EJW
ECON JOURNAL WATCH
Scholarly Comments on
Academic Economics

ECON JOURNAL WATCH 21(2)
September 2024: 339–356

Can We Detect the Effects of Racial Violence on Patenting? Reanalyzing an Article by Lisa Cook

Michael Wiebe

LINK TO ABSTRACT

In 2014, Lisa Cook published an article in *Journal of Economic Growth* titled, "Violence and Economic Activity: Evidence from African American Patents, 1870–1940." Cook studied the effect of racial violence on innovation by Black Americans over the period 1870–1940. The article is widely cited,[1] is discussed in the media (Duffin and Childs 2020), and has been described as 'seminal' (Center for Economic and Policy Research 2022; Khang 2020). Alex Albright et al. (2022) cited the article with this description: "Cook (2014) studies the effects of race riots and lynchings between 1870 and 1940 and finds that these forms of violence and insecurity reduced patenting by Black people by more than 15% annually from 1882–1940." Is such a precise quantitative summary warranted?

In this article, I use the original data and code from Cook (2014) to investigate the reliability of the results.[2] The time-series results are not robust to using a more complete measure of patents, and are not consistent with the timing of patent applications and grants: since patents are granted roughly one year after application, violence should reduce applications in the same year and grants in the following year; this pattern does not hold in the data. The panel data results are based on a dataset where most observations are missing, and hence are not reliable.

1. As of May 2024, there are 167 citations on Google Scholar.
2. I am able to computationally reproduce most of the results in the paper; see Appendix B for details.

For example, Cook reports a negative effect of riots on patenting, but there are only five riots in the panel data (compared to 35 in the time-series data). While the broad conclusions from Cook (2014) may be true, they are not supported by the evidence in the article.

Data irregularities

Cook has two measures of patents per year: (1) using the year the patent was applied for, and (2) using the year the patent was granted.[3] Cook's Figure 1 reports Black patents per million using grant year, while Cook's Figure 2 shows Black patents per million using application year.[4] Comparing the two figures reveals that the scale differs by a factor of about 10, which seems like an error because the two measures of patents should be approximately equal, differing only in the timing of applications and grants. It turns out that the discrepancy is because Cook divided Black grant-year patents by the White population to calculate patents per million and divided Black application-year patents by the Black population.[5]

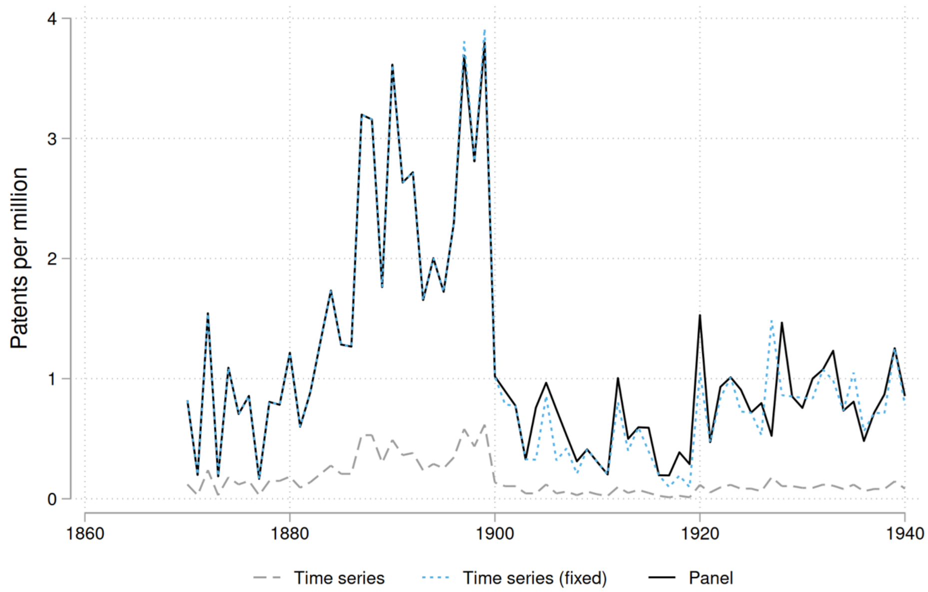
I correct the grant-year variable by obtaining the number of patents and dividing by the Black population. As we can see in Figure 1, the original time series variable is much smaller than the corrected variable due to dividing by the larger White population.

