Economics from the Top Down

new ideas in economics and the social sciences

Roger Pielke Jr.'s Appallingly Bad Analysis of Billion Dollar Disasters

Blair Fix

October 26, 2025

According to a simple linear trend, losses per disaster are down by about 80% since 1980, as a proportion of GDP.

— Roger Pielke Jr.

In the world of scientific disinformation, Roger Pielke Jr. is a well known player. A political scientist by training, Pielke has a long history of being a thorn in the side of climatologists who study natural disasters.[1]

Pielke's latest entry in this genre is a 2024 paper called 'Scientific integrity and U.S. "Billion Dollar Disasters"'. The paper takes aim at the 'billion-dollar disasters' dataset run by climatologists at the NOAA. As the name suggests, the database tracks the cost of US weather and climate-related disasters which have inflation-adjusted losses that exceed $1 billion. (Or rather, the database *tracked* these costs. The billion-dollar-disasters database was recently cancelled by the Trump regime. Afterwards, Pielke took to his blog to celebrate.)

Now, my goal here is not to defend the billion-dollar-disasters dataset from Pielke's criticism.[2] Instead, my aim is to show that Pielke's analysis is so flawed that it undermines his own appeal 'scientific integrity'. For his part, Pielke claims that putting climatologists in charge of disaster loss estimation

[1] To get a sense for Pielke's history, see his entry on the DeSmog Climate Disinformation Database

[2] As I see it, Pielke's most substantial criticism of the billion-dollar-disasters data is that the database has opaque sources and methods. It's a fair point. That said, poor documentation is actually the status quo for *many* government databases.

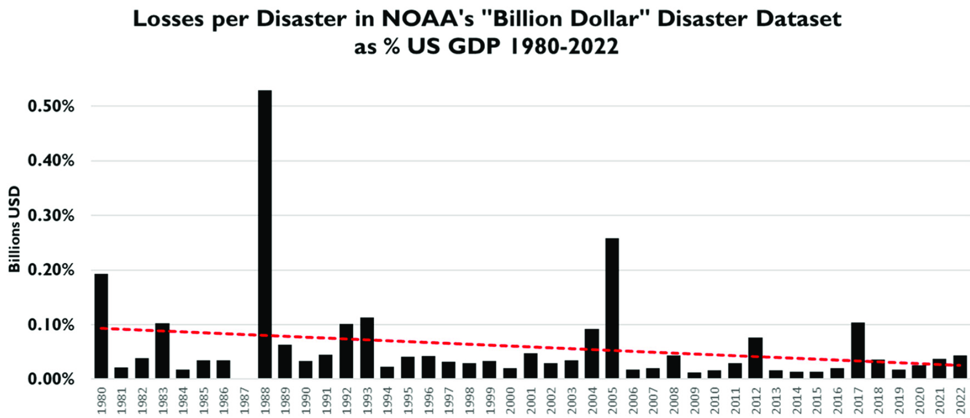


Figure 1: Roger Pielke Jr's analysis of the billion-dollar-disasters dataset

Source: 'Scientific integrity and U.S. "Billion Dollar Disasters"'.

---

is 'problematic', and that the job would be better left to 'proper economists'. Furthermore, Pielke argues that the billion-dollar-disasters dataset is so faulty that it violates the NOAA's own standards on 'scientific integrity'. Yet while Pielke sits on this high horse, he manages to so horribly botch his own analysis that one wonders if he is unintentionally writing satire.

In what follows, I'll spend a whole essay unpacking and debunking a single chart. Figure 1 shows Pielke's published analysis of the billion-dollar-disasters dataset. The graph seems to show a steady decline in average disaster costs as a share of US GDP. The implicit message is that when climatologists warn about worsening natural disasters, they're overreacting. If anything, economic growth seems to be making disaster costs more trivial. Or so Pielke claims.

On this front, the NOAA is somewhat unique in that it's run by actual scientists who have adopted an explicit policy on 'scientific integrity'. We should celebrate such standards. But if we want to be fair in our criticism, we should apply these standards not just to the billion-dollar-disasters dataset, but to *any* data we invoke in our corresponding analysis.

For his part, Pielke is happy to criticize the billion-dollar-disasters data at the same time that he unquestioningly uses data for 'real' GDP. Perhaps he is unaware that the divination of 'real' GDP is a scientific nightmare, with methodological problems that far exceeds any issues with the disaster data.

In reality, Pielke's chart reveals something quite different. It demonstrates that he does not understand the data he purports to analyze. You see, the trend line in Figure 1 has nothing to do with natural disasters themselves. Instead, it's generated by a series of mathematical artifacts introduced by flawed methods.

The first artifact is created by Pielke's uncritical analysis of the billion-dollar-disasters database itself — a database which is defined by a billion-dollar *threshold*. Unbeknownst to Pielke, this threshold creates an artificial skew in the disaster data — skew which makes it appear like average disaster costs are decreasing against GDP.

On top of this threshold effect, Pielke then adds a distortion created by botched inflation adjustment. (He warps history by using conflicting price indexes.) Next, Pielke forces his stack of artifacts through an inappropriate linear regression, which both straightens the trend line, but also (paradoxically) renders it statistically insignificant. He then finishes the job by not reporting his (insignificant) p-values, mis-indexing his GDP data, reporting data sources that are at best incomplete, and mislabeling his y-axis. In short, while Pielke's paper waxes about 'scientific integrity', his published analysis serves mostly to undermine his own credibility.

With scientific integrity in mind, here is the road ahead. We'll begin with the pitfall lurking within the billion-dollar-disasters dataset, which is the billion-dollar threshold itself. We'll see how this threshold creates measurement bias, and we'll look at ways to deal with the problem. Then, just when the disaster data is getting interesting, we'll throw it away and retrace Pielke's steps to producing Figure 1. Finally, once we've reproduced this comedy of errors, we'll reflect on the lessons learned.

The billion-dollar threshold

Any analysis of the billion-dollar-disasters dataset must grapple with a subtle but severe problem, which is that the database is defined by a *dollar-level threshold*: it records the costs of US weather and climate-related disasters with inflation-adjusted losses that exceed $1 billion.

Now this billion-dollar threshold makes for great PR, since everyone knows that a billion dollars is a big number. But press releases aside, the billion-dollar threshold is actually a burden for doing accurate analysis. The issue is that when it comes to natural disasters, what ultimately matters isn't the dollar value of disaster costs; what matters is disaster *solvency* — the scale of disaster costs relative to income. If disaster costs rise, but income rises faster, then there is no problem (at least for humans). But if disaster costs rise faster then income, the pattern points towards insolvency.

To his credit, Roger Pielke Jr. recognizes this issue, and he recommends a solution with which I agree. To put US disaster costs in the context of solvency, we should compare them to US aggregate income (i.e. GDP). As such, the road ahead looks straightforward. First, we peg billion-dollar-disaster costs to US GDP. Then we look at the results. Easy.

Except not.

The issue is that when we place the billion-dollar-disasters data in the context of GDP, the dataset itself creates an artificial trend over time. If we're not careful, we might mistake this artifact for something real.

Figure 2 illustrates the problem. Here, I've taken the most recent version of the billion-dollar-disasters dataset and measured the nominal cost of each disaster as a share of US nominal GDP. (Each blue point is an individual disaster.) The database pitfall lurks at the bottom of the figure. Looking at the red curve, notice that there are no disasters below it. This absence is by design. That's because the red curve represents the billion-dollar threshold used to define 'billion-dollar disasters'. By definition, a 'billion-dollar disaster' cannot sit below this value.

used to define 'billion-dollar disasters'. (By definition, 'billion-dollar disasters' cannot sit below this line.) When we index this threshold to the consumer price index and then pegged it against US nominal GDP, the value heads south with time. This moving threshold introduces a skew in the data — skew which produces an apparent downward trend in disaster costs. Sources and methods

Now, the problem is that when this ostensibly fixed threshold is indexed to the consumer price index and then pegged to US GDP, it *heads south with time*. This movement, in turn, creates a skew in our disaster data. If we're not careful, we can mistake this artificial skew for a real-world trend.

