	0.048	–0.040	–0.102	0.069
		(0.026)	(0.037)	(0.032)	(0.033)	(0.041)	(0.046)
Brazil	Years of schooling	0.846	0.883	0.870	0.906	0.256	0.906
		(0.358)	(0.565)	(0.596)	(0.458)	(0.790)	(0.729)
Colombia	Industrial income score	0.031	0.018	0.039	0.025	0.029	0.170
		(0.009)	(0.011)	(0.012)	(0.021)	(0.023)	(0.058)
Colombia	Literacy	0.020	0.009	0.018	–0.020	–0.006	–0.011
		(0.012)	(0.010)	(0.010)	(0.018)	(0.019)	(0.035)
Colombia	Years of schooling	0.368	–0.015	0.180	0.151	0.303	–0.079
		(0.156)	(0.176)	(0.151)	(0.382)	(0.352)	(0.759)
Mexico	Log earned income	0.250	0.133	0.199	0.255	0.274	–0.146
		(0.051)	(0.069)	(0.062)	(0.136)	(0.124)	(0.278)
Mexico	Literacy	0.015	–0.031	–0.052	0.012	0.019	0.138
		(0.030)	(0.030)	(0.025)	(0.021)	(0.023)	(0.051)
Mexico	Years of schooling	–0.386	–0.433	–0.511	0.542	0.895	1.266
		(0.276)	(0.424)	(0.351)	(0.403)	(0.439)	(0.605)

Estimates based on expanded data set, including women and, in the U.S. case, blacks as well as whites. Regressions weighted by IPUMS-provided sampling weights. Standard errors clustered by state of birth in parentheses.
Source: Authors' calculations.