Moreover, there is a discrepancy in the number of patents used by Cook (2014) in the time-series and panel-data analyses. Cook's Table 6 uses year-level data, regressing patents on lynchings, riots, and segregation laws separately by race. This time-series data has 672 Black patents. Cook's Table 7 used state-year panel data to run the same regression, but there are 702 patents.[6] I aggregate the panel-data variable across states to make a comparison with the time-series variable. Figure 1 above shows that the corrected time-series variable and the aggregated panel-data variable are identical up to 1896, after which they diverge. One possible explanation is that Cook revised the patent variable by adding more patents, but

3. That is, Cook has data on the subset of applications that are granted. There is no data on unsuccessful applications.
4. Appendix Table A1 summarizes the data structure (time series vs. panel data) and patent variable (grant-year vs. application-year) used in Cook's results.
5. When I multiply the original patent rate by the white population (interpolated by constant imputation), I obtain integer values, i.e., the raw patent counts. Hence, we can infer that Cook used constant imputation to fill in missing values of the decennial population data. For the application-year variable, Cook used the Black population and exponential interpolation. See my code for details. These data cleaning choices are not discussed in the article or documented in the text or code. Cook did not respond to any of my emails about the paper.
6. Cook reports collecting data on 726 patents. It is not clear why the totals in the replication files are lower.

updated only the panel-data variable and not the time-series variable. Below I test whether the time-series regressions are robust to using the aggregated panel-data patent variable.

Figure 1. Grant-year patents, Black inventors



Notes: *Time series* is the original grant-year patent variable used in Table 6, which divides by the White population. *Time series (fixed)* corrects this variable by dividing by the Black population. *Panel* is the panel data patent variable used in Table 7, aggregated to the year-level.

I document several data errors in Appendix C, where dummy or count variables take fractional values; I list reproducibility issues and coding errors in Appendix B.

Time series regressions

Cook (2014) used grant-year patents as the dependent variable for the main analysis and reported results from it in her Tables 6 through 8, and did not directly use application-year patents.[7] Cook's Table 6 uses time-series data to regress grant-

7. Cook says that the dependent variable used in her Table 6 is "patents per capita applied for in year t and granted to individuals of race i" (p.235). However, the code does not match the text, since the variable actually used is grant-year patents, i.e., patents granted in year t. Note that Cook's Table 9 uses application-

year patents on lynchings, riots, and segregation laws. I conducted a more nuanced analysis of the hypothesis that racial violence deterred Black patents, exploiting the differences in the year of application for the patent and the year the patent was granted. Cook reports that the average time gap between application and grant is 1.4 years.[8] Assuming that racial violence has a contemporaneous effect on Black inventors' activities, we would expect that violence in year t deters inventing in year t, which would show up as a decrease in patent applications in year t. Similarly, if violence in year t deters inventing in year t, then the number of patents granted should decline primarily in year $t+1$, as grants lag applications by roughly one year. The effect of violence in year t on patents granted in year t should be smaller than the effect on patent applications in year t, since the applications for patents granted in year t were made (on average) in year $t-1$ (and hence unaffected by violence in year t).[9] By the same reasoning, the effect of violence in year t on patents granted in year t should be smaller than the effect on patents granted in year $t+1$.

To test these predictions, in Table 1 I repeat Cook's Table 6 column 3 regression using data on Black patents by application year and grant year.[10] I use the three versions of the grant-year variable from my Figure 1: the original time-series variable that divided the number of patents by the White population, the corrected variable that divides by the Black population, and the aggregated panel-data variable (calculated from Cook's Table 7 panel data by summing patents across states). Since the application-year variable is missing in 1940, I omit that year from all regressions, which leads to slightly different grant-year results compared to the original (which has N=56). Column 1 presents the contemporaneous effect of violence on the number of patents to Black persons per million in the application year. There is no relationship between racial violence and application-year patents, contradicting the hypothesis that violence has a contemporaneous effect on inventing.

Column 2 replicates the original result from Cook (Table 6, Column 3). Estimates are very similar to the original paper, with lynchings and riots negatively correlated with contemporaneous patent grants.[11] Column 3 uses the corrected

---

year data to match Black and White patents, then aggregates over time.

8. I calculate the average gap by decade being approximately 1 or more years starting in the 1880s (matching the estimation sample in Table 1 below). The gap grows to roughly 2 years by the end of the sample.

9. If patent applications were processed immediately, then the time gap between application and grant would be 0, and violence would decrease both contemporaneous patent applications and grants by the same amount. Conversely, if the gap was 5 years (say), then violence should reduce patent applications in year t and patent grants in year $t+5$.

10. Table 6, Column 1 pools both races, while Columns 2 and 3 restrict the sample to White and Black patents, respectively. Since the main conclusions are for Black inventors, I focus on Column 3.

11. Cook's Table 6 incorrectly shows the lynching estimates in Columns 2 and 3 as being significant at the 5% level, when the p-values are larger than 0.05.

grant-year variable that divides by the Black population. The coefficients are nearly identical to Column 2, demonstrating that normalizing by the White or Black population does not have a big effect (since the transformation is similar to rescaling by a scalar). However, under the assumption that violence has a contemporaneous effect, there should be a smaller association between racial violence and patents granted (compared to patent applications). In contrast, when using the aggregated panel data variable in Column 4, the correlations are much smaller and nonsignificant. This is noteworthy because the aggregated panel data variable has 702 total patents in comparison to the 672 patents in the time-series variable. Hence, Cook's main result does not hold when using a more complete patent variable.