For example, consider the average cost of billion-dollar disasters as a share of US GDP. In 1980, losses from billion-dollar disasters averaged about 0.13% of US GDP. But by 2024, average losses had declined to 0.023% of US GDP — a fivefold decrease.

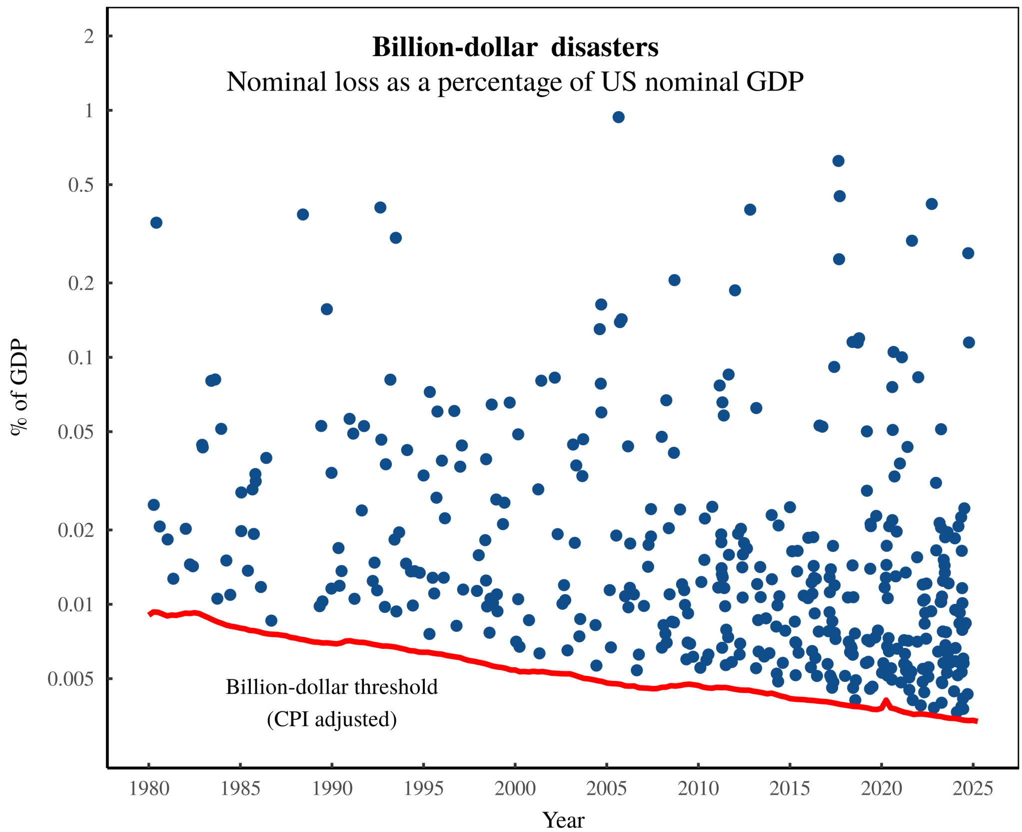


Figure 2: Billion-dollar disaster losses as a share of US GDP
Blue points show individual disasters contained in the billion-dollar-disasters dataset. The vertical axis shows the nominal cost of these disasters as a percentage of US nominal GDP. The red line shows the billion-dollar threshold,

---

Although seemingly impressive, this decline is driven by a threshold effect which makes a simple average misleading. The effect is similar to a university which gradually admits students of lower calibre, causing the average GPA to decline with time. While this pattern is 'real' in a statistical sense, it says nothing about the student population at large; it's purely an artifact of the university's admissions criteria.

In much the same way, the billion-dollar-disasters dataset has gradually (and unknowingly) changed its admissions criteria by including disasters with lower and lower costs as a share of US GDP. The consequence is that average disaster losses appear to decline against GDP. But like our decreasing student GPA, this apparent decline is a selection effect. It says nothing about the population of real-world disasters.

Looking ahead, Roger Pielke Jr. mistakes this billion-dollar threshold affect for a real-world trend in disaster costs. And from there, things only get worse.

A constant solvency threshold

We'll get to Pielke's cascade of errors in a moment. But first, we should analyze the billion-dollar-disasters data in a way that avoids the threshold effect.

Given the bias created by the billion-dollar threshold, there are two ways to deal with the problem. The best solution would be to order a whole new disasters database, one which lacks a cost threshold. The catch, of course, is that loss estimates almost always come with some sort of threshold, since government budget constraints (and data limitations) make it difficult to accurately track the losses from all disasters. So the question is not *if* there is a cost threshold, but *what* this threshold should be.

On that front, the second solution to the threshold problem is to introduce a *new* threshold, one which is fixed against GDP. In other words, we cull the billion-dollar-disasters data using a constant 'solvency threshold'.

Figure 3 illustrates this solvency approach. With our set of billion-dollar disasters, we exclude those that sit outside the red shaded region. The goal here is to keep disasters with losses that exceed a specific fraction of US GDP. In this case, I've chosen a solvency threshold of about 0.008% of GDP, as indicated by the dashed red line. Given this solvency threshold, we exclude disasters that sit below it. And we also exclude disasters to the *left* of the red box. We do so because in these years, there is the potential for missing data.

(Let me explain. When the original billion-dollar threshold sits above my new solvency threshold, a disaster might have losses that exceed the solvency threshold, but which do *not* exceed the original billion-dollar threshold. Such a disaster would therefore be absent from the existing billion-dollar-disasters database. If we want to avoid this missing-data bias, we need to exclude years when the billion-dollar threshold exceeds the new solvency threshold.)

Now I should acknowledge that this culling procedure is in some sense arbitrary. Just as there is no scientific meaning to a loss threshold of $1 billion, there is no scientific meaning to my solvency threshold of 0.008% of GDP. I have chosen this value because it is functional — it culls the billion-dollar-disasters database without throwing out too much data.

