



EJW

ECON JOURNAL WATCH
Scholarly Comments on
Academic Economics

ECON JOURNAL WATCH 21(2)
September 2024: 274–303

Temperature and Economic Growth: Comment on the Published Article by Kiley

David Barker¹

[LINK TO ABSTRACT](#)

In a previous issue of this journal, I commented (Barker 2023a) on a Federal Reserve working paper written by Michael Kiley (2021) about the relationship between GDP growth and temperature. I found that his results were influenced by a small number of unusual observations, and that simulated random data mimicking GDP growth and temperature but with no relationship between the two could falsely show a relationship using Kiley's techniques. I also found that Kiley's primary result is statistically insignificant.

My paper was published in March of 2023. Kiley was sent a copy of the paper and invited to respond in this journal, but he has not replied. Kiley submitted a slightly revised version of his paper to *Economic Inquiry* in February 2023, it was accepted by that journal in January 2024, and that journal published it in its July 2024 issue (Kiley 2024). It makes no mention of my critique. In the *Economic Inquiry* article, Kiley made no changes to his results presented in the Fed working paper, but he added a section containing two sets of robustness checks. One set is in response to Richard Newell et al. (2021), who had pointed out that regressions of growth on temperature are sensitive to the inclusion of time trends. The second set deals with the effect of outlier observations, which I discussed in my *Econ Journal Watch* comment (Barker 2023a).

In this paper I show that the new robustness checks are inadequate, and I strengthen my original critique. In Barker (2023a) I showed that a small number of influential observations drove Kiley's results. I also constructed simulated data

1. Views expressed in this comment are my own and not those of the Iowa Board of Regents or any other organization I am affiliated with.

in which there was no relationship between annual temperature fluctuations and growth, and demonstrated that using Kiley's estimation method on these data was likely to show that an effect existed. In this paper I respond to Kiley's robustness checks, add additional criticisms of Kiley (2024), and present an intuitive explanation of why simulated data can trick Kiley's econometric method into showing a false relationship between growth and temperature.

This article is the fifth in a series of critiques of literature claiming to find a connection between temperature and economic growth; the previous four are Barker 2022; 2023a; 2023b; 2024. The commented-on articles have received considerable media attention and appealed to certain agendas. It is important that voters, policymakers, and academics know the truth. Scholarly debate is the high road to truth. Unfortunately, for all of the articles I have commented on, none of the commented-on authors have replied to the critique. They are, accordingly, included in a list called *Sounds of Silence* ([link](#)).

Here, I first describe Kiley (2024). I discuss why his results are not statistically significant if standard errors are calculated properly. Then I discuss the data that Kiley uses, the characteristics of these data, and the difficulties arising from those characteristics. Next, I describe the robustness checks that Kiley added to his paper and show that they are inadequate. I add robustness checks showing that the nature, magnitude, and sign of Kiley's results are changed by different specifications and the elimination of outliers. Finally, I will discuss other reasons why Kiley's model is hopelessly flawed and illustrate those reasons with simulated data.

Description of Kiley (2021; 2024)

I described Kiley's original paper in Barker (2023a). The original paper (Kiley 2021) uses annual data on per capita GDP growth and temperature from 124 countries from 1961 to 2010. He performs quantile regression analysis, concluding that high temperatures reduce growth and that the effect is strongest at times when economic growth is already weak. Standard errors of the coefficients of the quantile regressions are obtained using a bootstrap resampling method.

The regressions in Kiley (2021; 2024) contain controls for year and country fixed effects, and also for country-specific quadratic trends. In other words, the regressions include dummy variables indicating country multiplied by time and time squared, where time is equal to one for the first year in the sample, 1961, up to 50 for the year 2010. The purpose of these control variables is to allow the measurement of the effect of temperature on growth holding constant quadratic trends in each country that are assumed to be independent of temperature.

Kiley's specific result is that the marginal effect on per capita GDP growth

of an additional degree of temperature for a country already at 25.6 degrees Celsius with growth in the bottom 10th percentile is negative with 99.1% confidence, and it is of larger magnitude for a country experiencing growth in the bottom 10th percentile than at the 90th percentile of growth. This result comes from quantile regression on data from all 124 countries, with each country allowed a different quadratic time trend of growth.

Kiley's results imply that a permanent increase of one degree Celsius in average temperature in an already warm country with median GDP growth will reduce growth by an amount that will result in GDP that is more than 50 percent lower after 50 years than it would be with no warming. They also imply 95.6% confidence that GDP would be 25 percent lower and 99.8% confidence that GDP would be 10 percent lower. At the average temperature of the United States, an increase of one degree Celsius would be expected to decrease GDP by 10.2 percent, although this estimate is not statistically significant. With a 3.7-degree increase in temperature, as is projected by the IPCC (Burke et al. 2015), the reduction in GDP in a warmer country would be 91 percent.

Kiley (2024) contains the results from Kiley (2021) and adds a section of robustness checks. I next discuss new critiques of Kiley's estimation method, and then I will examine his added robustness checks.

Statistical significance

In this section I show that an alternative bootstrap method and weighting of observations to correct for heteroskedasticity and outlying observations lead to much higher estimated standard errors of the effect of temperature on growth than those found in Kiley (2024). In addition to the magnitude of the estimated effect, Kiley (2024) uses statistical significance to support the relevance of his findings, so my finding of statistical insignificance is a direct refutation of his paper. However, lack of statistical significance does not necessarily mean a lack of importance (Ziliak and McCloskey 2008). In later sections of this paper I will show that reasonable alternative specifications lead to large differences in the oomph of the estimated effect of temperature on growth.

The Ziliak and McCloskey critique decries the use of statistical significance in economic research. McCloskey, however, has encouraged the use of robustness checks. McCloskey wrote the following, referring to historical economists:

He has in fact developed an art of creative self-doubt that is practiced in some other fields of economics and might be with profit practiced more widely. The habit of testing the sensitivity of one's argument to possible errors in its data or

possible mistakes in its analytical assumptions is widespread among scientists and historians, but it is not among economists. (McCloskey 1976, 445)

Other writers who are sympathetic to the McCloskey critique also emphasize the importance of robustness checking (e.g., van der Deijl 2023).

In this paper I examine the statistical significance of the results in Kiley (2024) because that is how they are presented in that paper. I also use robustness checks to show that the magnitude and even the sign of the effect are sensitive to alternative specifications.

Bootstrap standard errors

There is a formula to calculate the standard errors of quantile regression coefficients. The formula is different than the one for standard errors in OLS regression, but in both cases the standard errors can be calculated analytically, and they are calculated and displayed by quantile regression statistical packages. In the presence of heteroskedasticity, non-normality, and other data characteristics, however, these analytic standard errors will be inconsistent. Bootstrapping is a standard method of overcoming these difficulties (Davison and Hinkley 2009). In Kiley (2024) bootstrapping is clustered by country, and after running regressions on 200 synthetic samples the standard deviation of these estimates is taken as the standard error of the coefficient estimate. The bootstrap standard errors are much smaller than the analytic standard errors.

Kiley (2024) is correct to use bootstrapping to estimate standard errors, but there are at least two problems with Kiley's bootstrap estimation technique. First, clustering by country assumes that the observations of growth and temperature in each country are independent of each other. This is not the case, since temperatures and growth are similar in countries that are geographically near each other, and growth is correlated in countries that have trading relationships. Clustering by country when countries are not independent will tend to underestimate standard errors and inflate reported statistical significance (Cameron and Miller 2015).