TABLE 1. Timing of patents

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Application-year patents	Grant-year patents: time series	Grant-year patents: fixed	Grant-year patents: panel	Application-year patents	Grant-year patents: time series	Grant-year patents: fixed	Grant-year patents: panel
Lynchings	0.131 (0.466)	−0.844* (0.459)	−0.886* (0.473)	−0.234 (0.374)	0.356 (0.488)	−0.547 (0.348)	−0.592* (0.351)	−0.105 (0.319)
Major Riots	−0.018 (0.059)	−0.137* (0.069)	−0.135* (0.068)	−0.076 (0.049)	−0.040 (0.046)	−0.106** (0.053)	−0.101* (0.053)	−0.049 (0.057)
Segregation laws	−0.025 (0.035)	0.033 (0.026)	0.033 (0.026)	−0.000 (0.025)	−0.012 (0.038)	0.059** (0.025)	0.059** (0.024)	0.016 (0.028)
L.Lynchings					0.566 (0.488)	0.336 (0.416)	0.311 (0.412)	0.009 (0.486)
L.Major Riots					−0.012 (0.107)	0.251*** (0.061)	0.264*** (0.063)	0.177** (0.081)
L.Segregation laws					−0.031 (0.032)	−0.036 (0.032)	−0.034 (0.031)	−0.022 (0.036)
1921 dummy	0.195 (0.288)	−0.459*** (0.169)	−0.480*** (0.168)	−0.980*** (0.167)	0.307 (0.332)	−0.350* (0.196)	−0.369* (0.189)	−0.986*** (0.218)
Observations	55	55	55	55	55	55	55	55
R-squared	0.197	0.301	0.297	0.214	0.23	0.557	0.565	0.352

Notes: All models are estimated following Cook's specification using OLS in first differences. Heteroskedasticity-robust standard errors are in parentheses. Controls include a linear time trend; a post-1899 dummy; year dummies for 1910, 1913, and 1928; and the first-difference of the log of the Miron-Romer Industrial Production Index. *Application-year patents* is the number of patents per million by the year they were applied for. *Grant-year patents: time series* is the number of patents per million by the year they were granted; this variable incorrectly divides by the White population. *Grant-year patents: fixed* corrects *Grant-year patents: time series* by correctly dividing by the Black population. *Grant-year patents: panel* is the number of patents per million, calculated from the panel data by summing patents across states. Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

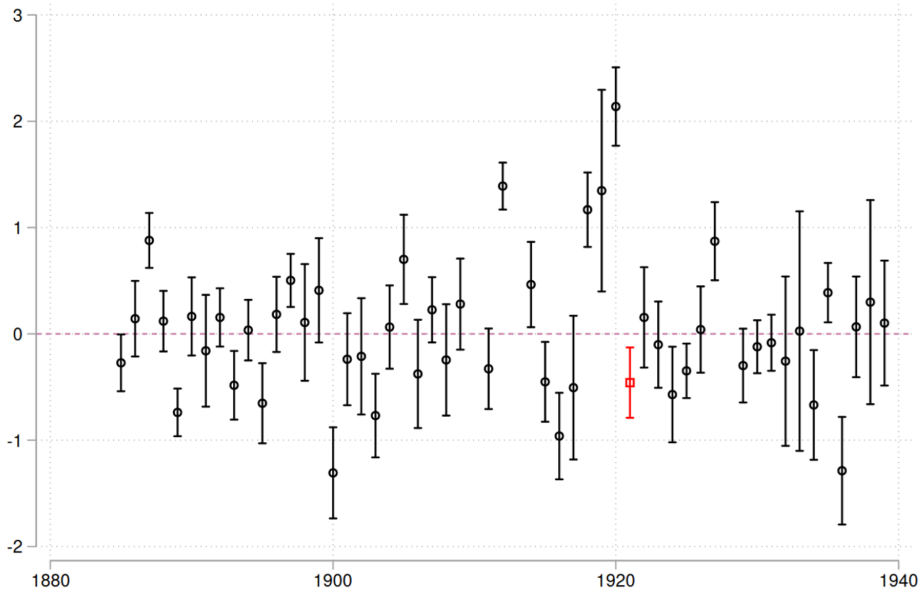
Columns 5–8 test for a lagged effect of violence on patents. Column 5 tests for a lagged effect on application-year patents, and shows no relationship. Against the prediction of a negative effect, riots in the previous year are positively correlated with patent grants for all three variables. The coefficient on lagged lynchings is positive for the two time-series grant variables (Columns 6 and 7), but the correlation is nonsignificant. Thus, if racial violence has a contemporaneous effect on patenting, three predictions are falsified: first, a decrease in contemporaneous application-year patents; second, a decrease in lagged patent grants; and third, a smaller effect on contemporaneous grants than on lagged grants. Overall, these results cast doubt on the main finding in Cook (2014).

An alternative hypothesis is that racial violence affects the behavior of the patent office, instead of influencing individual inventors. Patent examiners may have delayed the granting of Black patents during years with high racial violence, to avoid becoming a target themselves. This explanation is consistent with the effect on grants being negative for contemporaneous violence and positive for lagged violence. However, this hypothesis has several problems. First, if the panel data variable is preferred, then there is no negative effect on contemporaneous grants.[12] Second, as Cook noted, the time gap between application and grant was the same for White and Black inventors (2014, 226 n.15). If patent offices were delaying grants for Black inventors, the time gap should differ by race. Third, since race is not listed on the patent, examiners may not have been able to discriminate by race.