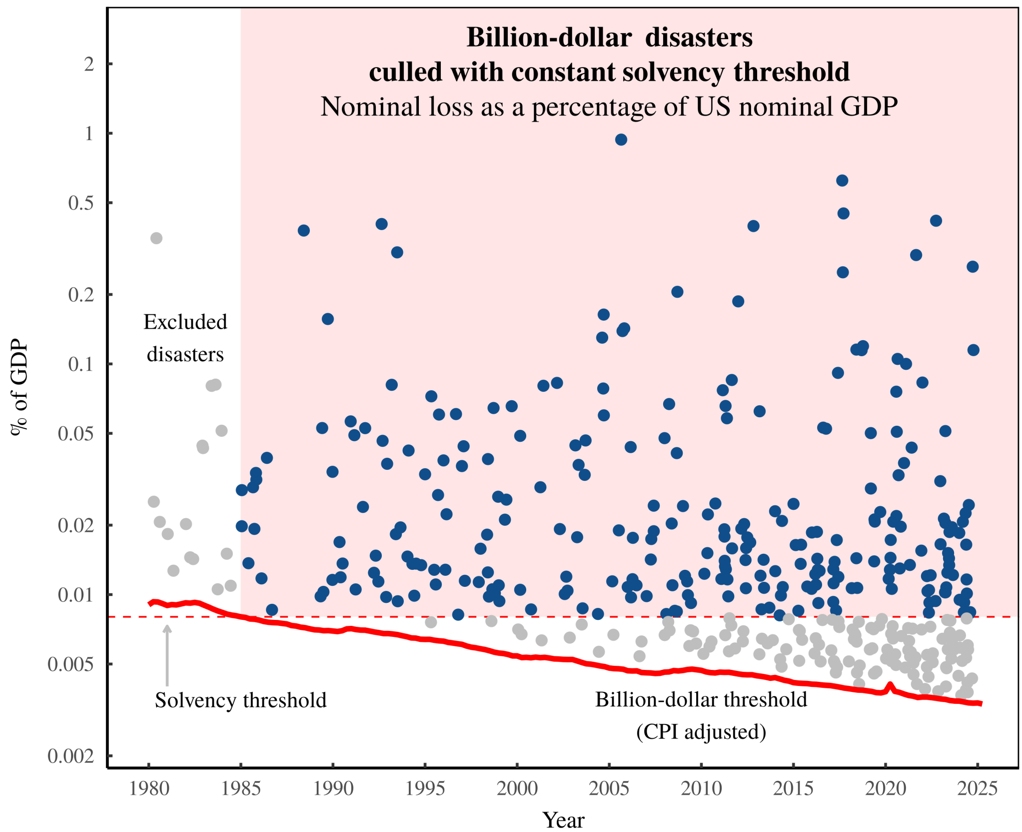


Figure 3: Culling billion-dollar disasters with a constant solvency threshold

This figure illustrates my solution for dealing with the billion-dollar threshold problem. Instead of a dollar-level threshold, I cull the disasters data using a 'solvency threshold' — a constant share of US GDP. Disasters with losses below this threshold are excluded. Also excluded are disasters which occur in years when the new solvency threshold exceeds the original billion-dollar threshold. Disasters that remain live inside the red shaded region. Sources and methods

An honest appraisal of natural disaster trends

With our culled billion-dollar-disasters dataset in hand, our first step will be to give the data an honest appraisal. Perhaps the most pressing question is the trend in disaster solvency. If we sum the annual disaster costs in our culled dataset, how do they behave against US income? As Figure 4 indicates, there is tentative evidence that disaster costs are *rising* with time.

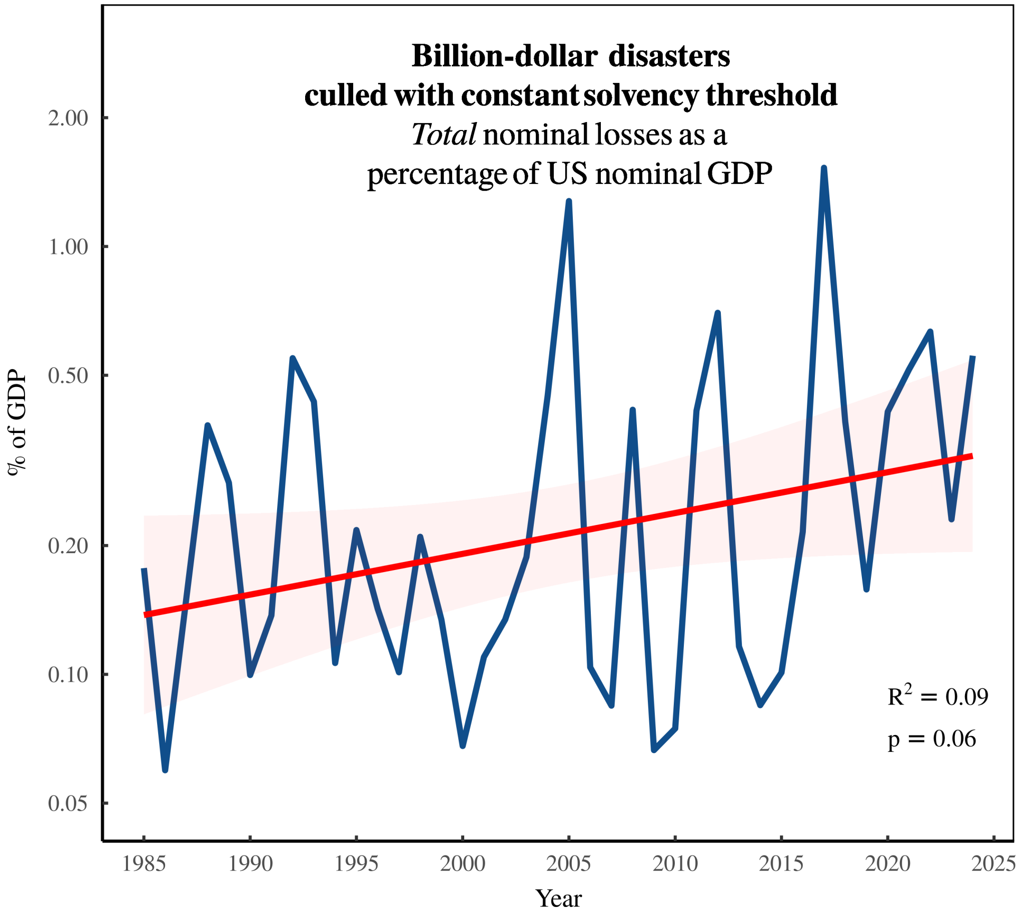


Figure 4: Total disaster losses as a share of US GDP — culled billion-dollar disasters

Looking at my culled billion-dollar-disasters dataset (see Figure 3), I find a slight upward trend in total annual losses measured as a share of US GDP. (Note the log scale on the vertical axis.) The red line and shaded region show a log-linear regression and associated confidence interval. Sources and methods

---

Now, I'll be the first to admit that this upward trend is not particularly compelling. (The regression p-value is 0.06, and the R^2 value is 0.09.) That said, even the *hint* that disaster costs are rising against income should give us pause for thought.

One of the great themes of the industrial revolution has been humanity's rising power over the natural world. (Once the pawns of nature, now we are the gods.) As such, it seems like our technological prowess should *isolate* us from the costs of natural disasters, making these losses increasingly trivial. Yet judging by the pattern in Figure 4, the evidence cuts (tentatively) in the opposite direction.

Actually, the evidence for worsening natural disasters is more compelling than I'm letting on. That's because one of the features of natural disasters is that the loss data is incredibly noisy. The reason is straightforward: annual disaster losses are dominated by rare but calamitous events — for example, a massive hurricane hitting a big city. (See the appendix for details.) Since these calamitous events are unpredictable, the effect is that total disaster losses vary wildly year to year. Against this short-term noise, a subtle trend can easily get swamped.

To separate the noise from the signal, let's take our culled billion-dollar-disasters data and measure the trend in average losses per disaster. Or rather, we'll measure the *lack* of trend. As Figure 5 illustrates, when we peg average annual disaster losses against US GDP, there is plenty of short-term noise, but no hint of a long-term trend.

The message in Figure 5 is that the average severity of big natural disasters is essentially unpredictable — it's a game of roulette played between humans and the gods of weather. Still, this game doesn't leave everything to chance. In fact, if we ignore the severity of big disasters and instead count their *frequency*, we see a clear pattern over time.

Figure 6 runs the numbers. Here, I count the annual number of disasters in my culled billion-dollar-disasters database. (To reiterate, these are disasters with losses that exceed 0.008% of US GDP.) It seems that the frequency of these big disasters is rising with time.

Now, if this essay was about good science, I'd pivot here and ask a bunch of questions. First, is the pattern in Figure 6 robust? Are big natural disasters becoming more frequent everywhere? Or is the pattern unique to this particular dataset?