The second problem with Kiley's bootstrap technique is that he assumes that the error associated with his coefficient estimates for each synthetic sample is the same. In doing so, he discards available information about each synthetic sample. This is particularly relevant because of the presence of extreme outliers in his data, which cause different synthetic samples to have different variances of growth. For example, if there is a large outlier in one country, then some bootstrapped samples will not contain that outlier at all, some will contain it once and others will contain it multiple times. This means that different bootstrap samples will have different variances. Kiley is combining estimates (calculating means) from different

samples, each of which has a different standard error. In this situation, it is best to combine studentized estimates (Cochran 1954; Hall 1992). Studentizing each bootstrap estimate requires an estimate of the standard error of each bootstrap estimate. This estimate can be obtained by using a double bootstrap to estimate the variance of each of the 200 estimates of the effect of temperature on growth (Hall 1992; McCullough and Vinod 1998; Davison and Hinkley 2009). In a double-bootstrap estimation, each synthetic sample is itself bootstrapped to find the variance of the estimate from that sample. With 200 outer bootstraps and 200 inner bootstraps for each outer bootstrap, a total of 40,000 quantile regressions are run for each percentile of interest. The standard error of the estimate of the effect of temperature on growth obtained using the double bootstrap is substantially higher than Kiley's estimate of the standard error, which is simply the standard deviation of the 200 bootstrap estimates. Table 1 shows the p-values reported in Kiley (2024, 1141) and those obtained using a double bootstrap to estimate the standard error. Using the double bootstrap method, the p-value for the 10th percentile is substantially higher, and the effect of temperature on growth is not statistically significant at the 5% level. For the 50th percentile, the p-value is also higher but still statistically significant. At the 90th percentile the p-value is also higher than Kiley's estimate and the effect of temperature on growth is not statistically significant.

TABLE 1. p-values of the effect of temperature on growth

	10th percentile	50th percentile	90th percentile
Kiley	0.00926	0.00018	0.05218
Double bootstrap	0.07023	0.01070	0.18490

Calculating p-values for each of the 200 bootstrap estimates for the 10th percentile, I find that the median p-value is 0.058, and that for more than a third of the bootstrap samples, the p-value is above 10%. These results suggest that something odd may be happening in the highest and lowest percentiles, perhaps heteroskedasticity and outlier observations. I will investigate these possibilities in later sections of this paper.

It is worth repeating that the primary result of Kiley's paper is that growth in the 10th percentile is adversely affected by higher temperatures. The double bootstrap result is that this effect is not statistically significant at the 5% level. Kiley reports that his main result is highly statistically significant, but it is not statistically significant at a conventional threshold.

Weighted regression

The double bootstrap results discussed above suggest that heteroskedasticity and/or outlier observations might be causing Kiley's finding of a statistically significant effect of temperature on growth. To further investigate this possibility, I weighted observations using a variety of methods. Weighted quantile regression is discussed by Roger Koenker (2005), but is not supported in the Stata procedure, *xtqreg*, used by Kiley (2024). I tried manually weighting observations using *xtqreg* and I used the weighting option that is available in the Stata procedure *qreg*. I weighted observations by the inverse of the standard deviation of growth by country, year and both country and year, and I also weighted observations by the conditional density evaluated at different percentiles.

Weighting by the inverse of the standard deviation by country using *xtqreg*

The usual way to deal with heteroskedasticity is to weight observations by the inverse of their standard deviations (Garson 2013). The idea is to give more importance to observations that are measured more precisely. Differences in results between weighted and unweighted regression can indicate that the unweighted results are driven by observations or groups of observations that are poorly measured. As Koenker (2005) points out, in quantile regression it is more appropriate to calculate weights that are tailored to the percentile that is being evaluated, and I will attempt this in a later section. It seems reasonable, however, to check to see if the simple method of weighting by the inverse of the standard deviation of growth by country has an effect on Kiley's results.

The *xtqreg* procedure handles country fixed effects automatically, and does not have an option to weight them, so in this section I weight all other variables, but not the fixed effects. In other words, because of how the *xtqreg* procedure works, I was not able to weight the entire observation. I weighted all of the observation except for the fixed effect dummy variables. A test using the same weights for OLS regression showed that the results are similar if all variables including country fixed effects are weighted and if all variables except country fixed effects are weighted. Of course, it is possible that quantile regression reacts differently to fixed effects not being weighted, which is a reason that I also estimate the regressions using other quantile regression packages.

After obtaining the coefficients of the weighted quantile regressions, I calculated the effect of temperature at the 75th percentile of weighted temperatures, and at the highest weighted temperature in the sample. Standard errors were calculated using Kiley's bootstrap method.

TABLE 2. Quantile regressions weighted by standard deviation of growth by country and year

	1	2	3	4	5	6	7	8	9
Weighted by sd of country									
At 75th percentile weighted temp									
Effect	0.441	0.396	0.365	0.338	0.315	0.294	0.274	0.249	0.212
se	0.251	0.198	0.162	0.138	0.135	0.121	0.118	0.132	0.153
P	0.078	0.045	0.024	0.014	0.020	0.015	0.020	0.059	0.165
At maximum weighted temp									
Effect	-0.305	-0.264	-0.235	-0.210	-0.189	-0.169	-0.150	-0.127	-0.092
se	0.195	0.174	0.145	0.127	0.116	0.110	0.110	0.108	0.134
P	0.116	0.130	0.106	0.099	0.105	0.125	0.171	0.241	0.490
Weighted by sd of year									
At 75th percentile weighted temp									
Effect	0.180	0.161	0.149	0.138	0.128	0.119	0.110	0.099	0.083
se	0.217	0.177	0.158	0.130	0.122	0.117	0.108	0.158	0.183
P	0.408	0.363	0.347	0.290	0.292	0.311	0.310	0.531	0.652
At maximum weighted temp									
Effect	-0.005	-0.003	-0.002	0.000	0.001	0.002	0.003	0.004	0.006
se	0.134	0.107	0.097	0.079	0.072	0.072	0.069	0.097	0.125
P	0.968	0.977	0.987	0.997	0.991	0.978	0.965	0.964	0.960
Weighted by sd of country and year									
At 75th percentile weighted temp									
Effect	0.266	0.253	0.245	0.239	0.233	0.228	0.222	0.216	0.206
se	0.141	0.112	0.100	0.113	0.104	0.123	0.123	0.146	0.152
P	0.059	0.024	0.014	0.034	0.025	0.064	0.072	0.140	0.176
At maximum weighted temp									
Effect	0.143	0.143	0.143	0.143	0.143	0.143	0.143	0.143	0.143
se	0.107	0.086	0.076	0.088	0.080	0.095	0.095	0.114	0.119
P	0.183	0.097	0.061	0.102	0.073	0.130	0.131	0.207	0.227

Weighting observations by country, year, and both country and year switched the sign of the effect of temperature on growth in most of the specifications. In fact, the positive effect of temperature on growth is statistically significant at the 10th percentile and at the median when weighting by country, and when weighting by both country and year. At the very warmest weighted temperature in the sample, the effect was negative but statistically insignificant for the 10th and 90th percentiles with weighting by country, and barely significant at the 10% level for the median regression. Weighting by year, the effect was statistically insignificant for all three percentiles, and weighting by both country and year at the warmest temperature the effect was positive, barely significant at the 10% level for the median regression, and statistically insignificant for the 10th and

90th percentiles. These results are shown in Table 2.

Weighting by the inverse of the standard deviation by country using *qreg*

The *xtqreg* procedure in Stata optimizes estimation for panel data, but it does not allow weighting. The *qreg* procedure allows weighting, but is not optimized for panel data. However, manually adding country fixed effects using *qreg* produces results that are similar to those using *xtqreg*. Table 3 compares the results.

TABLE 3. Comparison of basic results using *xtqreg* and *qreg*

	1	2	3	4	5	6	7	8	9
<i>xtqreg</i>									
Effect	-1.900	-1.681	-1.540	-1.412	-1.300	-1.198	-1.098	-0.980	-0.802
se	0.754	0.635	0.485	0.404	0.362	0.368	0.346	0.392	0.458
P	0.012	0.008	0.001	0.000	0.000	0.001	0.002	0.012	0.080
<i>qreg</i>									
Effect	-1.407	-1.048	-0.666	-0.731	-0.762	-0.557	-0.561	-0.330	-0.520
se	0.552	0.328	0.324	0.267	0.237	0.238	0.244	0.284	0.429
P	0.011	0.001	0.040	0.006	0.001	0.019	0.022	0.245	0.225

The results using *qreg* are somewhat weaker but similar to those using *xtqreg*, so it is reasonable to explore weighted quantile regression using the *qreg* procedure. Table 4 shows the results from weighting by the inverse of the standard deviation of growth by country, year, and both year and country. The results are similar to those in Table 2.