Cook (2014) included a dummy variable for 1921 to capture the effect of the Tulsa Race Riot on patenting. However, the estimate of the 1921 dummy in Column 1 of Table 1 (patents by application year) is positive, inconsistent with the hypothesis that racial violence directly deters inventing and patent applications. To investigate this issue result further, in Figure 2 below I measure the distribution of year-specific shocks and compare it to the 1921 dummy. Specifically, for each year from 1885 to 1939, I rerun the Table 6, Column 3 regression using patent grants with a time fixed effect for that year (excluding the years of the three existing year dummies). I plot the coefficients and 95% confidence intervals below, with the original 1921 dummy noted in red. While Cook singles out the 1921 dummy as representing the negative effect of the Tulsa Race Riot, here we can see that the 1921 effect is not especially large. The large positive effect in 1919 is particularly puzzling, given that it is the year in the sample with the most riots (see Cook's Fig. 2).

12. And if the time series variable is preferred, then we require an explanation for the positive effect of segregation laws in Columns 6 and 7.

Figure 2. Time effects from Cook's Table 6, Column 3, for varying years



Note: For each year, this figure plots the coefficient and 95% confidence interval on the time fixed effect from the Table 6, Column 3 regression, where the 1921 dummy is replaced with a year dummy from the corresponding year.

Next, I check the robustness of Cook's time series results for Black patents in her Table 6, Column 3. I vary whether the dependent variable is log or level patents, whether the lynching variable is log or level, and whether the riot and segregation law variables are differenced. The original regression is in Column 1 of Table 2 below, using the first difference of log patents and log lynchings, with riots and segregation laws in levels and undifferenced. Cook does not justify taking the log of patents and lynchings, or leaving riots and segregation laws undifferenced.

The lynching effect remains significant at the 10% level in the level-log specification (Column 3), but loses statistical significance in the level-level regression (Column 4). The riot effect loses statistical significance when using level patents. When I difference the other violence variables (Column 5), the riot effect becomes stronger and the segregation law effect is positive and significant at the 10% level. The year-1921 effect is much smaller in the level patent regressions. Overall, the results are sensitive to whether the patent variable is log-transformed. Since this specification choice is not strongly motivated (note that the panel regressions in Cook's Table 7 use level patents), this lack of robustness suggests that annual data is too aggregated to be suitable in testing Cook's hypothesis.

TABLE 2. Robustness of time series regressions

	(1)	(2)	(3)	(4)	(5)
	Log y	Log y	Level y	Level y	Log y
D log lynchings	−0.908* (0.461)		−0.094* (0.052)		−0.592** (0.277)
D lynchings		−0.134** (0.064)		−0.015 (0.010)	
Riots	−0.132* (0.070)	−0.137** (0.067)	−0.014 (0.015)	−0.014 (0.015)	
D riots					−0.182*** (0.049)
Segregation laws	0.036 (0.026)	0.025 (0.025)	0.005 (0.004)	0.004 (0.003)	
D segregation laws					0.032* (0.018)
1921 dummy	−0.538*** (0.180)	−0.598*** (0.175)	−0.020 (0.031)	−0.025 (0.031)	−0.570*** (0.193)
Observations	56	56	56	56	56
R-squared	0.283	0.281	0.118	0.120	0.473

Notes: Replications of Cook's Table 6, Column 3. Robust standard errors. Controls include a linear time trend; a post-1899 dummy; year dummies for 1910, 1913, and 1928; and the first-difference of the log of the Miron-Romer Industrial Production Index. Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Finally, I further relax the assumption that the effect of violence on patenting is same-year, by testing for cumulative effects of violence. Specifically, I use a moving average of the violence variables, with the window extending backwards and varying in length from two to five years (including the current year). Table 3 shows that there is no consistent pattern. Hence, in this dataset we cannot detect either a contemporaneous or a lagged effect of violence on patenting.

TABLE 3. Cumulative effects of violence

	(1)	(2)	(3)	(4)
	Window=2	Window=3	Window=4	Window=5
Lynchings	−1.019 (0.853)	0.244 (0.880)	−2.392** (1.103)	−1.527 (1.666)
Riots	0.197 (0.241)	−0.042 (0.246)	0.235 (0.308)	0.225 (0.347)
Segregation laws	0.029 (0.033)	−0.020 (0.040)	0.033 (0.045)	−0.002 (0.052)
Observations	55	55	55	55
R-squared	0.170	0.117	0.186	0.154

Notes: Replications of Cook's Table 6, Column 3. Robust standard errors. The violence variables are a moving average calculated over the window extending backwards, including the current year. Window length is indicated in the columns. Controls include: a linear time trend; a post-1899 dummy; year dummies for 1910, 1913, and 1928; and the first-difference of the log of the Miron-Romer Industrial Production Index. Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Panel data regressions

In her Tables 7 and 8, Cook uses state-level panel data over 1870–1940 to run regressions of patents on lynchings, riots, and segregation laws. This analysis tests for local, contemporaneous effects, and does not consider spatial or temporal spillovers (for example, the Tulsa Race Riot affecting patenting outside of Oklahoma or in years after 1921). However, we can see that the panel is unbalanced: there are 49 states and 71 years in the data, but only 430 observations. A complete, balanced panel would have 3,210 observations, as the number of states grows from 38 in 1870 to 49 in 1940 (including DC; see code for details). So, Cook is using 430/3,210 = 13 percent of the potential sample.