Supposing that the pattern *is* robust, I'd then want to understand the cause. Now the obvious culprit is climate change, which is likely making weather more volatile. But another plausible culprit is the nature of modern capitalism itself. You see in recent decades, investors have turned to real-estate as a place to park their money. As a consequence, house prices have steadily risen against income. Now if property *values* are rising against income, it follows that property *losses* might also rise. As such, the rising frequency of costly natural disasters could be an artifact of modern investment patterns.[3]

[3]Continuing to think about good science, one could differentiate between these two scenarios (climate change vs rising property values) by looking at the long-term history of disaster losses. If rising disaster losses are being driven exclusively by climate change, then

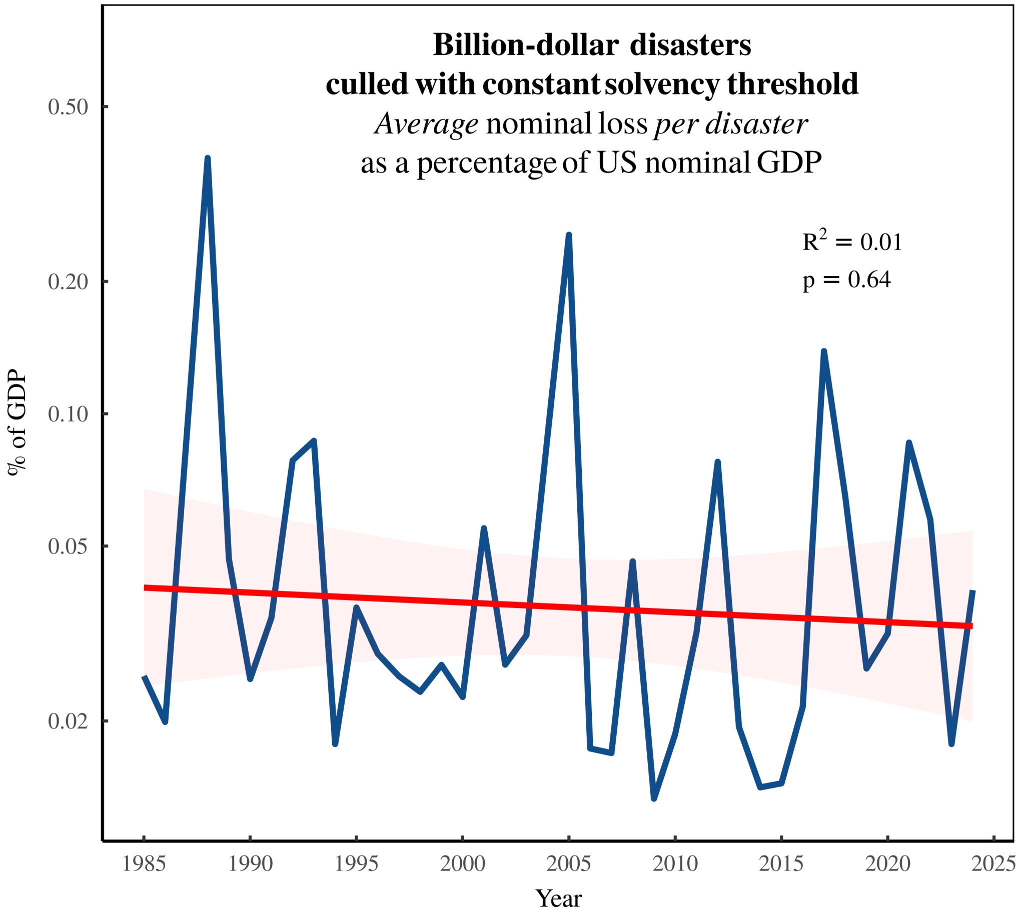


Figure 5: Average loss per disaster as a share of US GDP — culled billion-dollar disasters

Looking at my culled billion-dollar-disasters dataset (see Figure 3), I find no trend in the average cost per disaster, measured as a share of US GDP. (Note the log scale on the vertical axis.) The red line and shaded region show a log-linear regression and associated confidence interval. Sources and methods

---

Having piqued my interest, I'm now going to pivot and *not* investigate these questions. And that's because this essay is *not* about good science. It's about the error-riddled work of Robert Pielke Jr.

---

the secular rise in losses should date back centuries. If, however, the trend is being driven by the dynamics of house prices, we should see an L-shaped pattern — a decline in disaster losses until 1970 and a rise thereafter.

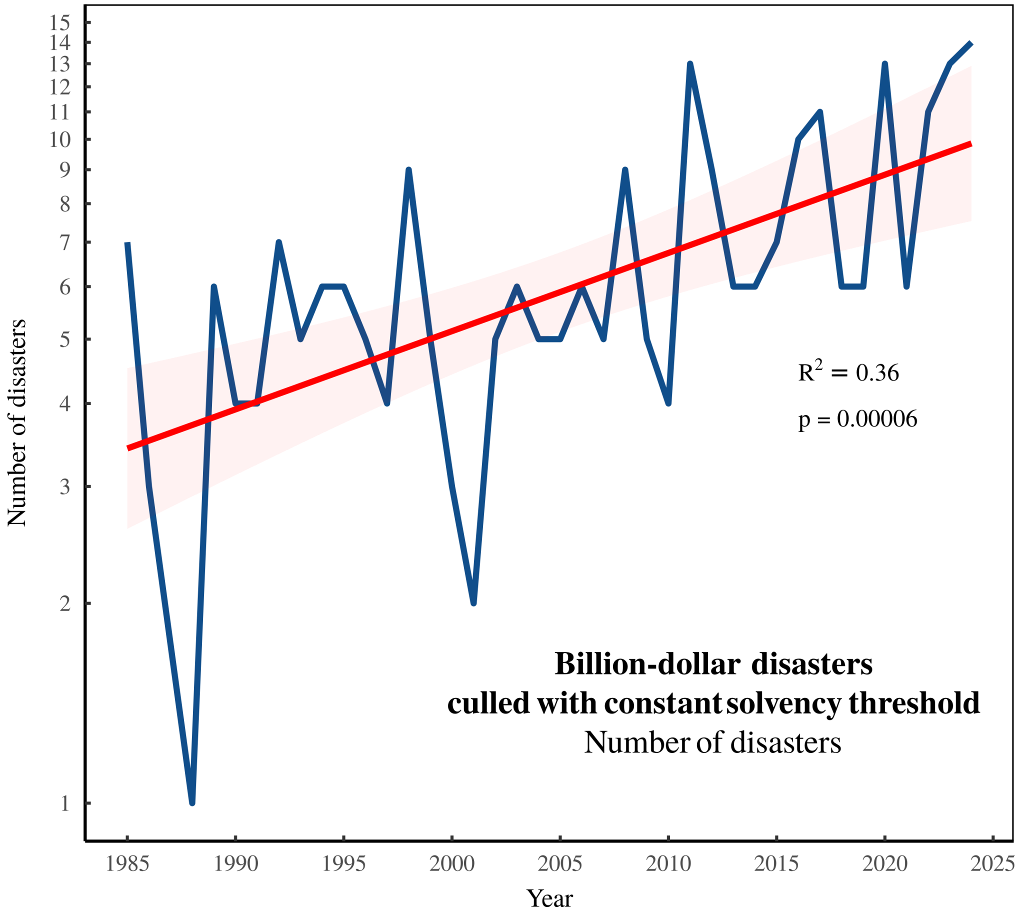


Figure 6: Disaster frequency — culled billion-dollar disasters
Looking at my culled billion-dollar-disasters dataset (see Figure 3), I find that the number of disasters has trended upward with time. (Note the log scale on the vertical axis.) The red line and shaded region show a log-linear regression and associated confidence interval. Sources and methods

---

Fooled by a mathematical artifact

To situate Pielke's analysis of the billion-dollar-disasters data, it's helpful to first describe the broader landscape of climate-change science. Throughout most of the terrain, we find natural scientists. These are folks who are trained in the scientific method and who genuinely want to understand the potential impacts of climate change — good, bad and ugly.

As we continue across the terrain, we find a small but vocal tribe of academics who call themselves 'climate economists'. These are folks who masquerade at doing science, but whose job is mostly to run interference for business interests. For example, when natural scientists find a pattern that looks bad, the ‘climate economists’ step in and manipulate the data until it confesses to a ‘better’ story. (For details about the sorry state of climate economics, see Steve Keen’s work.)

As I see it, Pielke’s analysis of natural disasters fits into the ‘climate economist’ camp. Unlike a ‘climate crank’, Pielke does not spout complete bullshit. He analyzes real data and he finds reproducible patterns that emerge from this data. But curiously, these patterns seem to always tell a happy story.