TABLE 4. Quantile regressions weighted by standard deviation of growth using *qreg*

	1	2	3	4	5	6	7	8	9
Weighted by sd of country									
Effect	-1.117	-0.790	-0.544	-0.655	-0.718	-0.547	-0.474	-0.288	-0.351
se	1.698	1.137	1.078	0.875	0.830	0.772	0.827	0.967	1.380
P	0.511	0.487	0.613	0.454	0.387	0.479	0.566	0.765	0.799
Weighted by sd of year									
Effect	-1.312	-0.826	-0.653	-0.599	-0.654	-0.606	-0.370	-0.381	-0.434
se	2.836	1.700	1.476	1.281	1.111	1.157	1.214	1.538	2.176
P	0.644	0.627	0.658	0.640	0.556	0.600	0.760	0.805	0.842
Weighted by sd of country and year									
Effect	-0.973	-0.749	-0.384	-0.572	-0.634	-0.552	-0.426	-0.306	-0.376
se	8.784	5.312	5.070	4.253	4.017	3.866	4.210	4.712	6.572
P	0.912	0.888	0.940	0.893	0.875	0.886	0.919	0.948	0.954

Weighting by the conditional density of growth using *qreg*

Table 5 shows the result of weighting observations using the principles outlined in Koenker (2005). To calculate weights for the n th percentile, that percentile of growth is first calculated from the data. Next, for each observation, a density function for growth using only observations with temperatures within five degrees of the temperature of the observation is constructed, and then the density function is evaluated at the relevant percentile of growth. This value is the weight for that observation.

TABLE 5. Quantile regressions weighted by conditional density using *qreg*

	1	2	3	4	5	6	7	8	9
Weighted by conditional density									
Effect	-1.308	-1.108	-0.610	-0.583	-0.696	-0.543	-0.266	-0.312	-0.372
se	19.070	11.910	12.252	10.054	8.439	9.405	9.332	10.645	15.607
P	0.945	0.926	0.960	0.954	0.934	0.954	0.977	0.977	0.981

These results indicate that using a variety of weighting methods to deal with possible heteroskedasticity and outlying observations increases the standard error of the effect measured in Kiley (2024).

The large standard errors associated with the effect of temperature on growth suggest that Kiley's estimates might be unreliable, and that further investigation of the effect of data characteristics such as outliers and heteroskedasticity is warranted.

Difference between 10th and 90th percentile

Kiley (2024) claims that a quantile regression for the tenth percentile of GDP growth on temperature shows a larger effect for warm countries than the quantile regression for the ninetieth percentile. Kiley (2024) shows the point estimates and standard errors for these regressions, but he does not show a test of the statistical significance of the difference. The result of a bootstrap test of this difference is shown in Table 6.

TABLE 6. Test of the difference between 10th and 90th percentile effect of temperature

	Effect	Standard error	p-value
10th percentile	-1.900	0.754	0.012
90th percentile	-0.802	0.419	0.055
Difference	-1.098	0.931	0.238

Kiley (2024) uses tests of statistical significance to establish that his results are important. He also uses bootstrap techniques that are easily adapted to test the statistical significance of his major result, which is that there is a difference between the effect of temperature on growth at the 10th and 90th percentiles. Kiley (2024), however, does not report this test.

Tables 1–5 show that the effect of temperature on growth at the 10th percentile is statistically insignificant, and Table 6 shows that the difference between the 10th and 90th percentiles is also statistically insignificant. Thus, these findings show that by the standard that Kiley (2024) uses to establish the importance of his results, statistical significance, his conclusions are unwarranted.

Data

Many characteristics of the data Kiley uses make accurate estimation difficult. In addition to containing extreme outlier observations, the data are multicollinear, autocorrelated, and heteroskedastic. These characteristics could cause spurious results. Of course, all economic data are imperfect, but Kiley's data are particularly flawed.

Autocorrelation in panel data can be evaluated using the Wooldridge test (Wooldridge 2002). For the residuals of Kiley's OLS specification, this test rejects the null hypothesis of no autocorrelation with an F statistic of 27.1. This value rounds to 27.1 with or without temperature and temperature squared in the regression. The test statistic for GDP growth itself is 28.1. The critical value at a 99% level of confidence is 6.8. The F statistic for a Wooldridge test on temperature is 266.5. These results suggest that an autocorrelated variable is being regressed on an autocorrelated variable, which means that Kiley's results may be spurious (Granger 1974).

Kiley's data are also spatially autocorrelated. A regression of GDP growth on GDP growth in the nearest 5 countries shows a strong relationship among contemporary growth rates in neighboring countries, even with fixed effects for years. Temperature shows even stronger relationships. The result is not surprising, since the same weather patterns can affect neighboring countries, and neighboring countries trade with each other and are affected by regional issues. This spatial autocorrelation could also produce spurious results.

Multicollinearity can be indicated by the Variance Inflation Factor (VIF). A VIF over ten is generally considered to suggest cause for concern over multicollinearity (Forthofer et al. 2007). Using Kiley's OLS specification, the VIFs for temperature and temperature squared are 253 and 252, more than twenty-five times the threshold for concern. The VIFs for year dummy variables from

2004–2010 are all over ten, and the VIFs for the other control variables are nearly all over thirty. Another way to illustrate multicollinearity is to regress temperature on all of the other independent variables in the model other than temperature squared. The R-squared value for this regression is 99.7%, indicating that temperature is highly correlated with the other independent variables.

Kiley's data are also heteroskedastic by country and by year. Growth has the lowest variance in Australia, with a standard deviation of 1.8, while growth in Equatorial Guinea has a standard deviation of 20.7. The standard deviation of growth in 2005 was 3.0, while in 1997 it was 10.2.

Kiley (2024) addresses heteroskedasticity by standardizing the variance of his observations by country, and finds that his results are unaffected. He makes no mention of autocorrelation or multicollinearity. I will discuss this standardization in a later section.

A variety of tests indicate that both GDP growth and temperature in Kiley's panel data contain unit roots, which indicates that spurious results are likely. I first calculate the Levin-Lin-Chu (LLC) test for a unit root in panel data (Levin et al. 2002) which takes account of cross-sectional correlation. The test statistic for growth is 4.54, much larger than the critical value of -1.65 for a one-tailed test, meaning that the null hypothesis of the panel containing unit roots cannot be rejected. For temperature, the test statistic is 8.99, even more strongly indicating failure to reject the presence of a unit root. The failure to reject occurred using a variety of lag structures and testing methods. The null hypothesis of the LLC test is that all panels contain unit roots. Even if this null hypothesis could be rejected, it is possible that enough panels contain unit roots to cause spurious regression results. The Hadri test (Hadri 2000) has as a null hypothesis that all panels are stationary. This hypothesis is rejected, again using a variety of lag structures and testing methods for both growth and temperature. The test statistic for growth is 6.71, with a 95% level of confidence critical value of 1.65, indicating a strong rejection of the null hypothesis. For temperature the test statistic is 12.15. Both tests require balanced panels, so I only used countries with 50 years of data, which eliminated 38 out of 124 countries.²

Kiley's revisions

Much of Kiley (2024) is identical to Kiley (2021). It contains additional discussion, and a scatterplot is added that is very similar to a scatterplot in Barker (2023a). There is, however, an important addition; section 3.3, titled "Robustness

2. This discussion is similar to a section in Barker 2024.

checks.” In that section, Kiley (2024) does three things: he experiments with leaving out controls for country-specific quadratic trends, he uses detrended per capita GDP as the dependent variable, and he attempts to deal with outlier observations.