And the pattern of missing data is not random. In Figure 3 I plot the number of observations by state and year. First, in Figure 3a we see that the majority of states have fewer than 10 observations over 71 years. Next, in Figure 3b the sample size is increasing until 1900 before dropping off and rising again starting in 1920. Appendix Figure A1 plots the raw panel data, showing how the composition of states varies by year. Decomposing by region, Appendix Figure A2 shows that the Midwest and Mid-Atlantic regions are relatively overrepresented, while the South and West are relatively underrepresented.

Moreover, consider how this unbalanced panel compares to the full time series. There are 35 riots in the time series data, but only 5 in the panel data (for 14 percent coverage). There are 290 new segregation laws in the time series data, but only 19 in the panel data (for 7 percent coverage).[13] We cannot say whether the same problem applies to the lynchings variable, since the replication files do not have raw count data.

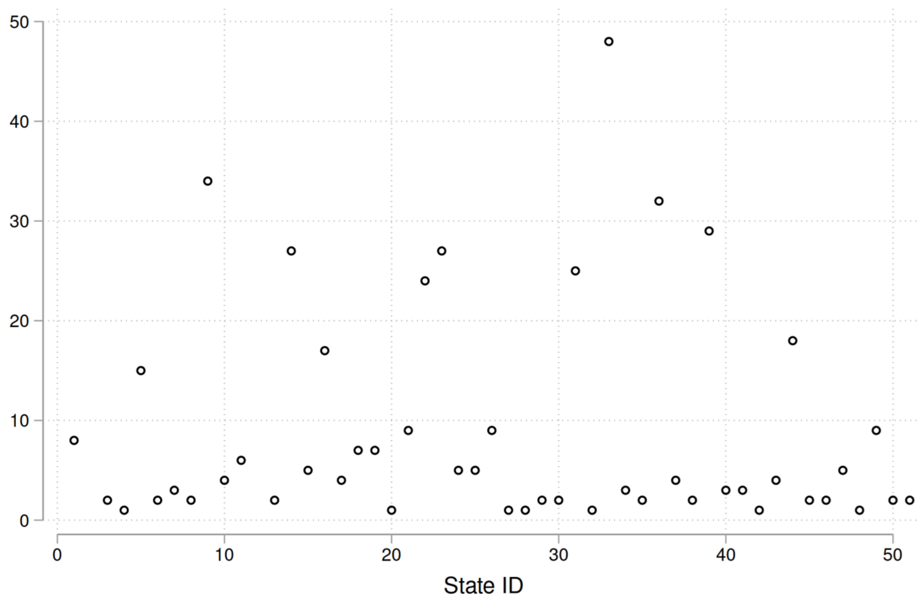
What explains the missing data? One possibility is that Cook (2014) dropped any state-year observation that had a variable with a missing value, as evidenced by the fact that her data has no variables with missing values, but many missing state-year observations.[14] Another explanation is that Cook (2014) did not collect complete data on observations with zero patents. There are 24 observations with patent=0, corresponding to twelve states, each with exactly two observations in the sample: one in 1900 and one in 1930. It appears that Cook collected data on zero-patent states only in 1900 and 1930.

13. The actual number is 19.33. Somehow, one state-year observation has a value of 0.33 for the number of new segregation laws.

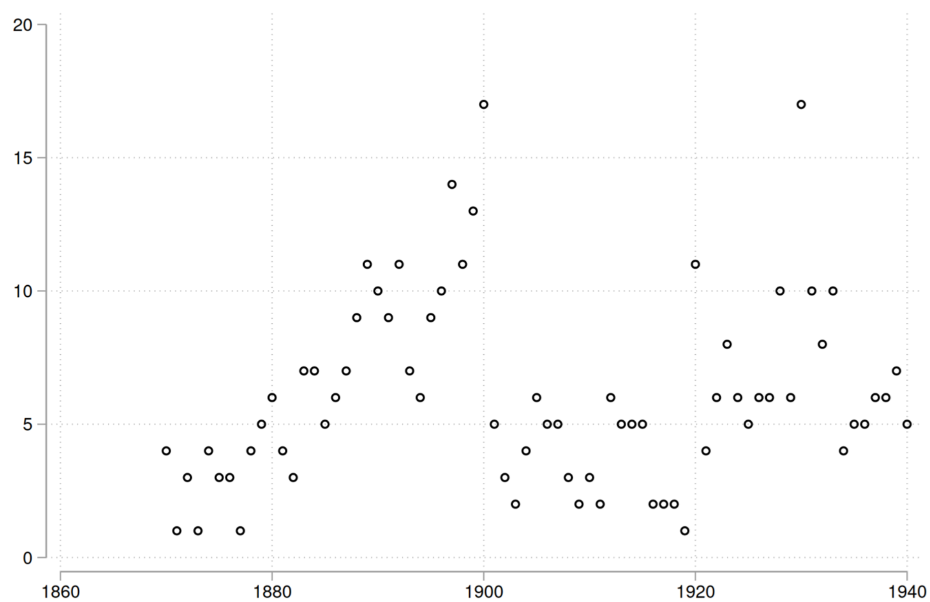
14. That is, the row corresponding to a state-year observation is missing. As a result, replacing missing patent values with 0 is not feasible, since the right-hand-side variables are also missing.