Pielke’s analysis of the billion-dollar-disasters data is a case in point. He finds a trend which is ‘real’, in the sense that it can be reproduced from real data. And the trend is also a ‘happy’ one — it suggests that natural disaster costs are decreasing against income. Unfortunately, in his rush to spread good news, Pielke ignores Richard Feynman’s first principle of science, which is that (a) you must not fool yourself; and (b) you are the easiest person to fool.

Here is how Pielke gets fooled.

Although Pielke is critical of the billion-dollar-disasters dataset, he does not analyze the data itself with a critical lens. Seemingly unaware of the billion-dollar threshold problem (outlined above), Pielke naively takes the whole billion-dollar-disasters database and dumps it through a simple average. Out pops the ‘good news’, which I’ve visualized in Figure 7. When we measure the average annual cost of billion-dollar disasters, we find that the losses have steadily declined against US GDP.

(Note: Pielke uses a different version of the billion-dollar-disasters dataset then the one used here. I replicate his exact results in the appendix.)

Although seemingly convincing, the trend in Figure 7 has a slight problem, which is that it has nothing to do with real-world natural disasters. Instead, the downward trend is entirely an artifact of the billion-dollar threshold on which the disaster data is based.

Figure 8 illustrates the principle. Here, the red line shows the trend from Figure 7 — the downward pattern in average disaster costs as a share of US GDP. Next to this trend, the blue line shows the billion-dollar threshold used to exclude disasters from the billion-dollar-disasters database. When we tie this billion-dollar threshold to the consumer price index and then peg it against US GDP, the threshold moves south with time. Indeed, it moves south at the *same rate* as the apparent trend in average disaster losses.

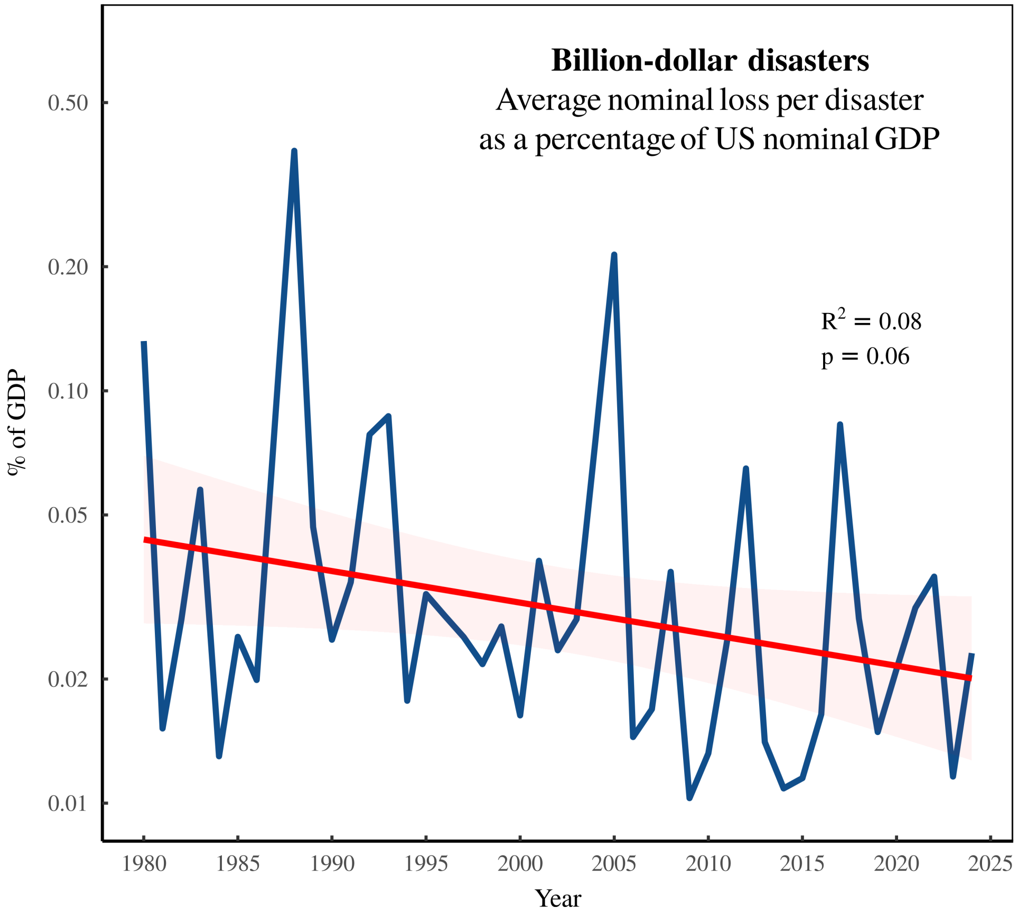


Figure 7: Average loss per billion-dollar disaster measured as a share of US GDP

Unlike previous figures (which use my culled dataset), this chart looks at trends in the *full* billion-dollar-disasters dataset. As expected (from the billion-dollar threshold effect), we find that average losses per disaster have declined against US GDP. (Note the log scale on the vertical axis.) The red line and shaded region show a log-linear regression and associated confidence interval. Sources and methods

This similarity is no coincidence. It is cause and effect. When we peg the billion-dollar-disasters data against US GDP, the role of the actual disaster costs is to provide statistical noise. When we then feed this noise through a threshold that moves relative to GDP, it introduces a skew in the data. Finally, when we run a regression on the skewed data, we get back a (rescaled) version of the billion-dollar threshold itself.

In short, the trend line in Figure 7 tells us nothing about natural disasters. Instead, it circuitously tells us what we already knew: that the billion-dollar-disasters dataset is defined by a billion-dollar threshold.

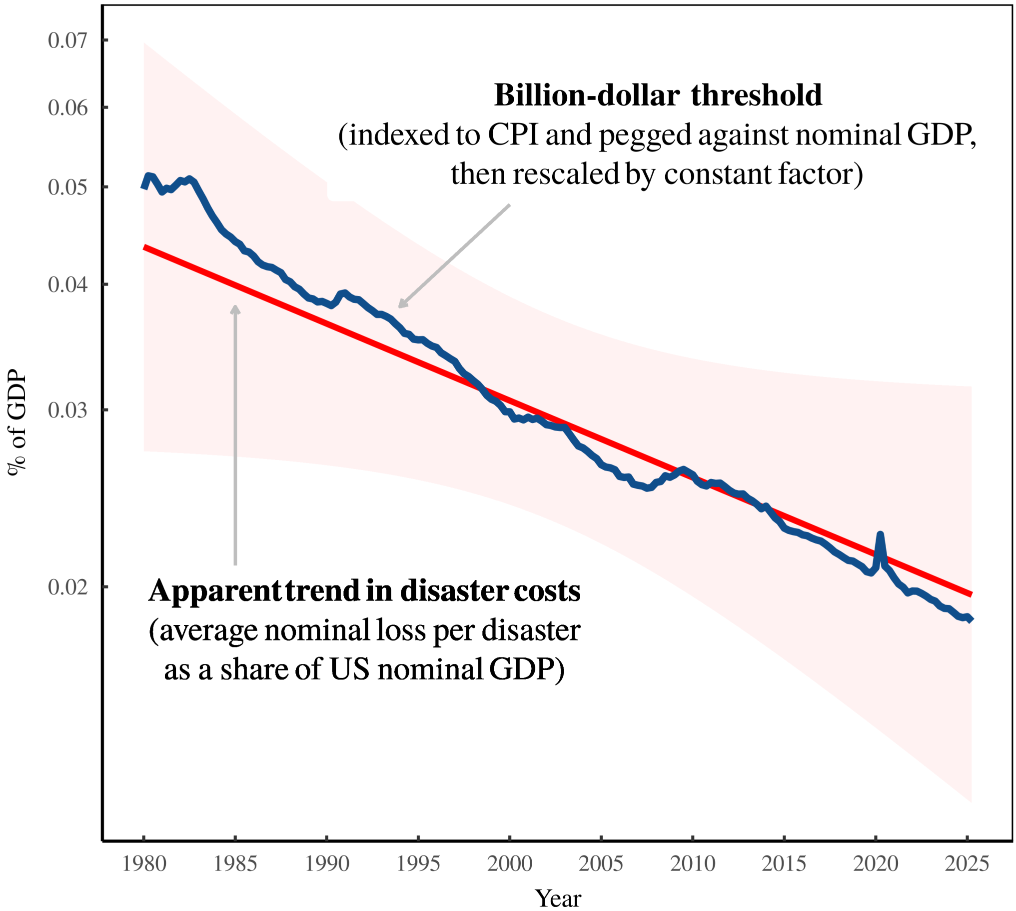


Figure 8: A billion-dollar artifact

This figure illustrates how the billion-dollar threshold creates an apparent downward trend in average disaster losses. When we index $1 billion (circa 2024) to the consumer price index and then peg this threshold to US GDP, we get the blue curve. As it happens, this curve has the same slope as the apparent downward trend in average disaster costs (red line, see Figure 7 for underlying data). This similarity not chance; it's cause and effect. The moving threshold creates a skew in the disaster data. When we then run a regression on the skewed data, we get back the slope of the threshold itself. Sources and methods