Country-specific quadratic trends

Kiley’s quantile regressions include controls for quadratic time trends. Kiley (2024) acknowledges that Newell et al. (2021) finds that results similar to Kiley’s are sensitive to the inclusion or exclusion of the quadratic trend control variables. Kiley says:

Newell et al. (2021) highlight how coefficient estimates are generally much smaller and/or imprecisely estimated if the quadratic time trends included in Table 3 are not included in least squares regressions. (Kiley 2024, 1143)

The statement in Newell et al. (2021) is stronger than this. The authors say the following:

Such parametric trends can result in over-fitting, as we demonstrate they do in this setting. ... Models...that are saturated with fixed effects...may also absorb variation necessary to identify some relationships. In the present context, saturation of fixed effects or parametric time trends can both lead to this problem. (Newell et al. 2021, 6)

A footnote adds that:

Removing the country-specific quadratic time trends from BHM [Burke et al. 2015]’s specification as well as adding them to DJO [Dell et al. 2012]’s specification changes the sign of the estimated impacts of warming on GDP in 2100. (ibid., 6 n.18)

Newell et al. also says:

Theory offers little guidance in controlling for trending unobservables, and the extant literature appears to take a fairly *ad hoc* approach to modeling trend heterogeneity. (ibid., 5)

Kiley (2024) shows the results of excluding the country-specific time trends in his Table 4. Doing so eliminates the statistical significance of his results and reverses the direction of the effect over the percentiles. Kiley’s attempted solution is to add back a simpler time-trend control, which is a dummy variable indicating

whether data is from after 1990 interacted with country fixed effects. In other words, he substitutes a simple country-specific time trend in which average growth differs before and after 1990. Adding back this time trend restores Kiley's results. But doing so begs the question of whether a country-specific time trend should be included at all, whether it is quadratic or a simple step function.

The reason Kiley (2024) gives for including a post-1990 country-specific dummy variable is that

many countries have experienced demographic transitions, shifts in policy regime, or changes in productivity trends that may affect economic growth at a low frequency. (Kiley 2024, 1143–1144)

Table 7 shows the results of using a two-step procedure to estimate the effect of temperature on growth. GDP growth is regressed on a dummy variable indicating whether the data are from before or after 1990 multiplied by country fixed-effects variables. The residuals from this regression are simply GDP growth minus the mean of growth for the country and time period (pre- or post-1990) of each observation. These residuals are then regressed against temperature and temperature squared. The results show that simply adjusting growth by pre- and post-1990 mean and using adjusted growth as the dependent variable eliminate Kiley's results.

TABLE 7. Regressions of growth with post-1990 adjustment on temperature

	OLS	1	2	3	4	5	6	7	8	9
Kiley										
Effect	-1.232	-1.698	-1.518	-1.400	-1.301	-1.213	-1.132	-1.050	-0.957	-0.824
se	0.398	0.652	0.530	0.464	0.422	0.394	0.382	0.384	0.402	0.453
P	0.001	0.012	0.004	0.003	0.002	0.002	0.003	0.006	0.017	0.069
Two-stage										
Effect	-0.013	-0.850	-0.889	-0.912	-0.931	-0.947	-0.963	-0.978	-0.999	-1.029
se	0.025	0.752	0.532	0.447	0.390	0.335	0.299	0.305	0.313	0.36
P	0.597	0.258	0.095	0.041	0.017	0.005	0.001	0.001	0.001	0.004

Another way to describe this procedure is that the mean of early- and late-period GDP growth is calculated for each country, and then that mean is subtracted from each observation before running Kiley's percentile regressions without the whether-post-1990 dummy variable. This dummy variable is not needed because the observations have already been adjusted for pre- and post-1990 mean differences. There is no strong theoretical justification for one method or the other, but the results are different. In one, the effect of temperature in the lowest percentile is statistically significant and the effect is weaker in higher percentiles,

and in the other the effect of temperature in the lowest percentile is statistically insignificant, and the effect is stronger in higher percentiles. Kiley's headline result is that the effect of temperature is strongest in the lowest 10th percentile of growth, but this result is not robust to a different model specification.

The magnitude of the effect of temperature on growth in the 10th percentile is cut in half, and it is almost eliminated in the OLS specification.

Later on I use simulated data to illustrate that time trends can cause spurious results. In the simulated data using Kiley's estimation method there is an apparently large and statistically significant result, but when time trends are not included, there is no statistical or practical relationship between growth and temperature. When detrended growth is used as the dependent variable the statistical significance of the effect of temperature on growth is eliminated, whether growth is detrended using quadratic trends or a simple two-period mean adjustment.

Detrended GDP

One of the robustness checks in Kiley (2024) is to use growth in detrended GDP instead of growth in unadjusted GDP as the dependent variable in quantile regressions. Kiley's results survive this adjustment because he includes quadratic time-trend control variables in his regressions. However, these results have the same problem I described in the previous section, which is that when the quadratic time-trend control variables are not included, the results disappear. Including quadratic control variables has even less justification in this case because the dependent variable has already been detrended. Even if I calculate growth and detrend that series and then use it as the dependent variable, the effect of temperature on growth is large and statistically significant only if the quadratic time-trend control variables are included.

The control variables in Kiley (2024) are designed to capture country-specific trends in per capita GDP growth. In other words, Kiley (2024) assumes that there are quadratic trends in GDP growth that differ by country. He controls for them because these trends might be independent of temperature changes. For example, one country might have implemented pro-growth policies that led to steady or even exponential increases in the GDP growth rate over time. Other countries might have had series of mishaps that caused decreasing growth. Other countries might have had trends that reversed during the time period of the sample. Kiley's control variables allow for these possibilities. But what if temperature also has quadratic trends or additional, unobserved variables are correlated with both growth and temperature? If this is the case, then Kiley's parameter estimates may be inefficient. A reasonable robustness test would be to detrend GDP growth using the same control variables as in the original regression, and then use detrended GDP growth

as the dependent variable in a regression on only temperature and temperature squared. Including the quadratic time-trend control variables in the regression is unnecessary because the dependent variable has already been detrended using these same quadratic time-trend control variables.

Table 8 shows the results of this robustness check along with Kiley’s results. Kiley’s results are shown in rows 2–4 of Table 8, and “Effect” is the derivative of the dependent variable with respect to temperature. Since temperature is a quadratic function, the derivative is the coefficient on temperature plus two times the coefficient on temperature squared multiplied by temperature. Kiley (2024) evaluates this derivative for warm countries, defined as the 75th percentile of average temperature, which is 25.64 degrees Celsius. So “Effect” is the total effect of an extra degree of temperature on growth for a warm country on the indicated percentile. The column labeled “5” represents the effect on the median of GDP growth, while the column labeled “1” represents the effect on the 10th percentile. In other words, if temperature increases by one degree, the median growth rate decreases by 1.3 percentage points, and the 10th percentile decreases by 1.9 percentage points. Both of these decreases are statistically significant at a 95% confidence level, while the decrease of 0.8 percentage points of the 90th percentile is not statistically significant. Kiley’s main result is that the effect of temperature in the 10th percentile is more than twice the effect in the 90th percentile.

Rows 6–8 of Table 8 show the results of using the two-stage procedure described above. Growth is regressed on the control variables, and the residual from that regression becomes the dependent variable in a regression on temperature and temperature squared. The sign of Kiley’s key result reverses, with the effect of temperature larger in the 90th percentile than in the 10th percentile. The effect of higher temperatures is actually positive for the 10th percentile, although not statistically significant. The OLS results, which Kiley (2024, 1140) describes as “a baseline with which to compare the full set of quantile regressions,” show that the effect of temperature is small and statistically insignificant.

TABLE 8. Regressions of growth and detrended growth on temperature

	OLS	1	2	3	4	5	6	7	8	9
Kiley										
Effect	-0.257	-1.900	-1.681	-1.540	-1.412	-1.300	-1.198	-1.098	-0.980	-0.802
se	0.090	0.754	0.635	0.485	0.404	0.362	0.368	0.346	0.392	0.458
P	0.004	0.012	0.008	0.001	0.000	0.000	0.001	0.002	0.012	0.080
Two-stage										
Effect	-0.020	0.612	0.117	-0.180	-0.418	-0.621	-0.832	-1.037	-1.310	-1.749
se	0.025	0.458	0.272	0.253	0.209	0.202	0.207	0.240	0.290	0.377
P	0.409	0.182	0.667	0.477	0.045	0.002	0.000	0.000	0.000	0.000

In summary, one of Kiley’s robustness checks is to use detrended GDP instead of actual GDP to calculate his dependent variable. But he includes quadratic time-trend control variables in these regressions, which is odd since the dependent variable has already been detrended. Not including these quadratic time trend control variables reverses Kiley’s main result, and causes his baseline OLS result to be statistically insignificant.