Figure 3. Numbers of observations in the Cook (2014) panel data

(a) Observations by state



(b) Observations by year



With this low level of data coverage and not-at-random missingness, it is unclear how to interpret the results in Cook's Tables 7 and 8. It is possible that the results are unbiased estimates of the population effects, and would remain stable as the missing data was filled in. Especially considering the high prior probability that racial violence and patents are negatively correlated, we should place some weight on this. On the other hand, it is possible that the results are false positives. We don't know. When working with small effects and noisy data, statistically significant results are expected (Gelman and Carlin 2014).

Aside from the problem of missing data, there also arises the issue of application and grant timing. As hypothesized, racial violence should affect patent applications in the same year, and grants in the following year, given the one-year lag between applications and grants. Cook has different data files for patents by application year and grant year, and of the three violence variables, only the riots variable is similar across datasets.[15]

TABLE 4. Timing of patents: panel data

	(1)	(2)	(3)	(4)
	Grant-year patents	Application-year patents	Grant-year patents	Application-year patents
Major riots	−0.362*** (0.070)	0.334 (0.578)	−0.370 (0.243)	1.685*** (0.083)
L.Major riots			1.502*** (0.236)	−0.049 (0.336)
Observations	422	433	193	205

Notes: All models are estimated using random effects. Standard errors are clustered by state. Control variables include illiteracy rate, share of African Americans by state, number of firms per capita, region dummies, and year dummies for 1910, 1913, and 1928. Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Hence, in Table 4 I regress patents by grant year and application year on contemporaneous and lagged riots (omitting lynchings and segregation laws).[16] As with the time series data, the deterrence model is not supported by the data. Grants are negatively correlated with violence in the same year, while the applications for

15. There are 5 riots in the estimation sample when using the grant data, and 4 riots when using the application data. The other two variables have different definitions. The application dataset uses cumulative segregation laws, while the grant dataset uses annual segregation laws. The lynchings variable is different across datasets, with no clear explanation. Moreover, the industry participation variable is only in the grant data, so I do not control for it (grants and applications mostly occur in different years, so the datasets cannot be merged).

16. There are 425 observations in the original Table 7, Column 3 regression using grant-year patents. My Column 1 has 422 observations because I have to merge the 'number of firms' variable from the application data, which is missing years 1870–1872. It is not clear why there are more observations in Column 2 when using the application data.

the patents are not (Columns 1 and 2). In Column 3, we again see that lagged riots are positively correlated with patent grants, even though the sample size is half as large because of the unbalanced panel. In Column 4, including lagged riots reveals a large positive correlation between application-year patents and contemporaneous riots. However, the number of riots in the estimation sample varies from one to five, which makes noisy data the most plausible explanation for the positive effect in both datasets.

Finally, I check the robustness of Cook's panel regressions in her Table 7. Cook uses a random-effects estimator on the basis of a Hausman test, but fixed-effects estimators address time-invariant omitted variables (and are more commonly used for panel data), so I repeat Cook's Table 7 using state and year fixed effects. The results are presented in Table 5 below; the sample size is smaller due to singleton observations being dropped. The lynching effect is larger, but loses significance when controls are added. The original riot effect is significant at the 1% level, but here loses significance with added controls. Also, the original riot effect was driven by the 1870–1917 period (Column 5), while here it is driven by the 1918–1940 period (Column 6). More seriously, the segregation law effects are positive and statistically significant, contrary to Cook's hypothesis. Overall, the panel regressions are not robust, which is consistent with the low level of data coverage.