The plot thickens

Backing up a bit, note that the regression line in Figure 7 has a p-value of 0.06. Now since the trend in the data is itself a statistical artifact, so too is the p-value. Still, if we mistook the trend line for a real-world pattern, then the high p-value would be a problem; it would suggest that the decline in average disaster costs is 'statistically insignificant', which means we shouldn't make much of it.

Here is where the plot thickens.

Pielke *does* make a big deal about the downward trend in average disaster costs. But his underlying regression is *not* statistically significant. (See the appendix.) So how did Pielke slip this contradiction past peer review? Simple. It seems he didn't report any p-values.

Here, the plot gets even thicker. Pielke *could* have published a 'statistically significant' pattern had he run the correct regression. But this 'significant' pattern is itself created by a *second* mathematical artifact which distorts the *first* mathematical artifact on which Pielke's results are based. This comedy of errors is what I mean when I say that Pielke's analysis resembles satire.

History, adjusted

Continuing to reproduce Pielke's work, our next step will be to understand his second mathematical artifact, which is created by the use of conflicting price indexes. The effect of this artifact is to steepen the downward trend in average disaster losses, thereby making the regression 'statistically significant'.

In my mind, this price-index distortion needs some explanation to be comprehensible. As such, I'll begin with a brief tutorial on how mismatched price indexes can be used to rewrite history.

Let's start with some hypothetical facts. Suppose that in 1980, my annual income was $10,000. And suppose that I used this income to purchase a car that also cost $10,000. While these dollar values are themselves historical facts, they are not meaningful on their own. What gives them meaning is their relationship — the fact that in 1980, my car purchase represented 100% of my annual income.

Now, the thing about historical facts is that (assuming they are accurate), they *should not change*. So regardless of how we study history, we should always find that my 1980 car purchase represented 100% of my 1980 income.

Having started with clear thinking, let's now introduce some confusion, courtesy of mainstream economics. "Hold on," economist say, "if we wish to study historical prices, we must adjust for inflation." Bowing to economists' authority, we decide to look at my 1980 car purchase through an inflation-adjusted lens. To do that, we adjust the 1980 dollar values to reflect modern prices (circa 2025). We use the consumer price index to adjust the value of my car purchase, which gives a modern price of about $40,600. And we use the GDP deflator to adjust my income, which gives a modern value of about $33,800.

(Why are we using the GDP deflator here? Well, because my 'income' is a metaphor for GDP, which is implicitly indexed to the GDP deflator.)

Having dutifully adjusted for inflation, we once again calculate the cost-to-income ratio for my 1980 car purchase. Intriguingly, we find that the ratio is not 1:1, as we once thought. According to our inflation-adjusted values, my 1980 car actually cost *120%* of my 1980 income.

Confused, we decide to rerun our calculations using different reference years. For example, we adjust the 1980 dollar values into 2024 prices, and then calculate the car's cost-to-income ratio. We do the same for the reference years of 2023, 2022, and so on, all the way back to 1980. When we plot our results, we find the pattern shown in Figure 9. The cost-to-income ratio for my 1980 car purchase appears to grow with time. History, it seems, can be steadily rewritten.

Of course, this 'rewrite' is a joke. In reality, we've been led astray by inflation 'adjustments' which are both *unnecessary* and *invalid*. Let's start with the 'unnecessary' part. Just as we don't need to 'adjust' for inflation when we purchase today's groceries using today's pay check, we don't need to 'adjust' for inflation when we compare historical transactions at the same point in time.

Now to the 'invalid' part. Why does inflation adjustment distort the relative value of my car purchase? Well, it doesn't *have* to act this way. The trick in Figure 9 is that we used two different price indexes which, unbeknownst to us, give conflicting accounts of US inflation.

Let's have a look at this wrinkle in history. Figure 10 shows the historical movement of the US consumer price index (red) and the US GDP deflator (blue). Until the late 1970s, these two indexes moved together. During this era, we could mix the CPI with the GDP deflator without creating much trouble. But from 1980 onward, our two price indexes parted ways, with the CPI rising much faster than the GDP deflator. During this post-1980 era,

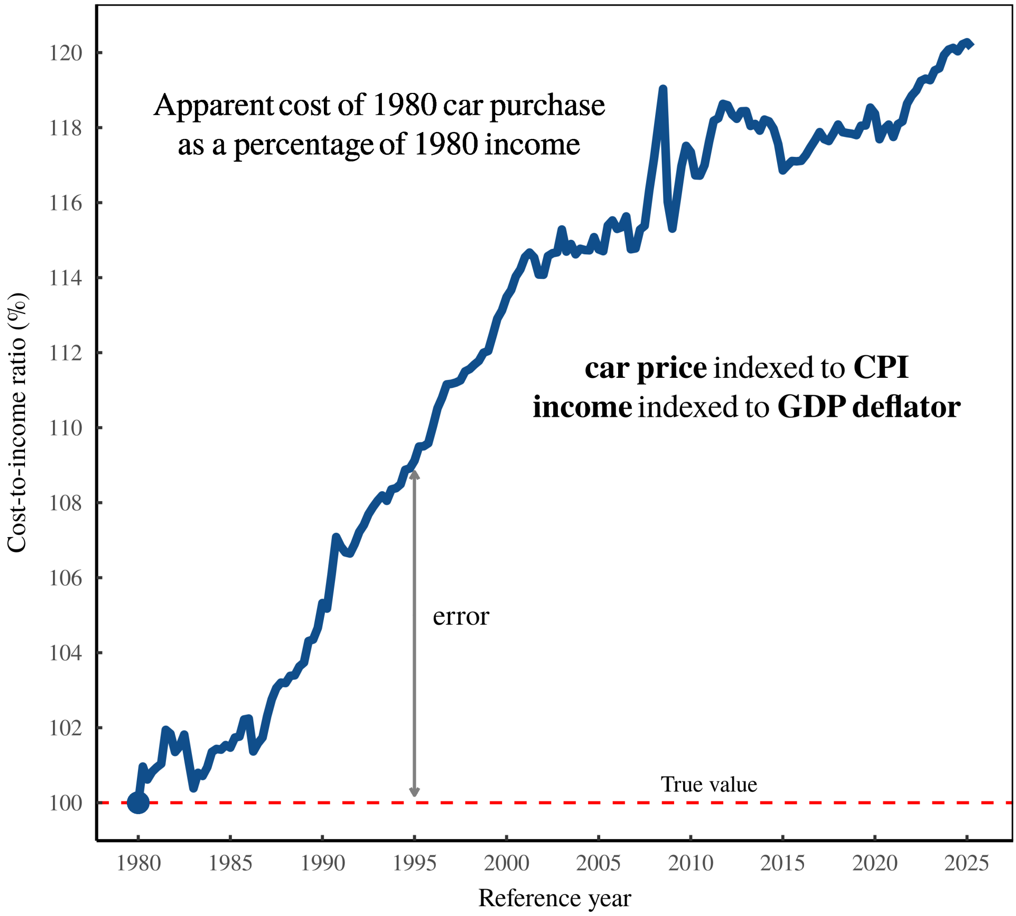


Figure 9: Rewriting history with inflation adjustment

Here, we suppose that in 1980, I purchase a car that costs 100% of my annual income. Then we view this historical purchase through an inflation-adjusted lens. We index the price of my 1980 car to the consumer price index. And we index my 1980 income to the GDP deflator. We find that the car's price seems to move against my income, depending on the 'reference year' used for inflation adjustment. (The 'reference year' is the point where we update the 1980 prices to the associated inflation-adjusted values.) Sources and methods