FIGURE 1. REPLICATION AND EXTENSION OF BLEAKLEY (2010) FIGURE 4: ORIGINAL DATA SETS

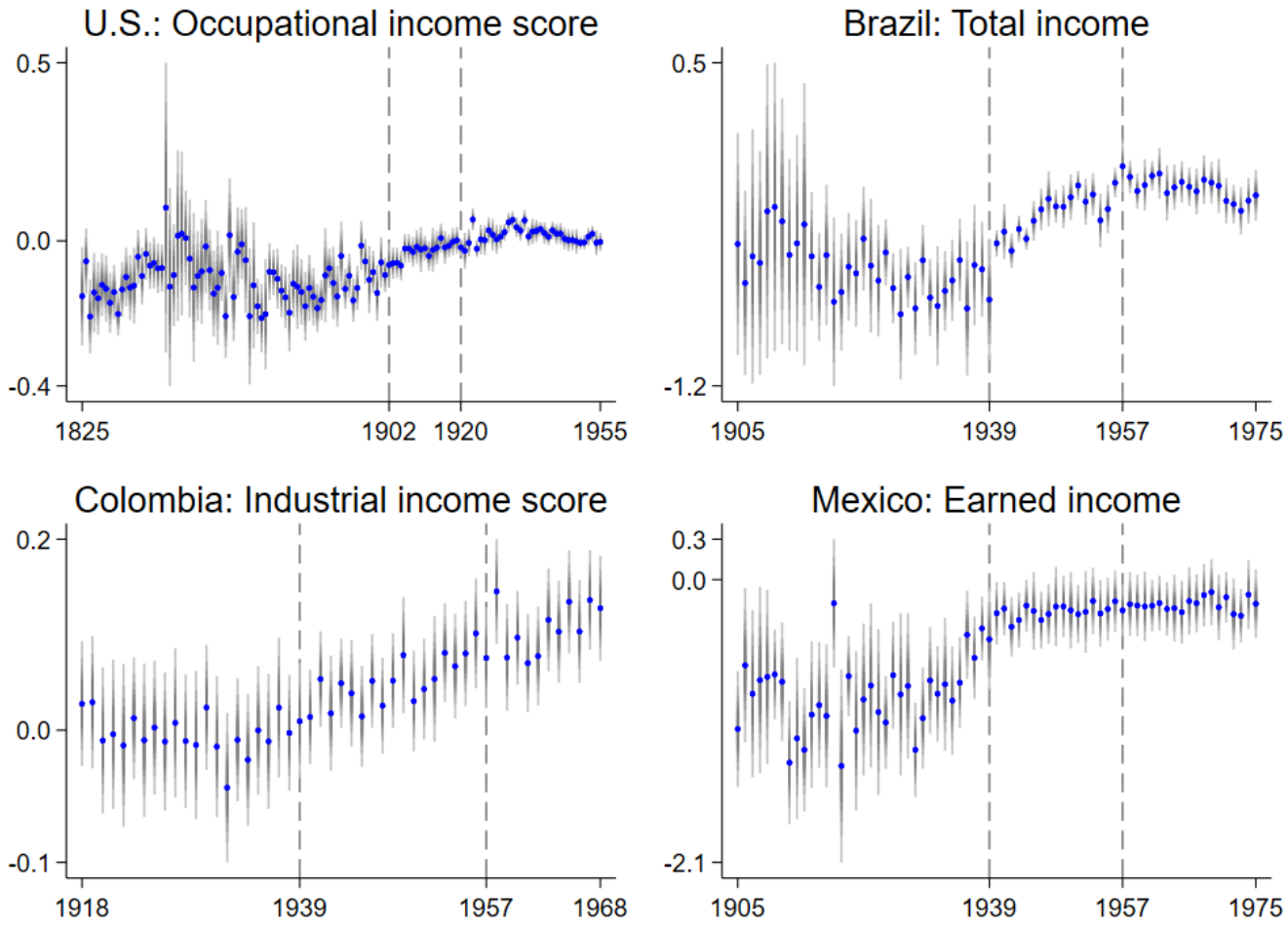

FIGURE 2. REPLICATION AND EXTENSION OF BLEAKLEY (2010) FIGURE 4: EXPANDED DATA SETS

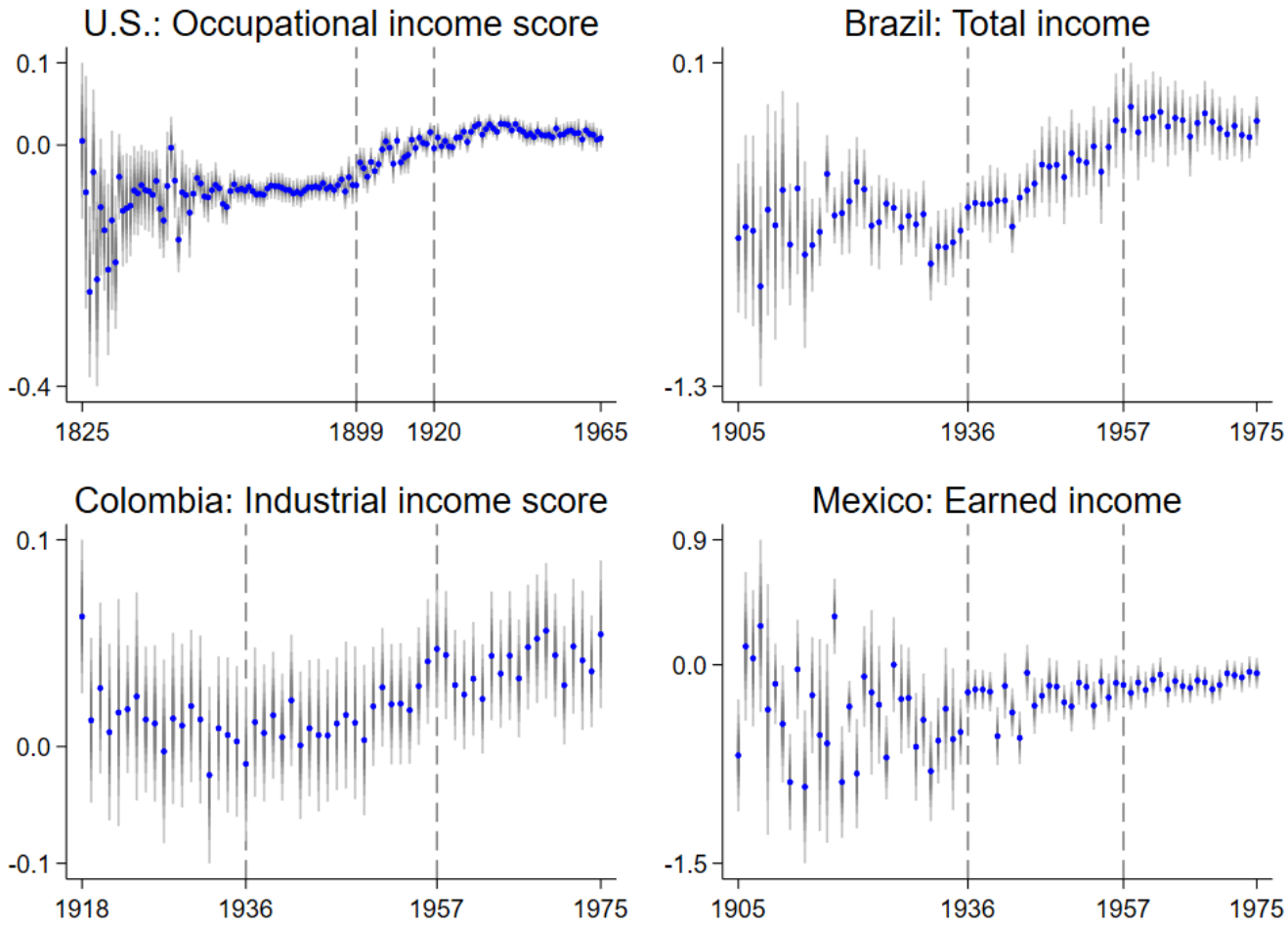