TABLE 5. Robustness of panel regressions using fixed effects

	(1)	(2)	(3)	(4)	(5)	(6)
Lynchings	−0.257* (0.128)	−0.262** (0.129)	−0.146 (0.101)	−0.150 (0.101)	−0.293* (0.157)	3.359 (2.424)
Major riots	−0.308* (0.180)	−0.283 (0.179)	−0.399* (0.233)	−0.373 (0.226)	−0.159 (0.311)	−1.886** (0.708)
Segregation laws	0.175* (0.100)	0.182* (0.107)	0.201** (0.090)	0.204** (0.092)	0.321 (0.243)	0.244 (0.298)
Illiteracy rate	−1.722 (1.041)	−1.704 (1.059)	−0.707 (0.841)	−0.689 (0.864)	0.343 (1.482)	−14.417 (8.933)
Number of firms			211.438 (130.155)	213.441 (131.118)	102.789 (157.360)	398.450 (461.568)
Industry part. rate		−0.280 (0.572)		−0.294 (0.614)	−0.431 (0.674)	−0.825 (1.017)
Observations	419	418	412	411	248	133
R-squared	0.370	0.371	0.395	0.396	0.399	0.304

Notes: Dependent variable is the number of patents granted to Black inventors in a state-year. State and year fixed effects. Standard errors are clustered by state. Control variables include illiteracy rate and share of African Americans by state. Column 5 restricts the sample to 1870–1917. Column 6 restricts the sample to 1918–1940. Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Conclusion

To summarize, the results from the time-series analysis in Cook (2014) are not robust to using a more complete patent variable and are not consistent with the hypothesis that racial violence reduces patents in the year for which they are applied and patent grants in the following year. The results are not robust to alternative model specifications, and I find no evidence of lagged effects of violence. The results from the panel-data analysis are highly questionable because most of the data is missing.

Cook's conclusions nevertheless remain plausible. Lynchings, race riots, and segregation laws were a severe problem, and it would be astonishing if they did not have pervasive effects on the lives of Black people.

But with the data available, it is unrealistic to think we can statistically detect causal effects. Credible causal inference would require more complete data as well as an identification strategy more convincing than regression with controls. Descriptive analysis is the most that this dataset can support, and is a valuable contribution in itself, along with the rich qualitative and historical evidence in the paper. Cook deserves credit for pursuing this important research question and putting in years of effort to collect the patent data.

Recent research reports findings that are consistent with Cook's claims: Jhacova Williams (2022) shows that historical lynchings reduce contemporary Black voter registration;[17] Albright et al. (2022) studies the Tulsa Race Massacre and finds that it had persistent negative effects on Black Americans; and Abhay Aneja and Guo Xu (2021) show that Woodrow Wilson's segregation of the federal government increased racial inequality. Future research should continue to bring attention to the consequences of racism in America's past.

17. However, the results in Williams (2022) have been questioned by Haddad, Kattan, and Wochner (2023), who find the effect to be driven by four outlier counties.

Appendix A. Data characteristics

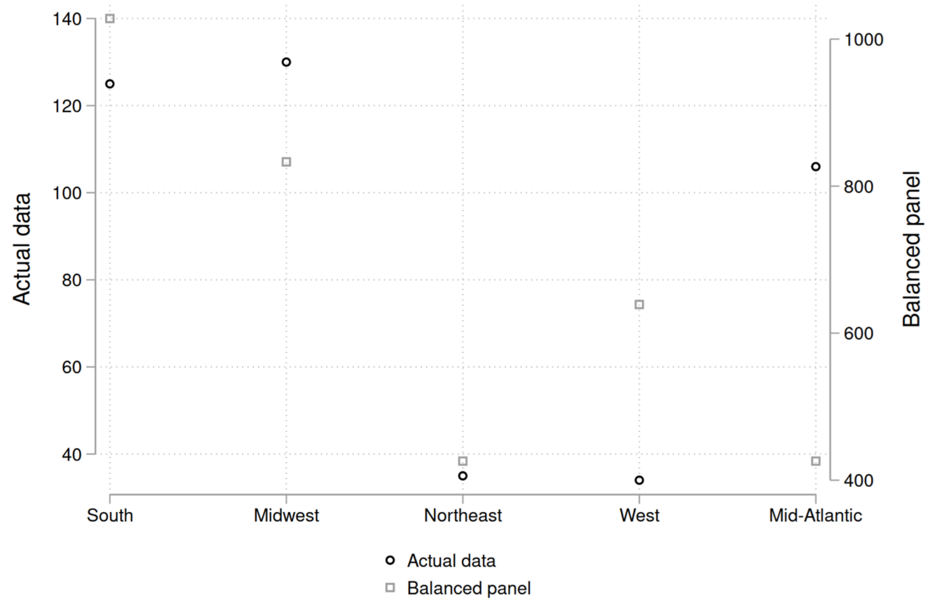
TABLE A1. Datasets and patent variables used in Cook (2014)

	Structure	Patent variable
Figure 1	Time series	Grant-year
Figure 2	Time series	Application-year
Table 6	Time series	Grant-year
Table 7	Panel data	Grant-year
Table 8	Panel data	Grant-year
Table 9	Panel data	Application-year

Figure A1. State-year observations

Note: State-year observations from the panel data used in Cook's Table 7 and 8.

Figure A2. Observations by region



Notes: Number of states by region: South 15, Midwest 12, Northeast 6, West 12, Mid-Atlantic 7. Eleven states enter after 1870, and hence have fewer than 71 years in the complete panel. See code for details.