Equation 1 is presented in Kiley (2024) to explain how he estimates the effect of temperature on detrended growth. Detrended growth is $y^{detrended}$, a_j is a constant, $A_D D$ is a set of dummy control variables multiplied by a set of constants, and $F(T(t, j))$ is a function of temperature. For Kiley’s primary estimation this function is quadratic.

$$\Delta y^{detrended}(t, j) = a_j + A_D D + b y^{detrended}(t - 1, j) + F(T(t, j)) \tag{1}$$

Without explanation, Kiley (2024) states that the long-run effect of temperature on the level of GDP is $\frac{F(T(t, j))}{1 + b}$. Kiley is correct, but this result is not immediately obvious. Kiley then says that the estimate of b is approximately -0.125 , “implying a long-run level effect about 8 times the reported impact effects” (2024, 1142).

Here is Kiley’s reasoning. Equation 1 can be rewritten as follows:

$$y^{detrended}(t, j) - y^{detrended}(t - 1, j) = a_j + A_D D + b y^{detrended}(t - 1, j) + F(T(t, j)) \tag{2}$$

$$y^{detrended}(t, j) = a_j + A_D D + (b + 1) y^{detrended}(t - 1, j) + F(T(t, j)) \tag{3}$$

If $y_0 = 1$, and if for simplicity we make $X = a_j + A_D D + F(T(t, j))$, then subsequent values of y will be:

$$y_1 = (b + 1) + X \tag{4}$$

$$y_2 = (b + 1)^2 + (b + 1)X + X \tag{5}$$

$$y_3 = (b + 1)^3 + (b + 1)^2 X + (b + 1)X + X \tag{6}$$

$$y_4 = (b + 1)^4 + (b + 1)^3 X + (b + 1)^2 X + (b + 1)X + X \tag{7}$$

$$y_n = (b + 1)^n + (b + 1)^{n-1} X + (b + 1)^{n-2} X + \dots + (b + 1)X + X \tag{8}$$

If $b = -0.125$, Equation 8 can be simplified to:

$$y_n = 0.875^n + (0.875^{n-1} + 0.875^{n-2} + 0.875^{n-3} + \dots + 1)X \quad (9)$$

As n tends to infinity, the first term in Equation 9 disappears, and the expression inside the parentheses becomes $\frac{1}{1-0.875}$ which is equal to 8. This means that long-run GDP is equal to 8 times $a_j + A_D D + F(T(t, j))$, which means that the component of GDP that changes with temperature is $8F(T(t, j))$, assuming that temperature does not affect the constant or the control variables. The effect of a unit change in temperature on GDP will be $8\frac{dF(T(t, j))}{dT}$, which is 8 times the effect shown in Table 8.

The important point to grasp is that if the magnitude of the effect cannot be distinguished from zero, then the entire effect cannot be distinguished from zero. The results shown in Table 8 suggest that, indeed, the magnitude of the effect cannot be distinguished from zero. The effect measured by the parameter b that Kiley (2024) describes does not change this conclusion.

Outliers

In Barker (2023a) I showed that dropping a small percentage of Kiley's observations eliminated the statistical significance of the effect of temperature in the lowest percentile. Without citing Barker (2023a), Kiley revised his original working paper to include a section on outlier observations. Kiley (2024) identified 44 out of his original 124 countries with at least one year of growth more than three standard deviations from the global mean. He deleted these countries as a robustness check for the effect of outliers and found that his results strengthened, with a p-value of 0.001 for the 10th-percentile quantile regression compared to 0.009 for the entire sample. Interestingly, using a four-standard-deviation cutoff yields a p-value of 0.021—still statistically significant but weaker. There is no theoretical justification for using a threshold of three instead of four standard deviations, but using three gives the impression that removing outliers strengthens Kiley's result. In reality, removing only the most extreme outliers weakens the result, while removing some less extreme outliers strengthens it. Reporting only the test most favorable to his results suggests that he may have chosen a threshold for deletion that maximizes statistical significance.

To find true outliers, I use the Political Instability Task Force database of state failures to identify years in countries with major political upheaval (PITF 2019). Like Kiley (2024), I identified extreme observations and deleted all observations from the countries in which they occurred. PITF (2019) lists wars, revolu-

tions, coups and genocides and rates them on several criteria such as the number of combatants and fatalities, the portion of the area of the country that is involved in conflict, the magnitude of failure of state authority, and the overall magnitude of violence. Each of these ratings is on a scale of 1 to 4 to indicate the severity of the event.³ I identified 46 countries experiencing an event with any rating of 4 during the sample period. In order to match Kiley (2024), in which 44 countries were eliminated, I added back two countries. I identified these two countries by calculating the total rating from all scores for each year in each country, then finding the year in each country that had the highest total rating. The two countries added back were those with the lowest maximum total rating. In other words, of countries with an event rated at 4, these two had the least-bad worst years. Using the remaining 80 countries, the statistical significance of the effect of temperature on growth for the lower percentiles is eliminated, as shown in the first panel of Table 9.

One country in Kiley's sample, Greenland, stands out as a climate outlier. Sorting all countries by average temperature and taking the difference between each country and the next warmest country, the difference for Greenland is nearly 10 times the standard deviation of these differences. The next highest difference is less than four standard deviations of these differences, and the next highest difference is close to 2. Greenland's climate is very different from the warm countries on which Kiley (2024) bases his conclusions. The second panel of Table 9 shows results from replacing Greenland with the next-lowest rating of political upheaval so that the number of countries in the sample is still 80. The sign of the effect changes to positive for the 10th-percentile quantile regression.

TABLE 9. Quantile regressions removing countries with major political upheaval

	1	2	3	4	5	6	7	8	9
Dropping 44 countries with major events									
Effect	-0.169	-0.436	-0.603	-0.756	-0.909	-1.037	-1.173	-1.334	-1.568
se	0.538	0.473	0.439	0.419	0.434	0.428	0.395	0.525	0.542
P	0.753	0.357	0.170	0.071	0.036	0.015	0.003	0.011	0.004
Dropping 43 countries with major events plus Greenland									
Effect	0.063	-0.252	-0.453	-0.638	-0.820	-0.979	-1.141	-1.338	-1.616
se	0.547	0.458	0.427	0.436	0.471	0.444	0.482	0.523	0.617
P	0.909	0.583	0.289	0.143	0.082	0.027	0.018	0.011	0.009

The upper-percentile quantile regressions show a statistically significant effect of temperature on growth, but in Kiley's robustness check the magnitude of the effect in the 90th percentile is cut in half, and the p-value is 0.279. This is

3. One variable in the database, *deathmag*, is based on annual deaths in genocides, and is on a scale of 1–5. I subtracted one from *deathmag* with a floor of zero to make it compatible with the other variables.

in part due to the fact that several countries in the sample experienced enormous GDP growth in years when major oil discoveries were made. For example, per capita GDP growth in Equatorial Guinea in 1997 was 88.4 percent when oil began flowing from the Zafiro field, which was discovered in 1995. Equatorial Guinea had no events in the PITF (2019) database, and so it is included in the quantile regressions shown in Table 9, but is not included after Kiley (2024) deletes countries with growth more than three standard deviations from the mean.

Even after Kiley (2024) eliminates countries with instances of growth that are three or more standard deviations from global mean growth, many observations remain that appear to have limited value in an investigation of the effect of temperature on GDP growth. For example, it is difficult to justify including Syria in 1966, when a coup occurred in February, before any hot weather recorded in Kiley's data for that year occurred, causing a loss of 11.3 percent of GDP. Kiley (2024) included the observation because the loss was less than three times the standard deviation of growth in his sample, or 17.4 percent.

When the influential observations I identified in Barker (2023a) are not included in the sample, Kiley's results disappear. To verify that these observations are truly driving the results, I created predicted GDP growth rates based on a regression of growth on all of Kiley's independent variables other than temperature and temperature squared. I added random noise that matched the country-specific standard deviations of the residuals from that regression. By construction, these simulated growth rates have no relationship with temperature, but they closely match actual growth rates. Kiley's regressions run on these data show no results, as expected. But if the actual growth rates from these same 18 countries are substituted in for growth, then Kiley's results reappear. Just as in Kiley's sample, if the controls for quadratic time trends are eliminated, the results disappear.