FIGURE 3. REPLICATION AND EXTENSION OF BLEAKLEY (2010) FIGURE 4: MODEL WITH LINEAR SPLINE GENERALIZATION OF STEP FUNCTION

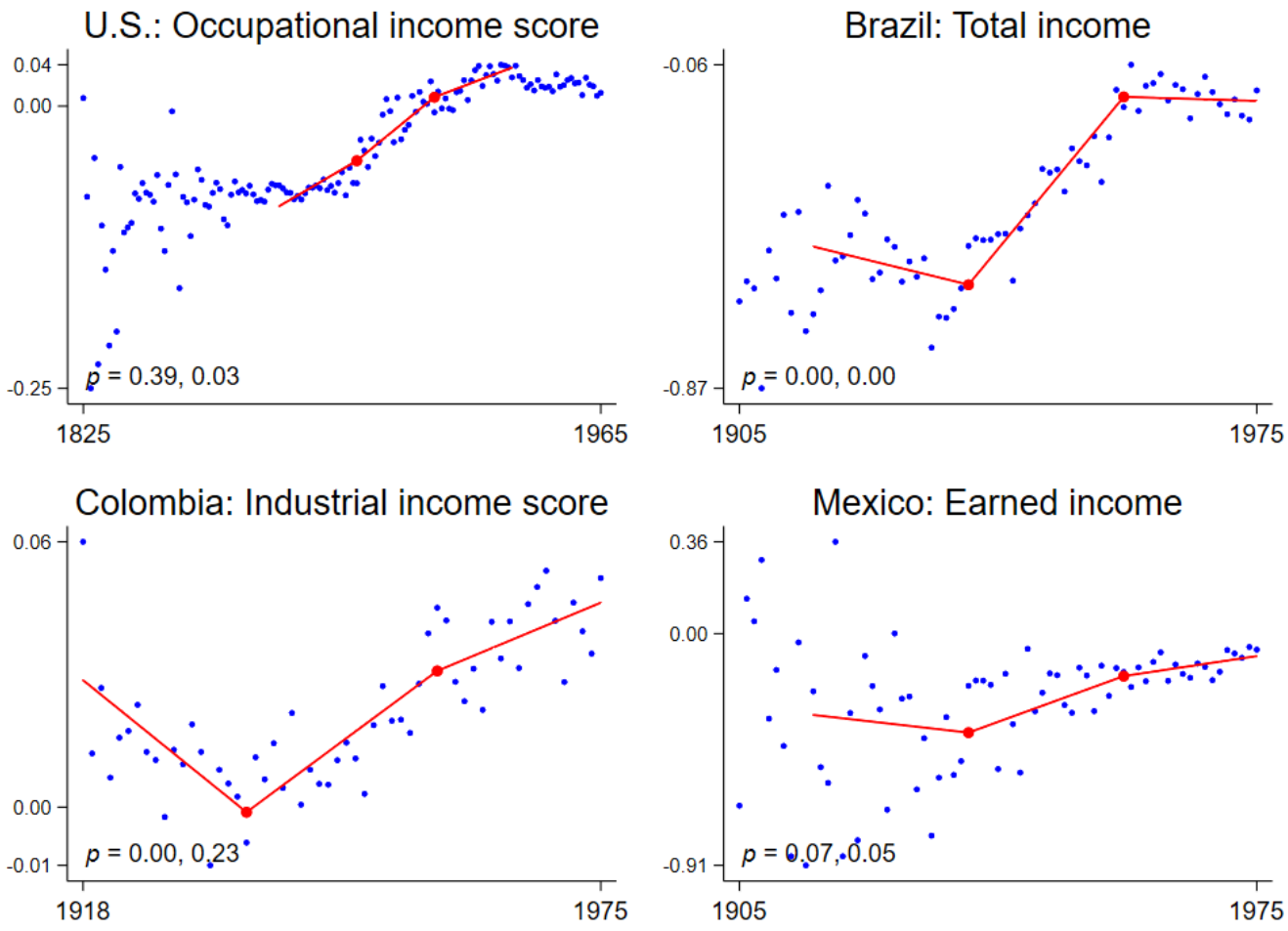

FIGURE 4. REPLICATION AND EXTENSION OF BLEAKLEY (2010) FIGURE 4: MODEL WITH POLYNOMIAL TIME CONTROLS, FIT TO EXPANDED DATA SET