---

mixing the CPI with the GDP deflator became a statistical no-no, because it provides a way to distort history. A past purchase that is tied to the CPI will appear to *move* relative to income that is tied to the GDP deflator.[4]

[4]A historical aside: in the mid 1990s, economists at the Bureau of Economic Analysis decide to change how they calculated 'real' GDP, switching from a 'base-year' approach to a 'chain-weighting' approach. The specifics of this change are not important. What *is* important is that the revision gave a nice boost to real-GDP growth, a boost achieved by reducing the rate of inflation, as measured by the GDP deflator.

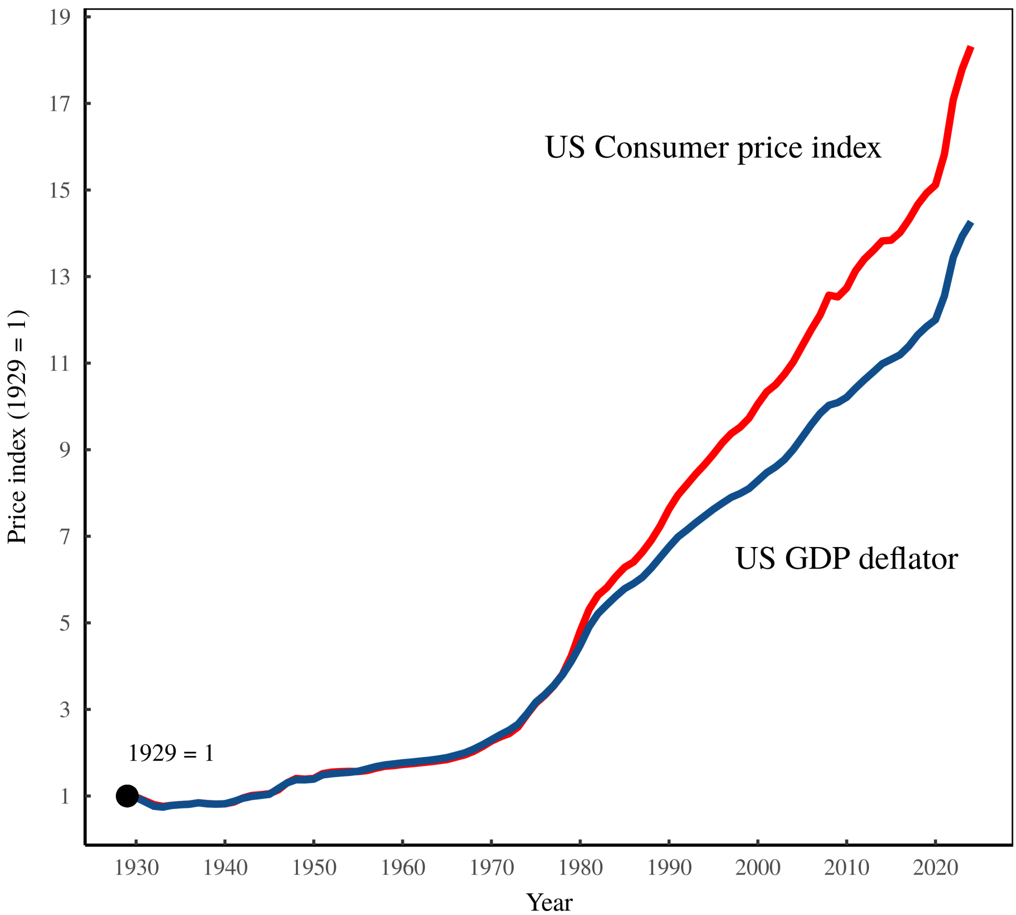


Figure 10: A fork in price-index history

This chart shows the long-term history of two measures of US inflation — the consumer price index (red) and the GDP deflator (blue). After moving together for a half century, the two series parted ways in 1980. Thereafter, the CPI gave a consistently higher measure of inflation. Because of this post-1980 divergence, mixing the CPI with the GDP deflator has become illegitimate, and provides a way to warp history. Sources and methods

Artifact atop of artifact

Back to Pielke's work. When we left off, we'd seen how the billion-dollar threshold effect created an apparent downward trend in the average cost of natural disasters (as a share of GDP). Apparently unaware of this effect, Pielke mistakes the downward pattern for something real.

Now to step two in Pielke's error cascade. On top of the threshold artifact, Pielke adds a distortion created by conflicting price indexes. (He mixes the CPI with the GDP deflator, just as we did in Figure 9.)

To (unwittingly) achieve this distortion, Pielke's thinking might have gone something like this. Putting on our 'proper economist' hat (Pielke's term), we see that the billion-dollar-disasters dataset provides CPI-adjusted values for disaster losses. In the language of mainstream economics, these adjusted losses represent 'real' monetary value. As such, we should put these 'real' losses in a solvency context by comparing them to 'real' GDP.

While this comparison seems reasonable, the language of economics misleads us. The so-called 'real' value inferred from the CPI is not the same 'real' value that is inferred from 'real' GDP. Under the hood, the latter data is indexed to the GDP deflator, which conflicts with the CPI. The result is that if we compare CPI-adjusted disaster losses to 'real' GDP, we unwittingly rewrite history.

Importantly, the effect of this rewrite is to bolster the apparent downward trend in average disaster costs (as a share of GDP). Figure 11 illustrates the distortion. Here, the dashed line shows the trend in average disaster costs derived from nominal monetary data. (See Figure 7.) The red line shows the updated trend after the data is warped by our conflicting price indexes. Although the change is subtle, it is 'significant' in the sense that it lowers our regression p-value below the magic threshold of 5%. According to standard statistical lore, our results are now 'publishable'.

Unfortunately, this reassuring p-value is, like the trend itself, a mathematical artifact. In fact, it is generated by *multiple* mathematical artifacts. To create the pattern in Figure 11, we start with the billion-dollar-threshold effect, which generates an apparent downward trend in average disaster losses as a share of GDP. On top of this artifact, we then distort history by indexing the data to conflicting price indexes.

The net result is that we've still said nothing about real-world natural disasters. As Figure 12 demonstrates, we can predict our downward trend in disaster losses simply by transforming the billion-dollar threshold itself. When we tie this threshold to the consumer price index, compare it to GDP, and then distort the result with price-index shenanigans, out pops the apparent slope in our disaster data. Magic!