Kiley (2024) standardizes growth in another set of quantile regressions in order to reduce the possible effect of heteroskedasticity and outliers. He divides growth by the standard deviation of growth by country and subtracts the mean, so that in each country the standard deviation of growth, the dependent variable, is one and its mean is zero. Kiley's results survive this robustness check. Standardizing the dependent variable, however, is only half of the usual correction for heteroskedasticity. In a weighted regression, the entire observation, dependent and independent variables, are divided by the standard deviation of the dependent variable. This is what I did in Tables 2 and 4 above, and Kiley's results did not survive. Having weighted the dependent variable by the inverse of the standard deviation of growth by country, Kiley surely tried the same weighting of both the dependent and independent variables, finding, as I report in Tables 2 and 4, that his results are eliminated. Similar to the Brady Rule for prosecutors, perhaps researchers should be required somehow to include in their publications evidence

that they find against their hypotheses.

To more clearly illustrate how outliers affect Kiley's model, I reduced the sample to five countries for which the model shows the same basic results as it does for the full sample. Using Bolivia, Bulgaria, Portugal, Rwanda, and the United States, the model shows a strong negative effect of temperature on the 10th percentile of growth at 25.6 degrees, and a less negative effect on the 50th and 90th percentiles. In addition, just as in the larger sample, the analytic standard errors show a statistically insignificant effect, while a clustered bootstrap procedure shows a much larger and statistically significant effect. The results are shown in Table 10.

Rwanda contains a large outlier of growth in 1994, the year of the Rwandan genocide. Removing that observation eliminates the statistical significance of temperature on growth. It is worth noting that the warm weather of 1994 occurred after the genocide took place.

TABLE 10. Results using only Bolivia, Bulgaria, Portugal, Rwanda and the United States

	All observations			Without one outlier		
	10th	Median	90th	10th	Median	90th
Estimate	-30.900	-13.998	-3.590	-1.602	-2.108	-2.474
p-value, BS	0.011	0.068	0.583	0.739	0.672	0.647
p-value	0.705	0.680	0.952	0.902	0.648	0.686
Observations	230	230	230	229	229	229

Other issues

Autocorrelation

As discussed above, GDP growth and temperature are both autocorrelated over time and space. Regressing an autocorrelated variable on another autocorrelated variable can produce spurious results (Granger 1974). A simple way to check whether autocorrelation might be producing spurious results is to perform a Wooldridge test for autocorrelation in panel data on detrended and demeaned growth residuals, incorporating the independent variables temperature and temperature squared. Including both temperature and temperature squared yields an F-statistic of 27.3; including only temperature yields an F-statistic of 27.4. The critical value of the F-statistic at a 95% level of confidence is 3.9. The possibility of spurious results because of autocorrelation is clearly present, but in Kiley (2024) there is no mention at all of autocorrelation or lagged variables.

To further illustrate the possible consequences of autocorrelation, I included a single lag of GDP growth along with growth in the geographically nearest country in the quantile regressions makes Kiley's results more fragile. Removing only nine

observations out of the total 5,741 observations eliminates the statistical significance of the 10th percentile quantile regression. These results are shown in Table 11, along with Kiley's results for comparison. The effect of temperature on growth increases from the lower to the higher percentiles, the opposite of Kiley's result.

The nine observations were identified by using the OLS regression diagnostic statistic $dfbeta$, which stands for difference in beta values. This statistic can be calculated for each observation and represents the difference, as a fraction of a standard deviation of an OLS regression coefficient, in the estimated coefficient if an observation is removed. David Belsey et al. (2004) suggest a cutoff of $\pm 2/\sqrt{n}$ for this statistic, where n is number of observations. They suggest that observations above this cutoff should be examined to see if they may be affected by factors outside of the model being tested. Kiley's sample contains 5,741 observations, so the cutoff value is -0.0264 . The nine observations all had $dfbeta$ values for squared temperature of less than -0.1 , nearly four times the cutoff value.⁴ Dropping 29 observations reverses the sign of the effect in the 10th percentile.

TABLE 11. Regressions of growth including lagged GDP growth

	1	2	3	4	5	6	7	8	9
Kiley									
Effect	-1.698	-1.518	-1.400	-1.301	-1.213	-1.132	-1.050	-0.957	-0.824
se	0.652	0.530	0.464	0.422	0.394	0.382	0.384	0.402	0.453
P	0.012	0.004	0.003	0.002	0.002	0.003	0.006	0.017	0.069
With temporally and spatially lagged GDP growth, without nine influential observations									
Effect	-0.862	-0.871	-0.877	-0.881	-0.886	-0.890	-0.894	-0.899	-0.906
se	0.546	0.422	0.384	0.358	0.327	0.285	0.307	0.329	0.347
P	0.114	0.039	0.022	0.014	0.007	0.002	0.004	0.006	0.009

Growth and temperature in individual countries

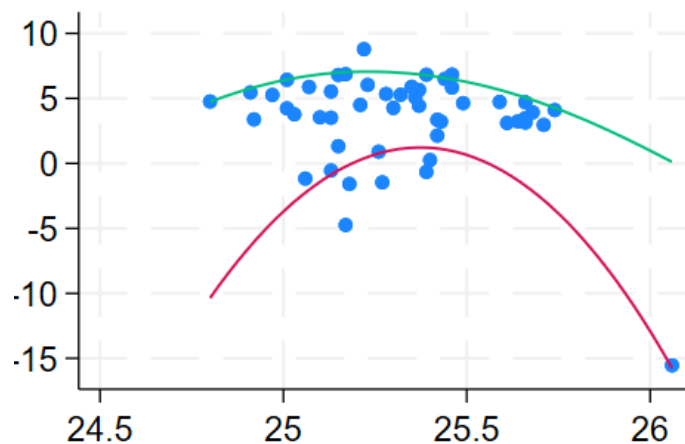
Running quantile regressions on all 124 countries in his sample, Kiley (2024) claims to find the following:

4. Four of the nine had level-4 events in the PITF (2019) database, and another, Niger in 1973, experienced a military coup that was not in the PITF (2019) database. Another was Thailand in 1998 following the Asian financial crisis, and another was Oman in 1977, a year in which the country's oil production fell by 7 percent (Ministry of Energy and Minerals 2024). The other two were from Greenland, which I discuss as an outlier country. In one of those years, 1990, a mine closed that had produced at levels representing 12.1 percent of Greenland's GDP (Barker (2024). In the other year, 1984, Greenland became the first territory to leave the European Economic Community since the Treaty of Rome established a common market in 1958 (New York Times, Feb. 4, 1985, Section d, 4). These nine observations were chosen because of their influence, not because of these events as was done for the results in Table 9, but it is worth noting that they all coincided with significant events unrelated to temperature that caused unusual economic performance.

1. A negative relationship between the 10th percentile of growth and temperature at 25.6 degrees Centigrade.
2. A negative relationship between the 90th percentile of growth and temperature at 25.6 degrees Centigrade.
3. An effect larger in absolute value at the 10th than at the 90th percentile.

I ran quantile regressions on each country individually to see how many produce these results. Only one of the 124 countries in Kiley's sample, Indonesia, produced all three. For many countries, all temperature observations are either above or below 25.6 degrees, so they do not meet the above conditions. (I did not extrapolate the effect to temperatures that are out of sample for a particular country.) The results for Indonesia are shown in Figure 1. The scatterplot shows annual observations of growth and temperature. The red line shows the estimate of quantile regression for the 10th percentile, and the green line shows the estimate for the 90th percentile. Each line is smoothed using local second-degree polynomial regression. This is done to isolate the effect of temperature without interference from the effects of the time trend control variables.