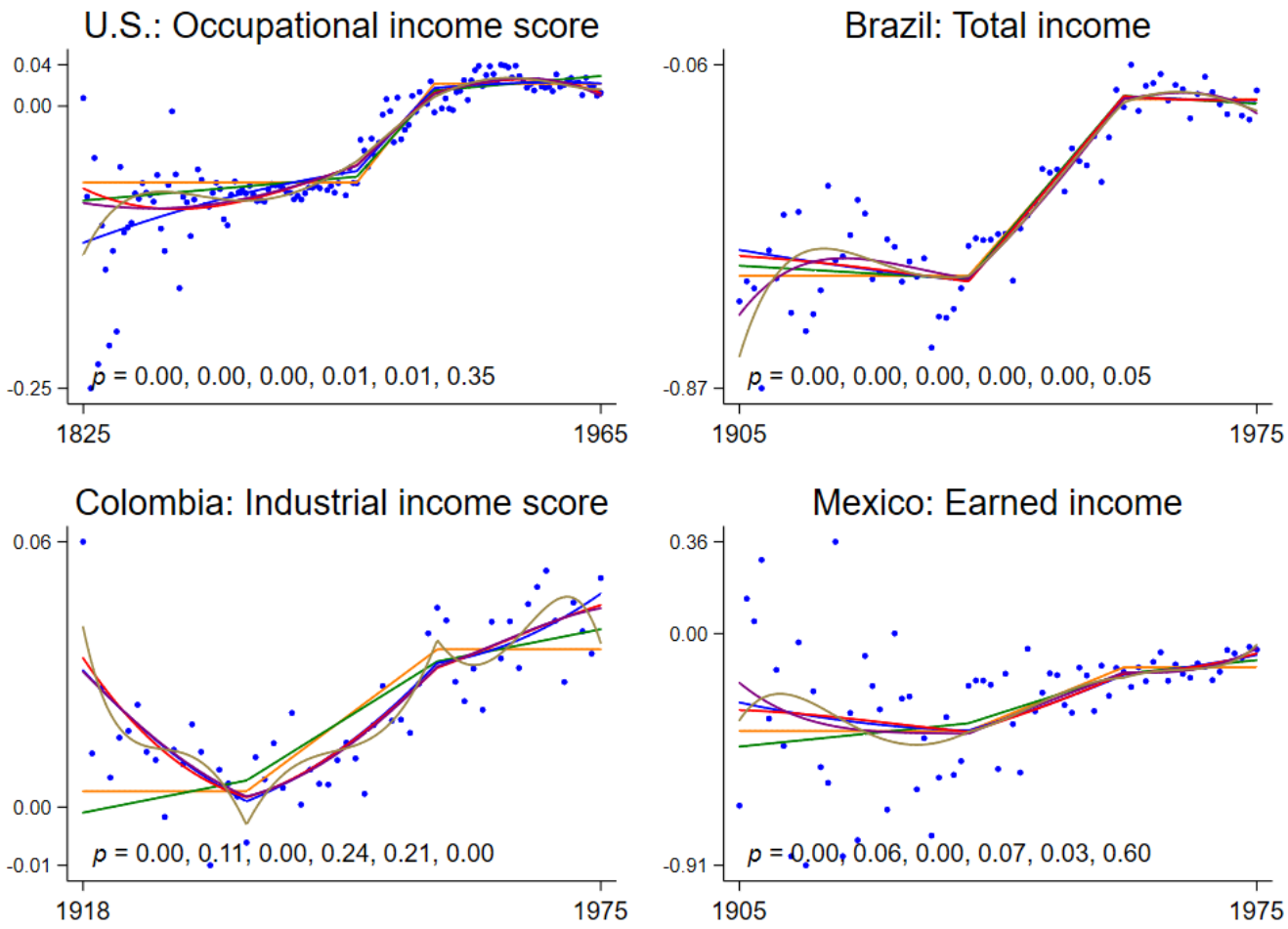

FIGURE 5. REPLICATION AND EXTENSION OF BLEAKLEY (2010) FIGURE 4: MODEL WITH LINEAR SPLINE GENERALIZATION OF STEP FUNCTION, ALTERNATIVE OUTCOME MEASURES