Appendix B. Reproducibility issues

Regarding computational reproducibility, Cook's code has several problems:

- The code for Figures 1, 2, and 3 is in Stata graph editor format, which cannot be run from a do-file.
- Figure 1 uses the variable *patgrntpc*, patents by grant-year per capita, but the graph refers to patents per million. Similarly, Table 5 reports 'Patents, per million', but the code uses *patgrntpc*. The variable should be named 'patents by grant-year per million'.
- There is no code for Table 4.
- Equation 1 and Table 6 refer to patents per capita, but the variable in the code, *patgrntpc*, has mean values of 0.16 for Blacks and 425 for Whites; this is patents per million, not per capita.
- The text says that the dependent variable used in Table 6 is "patents per capita applied for in year *t* and granted to individuals of race *i*" (Cook

2014, 235). However, the code uses grant-year patents, i.e., patents *granted* in year *t*. Application-year patents are used in Table 9.

- The code for Table 6 refers to a variable *LMRindex*, but the dataset contains *DLMRindex*.
- Section 3.2 mentions that the state-level regressions use data over 1882–1940, but the code uses data over 1870–1940.
- The code for Table 7 uses a variable, *estbnumpc*, for the number of firms per capita, but it is not included in the dataset. It is included in the Table 9 data, so I am able to control for it by merging with that dataset.
- Table 7 includes the 'number of firms' variable in columns 3–6, but the code controls for it in Column 1 as well.
- In the notes to Tables 7 and 8, Cook writes that "Standard errors robust to clustering on state and year are in parentheses." However, the code only clusters by state, using *vce(cl stateno)*.
- The code for Table 8 has an error in its clustering command, using the incorrect syntax *vce(stateno)* instead of the correct *vce(cl stateno)*.
- The code for Table 8 does not exactly reproduce the results in the paper. When I run the code, I get N=429, while Cook's regressions have N=428.
- The code for Table 9 does not reproduce the results in the paper.
- In the text, Cook says there are 714 patents used in Table 9, but the actual number is 712.
- The data for Table 9 has different variables than the data for Table 7. The Table 9 data includes cumulative segregation laws, while the Table 7 data has annual segregation laws. The lynching variable is also different. For example, California has no lynchings in the Table 7 data, but nonzero lynchings in every year in the Table 9 data.

Appendix C. Data errors

There are a few data errors, which I correct in my reanalyses:

- State 9 has the South dummy equal to 1 for all years, but also has the Mid-Atlantic dummy equal to 0.33 in 1888.
- State 14 has the Midwest dummy equal to 1 in all years except 1886, when both it and the South dummy are 0.5.
- State 31 in 1909 has a value of 0.33 for 'number of new segregation laws,' but the value should be an integer.

Data and code

Data and code used in this research is available from the journal website (**link**).

References

Albright, Alex, Jeremy Cook, James Feigenbaum, Laura-Thorne Kincaide, Jason Long, and Nathan Nunn. 2022. After the Burning: The Economic Effects of the 1921 Tulsa Race Massacre. Working paper, October 20. **Link**

Aneja, Abhay and Guo Xu. 2021. The Costs of Employment Segregation: Evidence from the Federal Government Under Woodrow Wilson. *Quarterly Journal of Economics* 137(2): 911–958.

Center for Economic and Policy Research. 2022. CEPR Statement in Support of the Nomination of Dr. Lisa Cook to the Federal Reserve Board of Governors. January 25. Center for Economic and Policy Research (Washington, D.C.). **Link**

Cook, Lisa D. 2014. Violence and Economic Activity: Evidence from African American Patents, 1870–1940. *Journal of Economic Growth* 19(2): 221–257.

Duffin, Karen, and Mary Childs. 2020. Patent Racism. *Planet Money* (National Public Radio, Washington, D.C.), June 12. **Link**

Gelman, Andrew and John Carlin. 2014. Beyond Power Calculations: Assessing Type S (Sign) and Type M (Magnitude) Errors. *Perspectives on Psychological Science* 9(3): 641–651.

Haddad, Joanne, Lamis Kattan, and Timo Wochner. 2023. Comment on "Historical Lynchings and the Contemporary Voting Behavior of Blacks." *I4R Discussion Paper Series* 32. Institute for Replication (Ottawa, Canada). **Link**

Khang, Hyun-Sung. 2020. The Accidental Economist. *Finance & Development Magazine* (International Monetary Fund, Washington, D.C.), December. **Link**

Williams, Jhacova. 2022. Historical Lynchings and the Contemporary Voting Behavior of Blacks. *American Economic Journal: Applied Economics* 14(3): 224–253.

About the Author

Michael Wiebe is an economist and data scientist based in Vancouver, Canada. He completed a Ph.D. in economics at the University of British Columbia, specializing in political economy and development economics. His dissertation studied whether mayors in China are promoted based on GDP growth, showing that certain results in that literature are not robust. He has also replicated research on air pollution, the measles vaccine, and agglomeration effects for innovation. His email address is maswiebe@gmail.com.

Go to archive of Comments section
Go to September 2024 issue