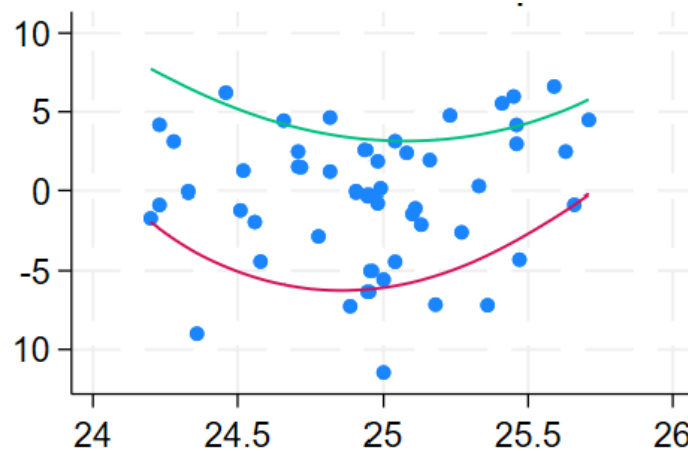
Figure 1. Scatterplot of Indonesia growth and temperature with predicted 10th and 90th percentiles



At 25.6 degrees, the downward slope of the red line is steeper than the green line, indicating a larger negative effect of temperature. From the scatterplot, it is obvious that this is due to a single point, which happens to represent the year 1998, the year of the Asian financial crisis. Per capita GDP growth was -15.5% that year in Indonesia. Temperatures in Asia were high that year, the highest of any year in Kiley's sample. No analysis of the 1998 financial crisis, however, discusses high temperatures as a cause or aggravating factor in the crisis. Without 1998 in

Indonesia, not a single country, analyzed individually, would produce Kiley's result. The patterns of temperature and growth are different for each country, and Figure 2 shows an example, Central African Republic, where the pattern is very different than that of Indonesia. In Central African Republic, GDP growth rises with temperature for both the 10th and 90th percentiles at a temperature of 25.6.

Figure 2. Scatterplot of Central African Republic growth and temperature with predicted 10th and 90th percentiles



The fact that no individual country shows the relationship between growth and temperature that Kiley hypothesizes is not conclusive proof that the hypothesis is wrong, but it does indicate that time-series variation within countries is insufficient to provide evidence of Kiley's effect. He is strongly relying on cross-sectional variation between countries for this evidence. But as Kiley points out:

The negative correlation between income and temperature has been observed for a long time (Montesquieu, 1750). (Kiley 2024, 1136)

Kiley's contribution is to add time-series data to the cross-sectional data so that we can get beyond Montesquieu's observation that many hot countries are poor. Kiley says that there is sufficient time-series variation to do so:

Nonetheless, the standard deviation of temperature within country from year-to-year is sufficient to be economically important. Generally, these standard deviations are on the order of 0.5°C–0.7°C. This magnitude of variation, with a two standard deviation move of 1°C to 1–1/2°C, is similar in magnitude to the anticipated increase in temperature associated with climate change in coming decades. (Kiley 2024, 1136)

The fact that none of the 124 countries in Kiley’s sample has time-series variation that demonstrates the effect he is trying to prove demonstrates that he is largely relying on the same cross-sectional variation that Montesquieu did. As Kiley (2024, 1136) and Melissa Dell et al. (2012, 66–67) point out, this cross-sectional variation has been insufficient to settle the long debate over a possible causal connection between temperature and economic well-being.

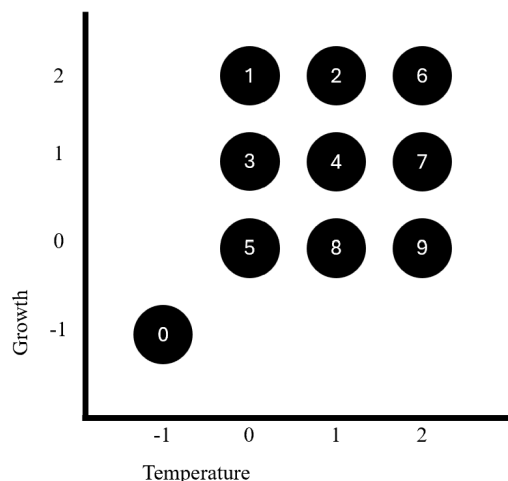
Illustrative simulations

Growth and temperature trends

Temperature in Kiley’s sample tends to increase over the sample period, and GDP growth tends to decrease. A simple regression of GDP growth on time, controlling for individual country averages, shows a negative trend with a p-value of 0.001. A similar regression of temperature on time shows a positive trend with a p-value very near zero. This is important because Kiley (2024) is measuring the effect of temperature on growth holding time trends in growth constant.

Let me create a simple analogy, to illustrate the problem. We consider ten data points for a single country. The data points are annual (temperature, growth) for that country, spanning ten years. In Figure 3, growth is on the y-axis and temperature is on the x axis. At time five, growth and temperature are equal to zero. At time six, growth and temperature are equal to two. At time zero, growth and temperature are equal to negative one.

Figure 3. An analogy: Ten annual data points for a country



The point at time one represents the first year of data, with growth of two and temperature of zero. The point at time nine represents the ninth year of data, with growth of zero and temperature of two. The point at time zero is an outlier observation.

Excluding the outlier, the remaining nine points are symmetric in growth-temperature space, so a regression of growth on temperature will show no relationship. But growth has a negative time trend and temperature has a positive time trend, and when time is added to the regression, the coefficient on temperature becomes large and statistically significant, falsely indicating a negative relationship between growth and temperature. Table 12 shows the results of regressing growth on temperature, regressing growth on temperature and time, and of taking the residuals from a regression of growth on time and regressing these residuals on temperature. The residuals represent detrended growth.

TABLE 12. Regressions using illustrative data, no outlier

Regression of...	Coefficient of Temp	Standard error of Temp	P-value for Temp
Growth, on Temp	0.000	0.378	1.000
Growth, on Temp and Time	0.885	0.190	0.003
Residual from Regression of Growth on Time, on Temp	0.469	0.211	0.061

Adding an outlier observation allows the same thing to happen with temperature squared. Neither temperature nor temperature squared are large or statistically significant by themselves, but when time is added to the regression, both are large and statistically significant at the 5% level. Using residuals from the regression of growth on time instead of growth reduces the statistical significance of temperature and temperature squared.

TABLE 13. Regressions using illustrative data, including outlier

Regression of...	Coefficient of...		Standard error of...		P-value for...	
	Temp	Temp ²	Temp	Temp ²	Temp	Temp ²
Growth, on Temp and Temp ²	1.000	-0.455	0.558	0.337	0.116	0.219
Growth, on Temp, Temp ² , and Time	1.899	-0.466	0.386	0.190	0.003	0.049
Residual from Regression of Growth on Time, on Temp and Temp ²	1.113	-0.456	0.508	0.306	0.065	0.180

In the illustrative data, there is no large or statistically significant relationship between growth and temperature. They both have independent time trends, however, and attempting to control for time creates a spurious relationship. This is similar to what happens in Kiley's model. When Kiley (2024) includes individual

country time trends in his regressions, temperature is statistically significant, but when these trends are not included, the results disappear. When the residuals from a regression of growth on time trends are regressed on temperature, the statistical significance is greatly reduced or eliminated, similar to the results in Table 12 and Table 13.

Quantile regression with simulated data

To illustrate how time trends and outliers can produce Kiley’s results using quantile regression with panel data, the simulated data needs to be a bit more complicated. I constructed data representing five countries, with 25 years of data for each country. Each country has a different mean temperature, and growth and temperature are functions of time, similar to the previous example. There is, however, no relationship between growth and temperature, except for a single outlier. For each country, a scatterplot of growth and temperature is a rectangle, just as in the previous example. Noise is introduced into the data, and each country has different functions of growth and temperature with respect to time. Figure 4 shows the simulated data.

Figure 4. Simulated data: Rectangles represent different countries; numbers indicate time

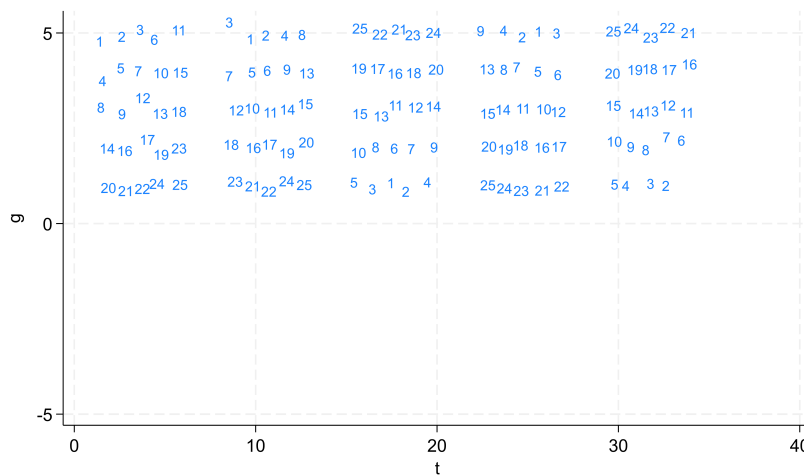


Table 14 shows the result of 10th-, 50th-, and 90th-percentile regression on these simulated data. With time-trend controls the effect of temperature appears to be large and statistically significant for the 10th percentile, and is less negative in the 90th percentile, just as in Kiley’s results. Also as in Kiley’s results, removing

the time-trend controls eliminates the statistical significance of the effect of temperature on growth. Eliminating the outlier observations eliminates the effect of temperature.