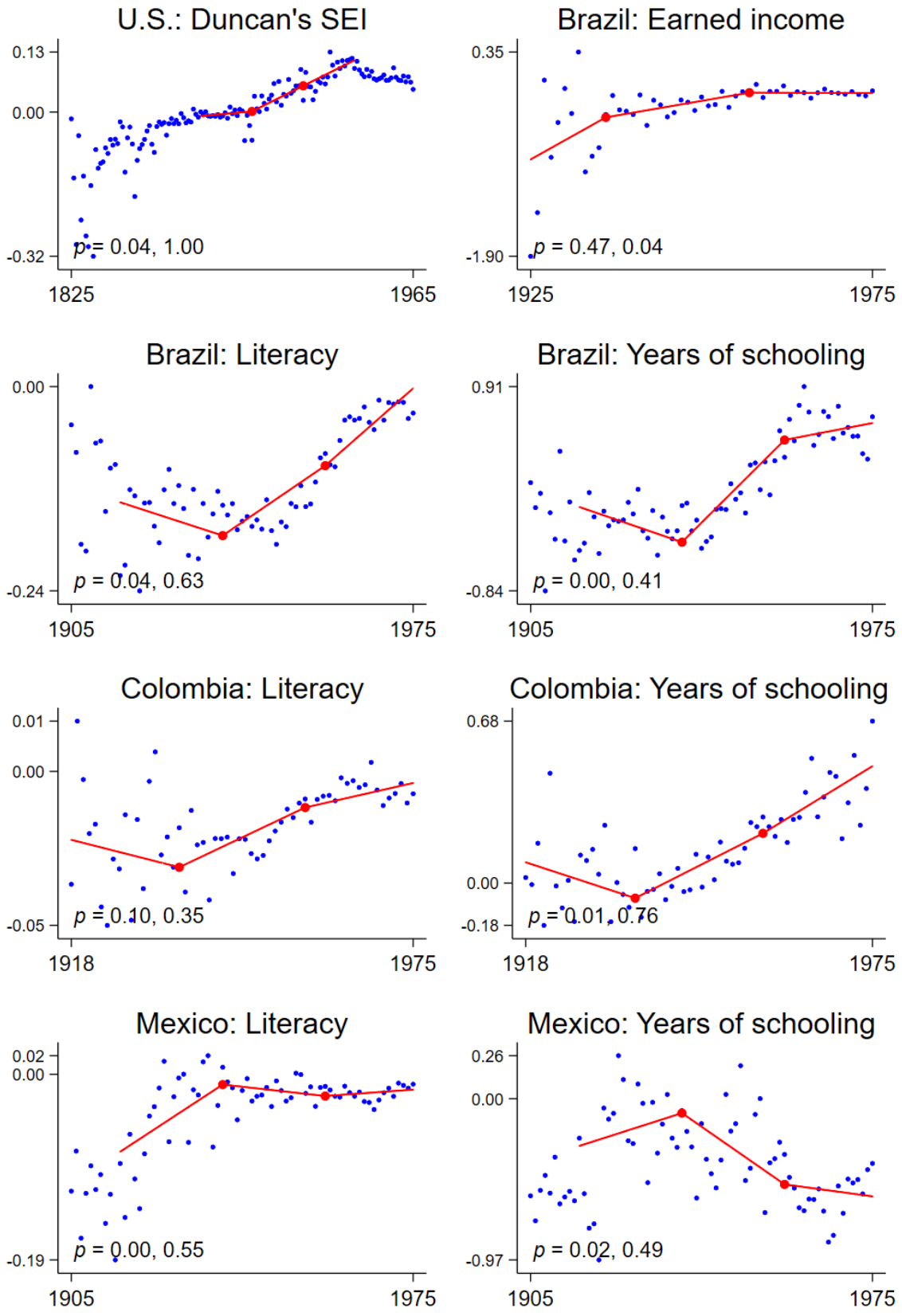

FIGURE 6. REPLICATION AND EXTENSION OF BLEAKLEY (2010) FIGURE 4: MODEL WITH POLYNOMIAL TIME CONTROLS, FIT TO EXPANDED DATA SET, ALTERNATIVE OUTCOME MEASURES