These results demonstrate that it is possible for data in which there is no true relationship between growth and temperature to show a relationship using Kiley's estimation method. It does not show that simulated data will always produce this result, since different random number seeds and different time trends will produce different results.

TABLE 14. Five countries simulated

With all time trend controls						
	All observations			Without outlier		
	10th	Median	90th	10th	Median	90th
Estimate	-0.171	-0.156	-0.133	-0.005	-0.001	0.003
p-value, BS	0.037	0.061	0.202	0.980	0.994	0.985
Obs	125	125	125	125	125	125
Without time trend controls						
	All observations			Without outlier		
	10th	Median	90th	10th	Median	90th
Estimate	-0.372	-0.349	-0.321	-0.029	-0.046	-0.073
p-value, BS	0.122	0.0568	0.044	0.804	0.571	0.450
Obs	125	125	125	125	125	125

Conclusion

The data analyzed in Kiley (2024) are hopelessly flawed for the purpose of statistical analysis. They are autocorrelated, heteroskedastic, and multicollinear, and they contain huge outliers. All economic data have these problems to some extent, but the problems are extreme in Kiley's data, and they are not adequately addressed. Kiley's model applied to these data produces unreliable coefficient estimates that predict gigantic effects of small temperature changes on growth. Kiley (2024) reports that his results are large and highly statistically significant, but they fail several robustness checks. Kiley (2024) cherry-picks a few robustness checks that his results survive, and, it seems, mainly ignores my previously published criticisms of his work.

Publication of unsound research is always damaging to the academic enterprise, but it might have few noticeable consequences if the topic is obscure, and the research has no policy implications. When it comes to the political

economy surrounding climate change, however, the issue is widely discussed and has enormous policy consequences. Given the stakes, the integrity of research on climate change is not just an academic concern but a societal imperative.

Kiley (2024) contains the standard disclaimer that “The views expressed herein are those of the author, and do not reflect those of the Federal Reserve Board or its staff.” However, a working paper version of Kiley (2024) was published by the Federal Reserve (Kiley 2021), and Kiley is an employee of the Federal Reserve, which has no definitive obligation to provide the kind of academic freedom offered by universities.

The weakness of Kiley’s results raises the question of why the Federal Reserve continues to publish such terrible research. Perhaps it believes that attaching itself to the climate change issue is a good political strategy for preserving its influence, resources, and independence. A more frightening possibility is that Federal Reserve economists believe that the research is sound, which raises the question of whether the core research of the Federal Reserve on economic and monetary issues is also of poor quality. A yet more frightening possibility is that the Federal Reserve is not concerned with the soundness of its research.

Data and code

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

References

- Barker, David.** 2022. Temperature and U.S. Economic Growth: Comment on Colacito, Hoffmann, and Phan. *Econ Journal Watch* 19(2): 176–189. [Link](#)
- Barker, David.** 2023a. Temperature and Economic Growth: Comment on Kiley. *Econ Journal Watch* 20(1): 69–84. [Link](#)
- Barker, David.** 2023b. Temperature Shocks and Economic Growth: Comment on Dell, Jones, and Olken. *Econ Journal Watch* 20(2): 234–253. [Link](#)
- Barker, David.** 2024. Global Non-Linear Effect of Temperature on Economic Production: Comment on Burke, Hsiang, and Miguel. *Econ Journal Watch* 21(1): 35–68. [Link](#)
- Belsey, David A., Edwin Kuh, and Roy E. Welsch.** 2004. *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. John Wiley & Sons.
- Burke, Marshall, Solomon M. Hsiang, and Edward Miguel.** 2015. Global Non-Linear Effect of Temperature on Economic Production. *Nature* 527(7577): 235–239.
- Cameron, A. Colin, and Douglas Miller.** 2015. A Practitioner’s Guide to Cluster-Robust Inference. *Journal of Human Resources* 50(2): 317–372.

- Cochran, William.** 1954. The Combination of Estimates from Different Experiments. *Biometrics* 10(1): 101–129.
- Davison, Anthony, and David Hinkley.** 2009 [1997]. *Bootstrap Methods and their Applications*. Cambridge University Press.
- Dell, Melissa, Benjamin F. Jones, and Benjamin A. Olken.** 2012. Temperature Shocks and Economic Growth: Evidence from the Last Half Century. *American Economic Journal: Macroeconomics* 4(3): 66–95.
- Forthofer, Ronald, Eun Sul Lee, and Mike Hernandez.** 2007. *Biostatistics: A Guide to Design, Analysis and Discovery*, 2nd ed. Elsevier Academic Press.
- Garson, David.** 2013. *Weighted Least Squares*. Raleigh, N.C.: Statistical Associates Publishers.
- Granger, Clive.** 1974. Spurious Regressions in Econometrics. *Journal of Econometrics* 2(2): 111–120.
- Hadri, Kaddour.** 2000. Testing for Stationarity in Heterogeneous Panel Data. *Econometrics Journal* 3(2): 148–161.
- Hall, Peter.** 1992. *The Bootstrap and Edgeworth Expansion*. Springer-Verlag.
- Kiley, Michael T.** 2021. Growth at Risk from Climate Change. *Finance and Economics Discussion Series* 2021-054. Board of Governors of the Federal Reserve System (Washington, D.C.). [Link](#)
- Kiley, Michael T.** 2024. Growth at Risk from Climate Change. *Economic Inquiry* 62(3): 1134–1151.
- Koenker, Roger.** 2005. *Quantile Regression*. Cambridge University Press.
- Levin, Andrew, Chien-Fu Lin, and Chia-Shang James Chu.** 2002. Unit Root Tests in Panel Data: Asymptotic and Finite-Sample Properties. *Journal of Econometrics* 108(1): 1–24.
- McCloskey, Deirdre N.** 1976. Does the Past Have Useful Economics? *Journal of Economic Literature* 14(2): 434–461.
- McCullough, Bruce, and Hrishikesh Vinod.** 1998. Implementing the Double Bootstrap. *Computational Economics* 12: 79–95.
- Ministry of Energy and Minerals.** 2024. Oil and Gas Statistics. Sultanate of Oman (Muscat, Oman). [Link](#)
- Newell, Richard, Brian Prest, and Steven Sexton.** 2021. The GDP-Temperature Relationship: Implications for Climate Change Damages. *Journal of Environmental Economics and Management* 108: 102445.
- Political Instability Task Force (PITF).** 2019. Political Instability Task Force (PITF) Consolidated Problem Set, Historical State Armed Conflicts and Regime Crises, 1955–2018. Center for Systemic Peace (Vienna, Va.). [Link](#)
- Van der Deijl, Willem.** 2023. *The Philosophy of Econometrics*. Open Press Tilburg University.
- Wooldridge, Jeffrey M.** 2002. *Econometric Analysis of Cross Section and Panel Data*. MIT Press.
- Ziliak, Stephen T., and Deirdre N. McCloskey.** 2008. *The Cult of Statistical Significance: How the Standard Error Costs Us Jobs, Justice, and Lives*. University of Michigan Press.

About the Author



David Barker taught economics and finance at the University of Chicago and the University of Iowa. His Ph.D. is from the University of Chicago and he worked as an Economist at the Federal Reserve Bank of New York. He currently runs a real estate and finance company in Iowa and is a member of the Iowa Board of Regents. His email address is drb@barkerapartments.com.

[Go to archive of Comments section](#)

[Go to September 2024 issue](#)