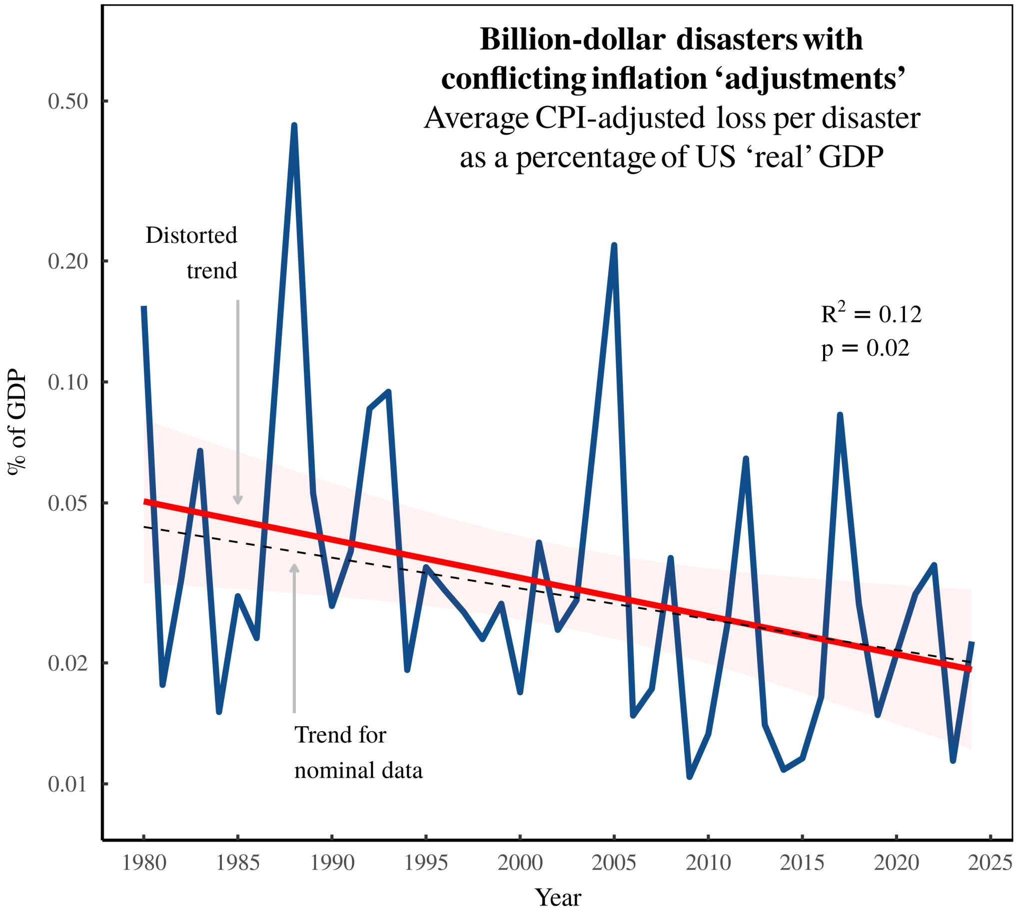


Figure 11: Disaster losses get 'real'

This chart illustrates how flawed inflation adjustment can bolster the downward trend in average disaster costs as a share of US GDP. The dashed grey line shows the apparent disaster trend calculated with nominal data. In contrast, the red line shows the revised trend, calculated using CPI-adjusted disaster losses and US 'real' GDP. The trend steepens, but only because our conflicting price indexes have warped history. Sources and methods

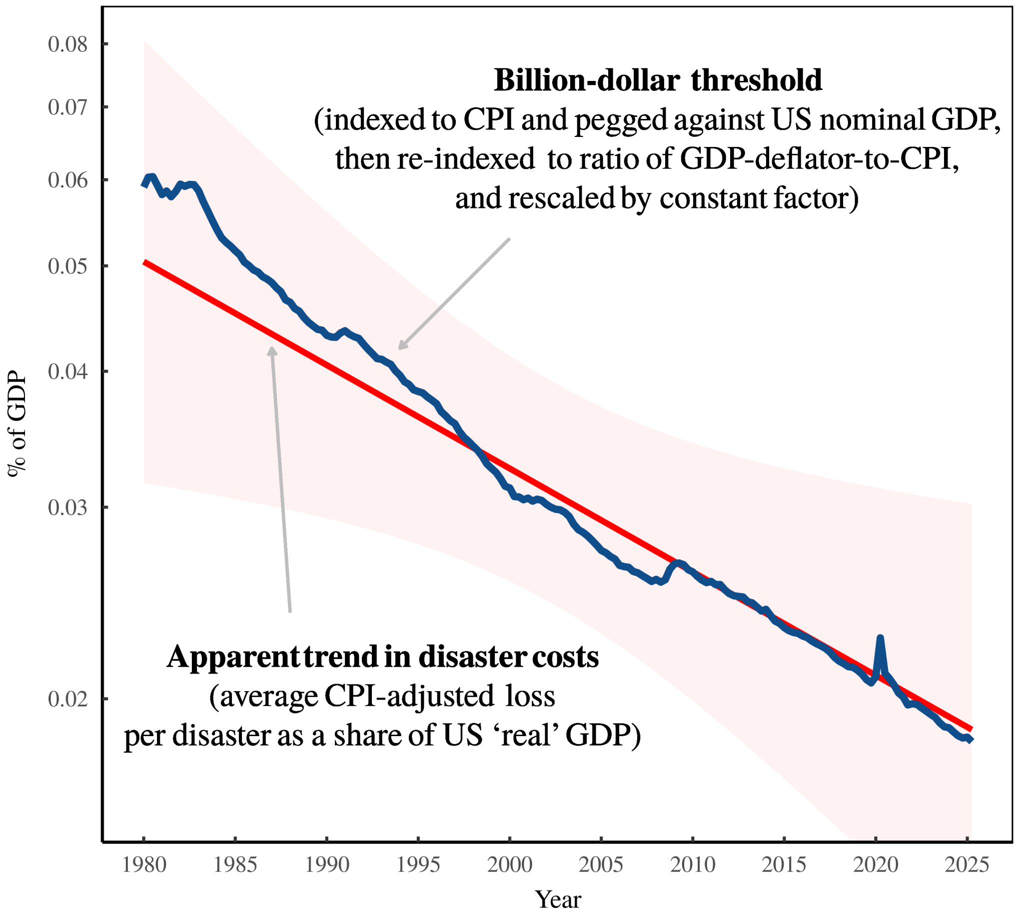


Figure 12: A billion-dollar artifact, distorted by conflicting price indexes

Although we've added more moving parts to our stack of mathematical artifacts, the apparent trend in inflation-adjusted losses per disaster (red line, see Figure 11 for underlying data) still says nothing about real-world disasters. It is produced by the billion-dollar threshold (blue line), which is indexed to the consumer price index, pegged against nominal GDP, and distorted by mismatched price indexes. Sources and methods

A linear bed of Procrustes

It's at this point that our plot line gets weird. In his paper criticizing the billion-dollar-disasters dataset, Pielke could have reasonably published a chart like Figure 11. And by 'reasonably', I mean that such a chart appears statistically valid. It plots a regression line that is a good fit to the data, and it reports a p-value that's in the correct 'zone'. (Of course, we know that these 'results' are a statistical artifact. But the average peer reviewer, busy as they are, would almost certainly miss this subtle but severe problem.)

Inexplicably, however, Pielke chooses to go a different route and publish a chart that is both a mathematical artifact *and* statistically dubious. Figure 13 visualizes this twist. Here I've taken the data from Figure 11 and plotted it on a linear scale. By doing so, I've implicitly changed my trend line from being a log-linear regression to a simple linear regression.

This scale transformation creates several problems. The first is that the new linear regression is a demonstrably poor fit to the data. Visually it looks bad. And we can confirm the poor fit by analyzing the regression residuals. (See the appendix.) Worse, the switch to a linear trend ruins our regression p-value, bumping it up to a non-publishable value of 0.09. Oddly, Pielke seems to have 'solved' this problem simply by not reporting his p-values. Finally, the linear trend implies that average disaster losses will soon become *negative*, meaning disasters losses transform into *gains*. In this implied world, hurricanes hand out cash to their victims. Nonsense.

Back to the running theme of Pielke's analysis. The effect of Pielke's inappropriate linear regression is to add a *third* mathematical artifact on top of his existing stack of errors. Let's review.

We start with the billion-dollar threshold effect, which generates an apparent decline in average disaster costs as a share of GDP. On top of this artifact, we steepen the downward trend by distorting the data with mismatched price indexes. Finally, we force the resulting artifact into an ill-fitting linear trend. The effect is to take a our curved artifact and warp it into a straight line.

Figure 14 illustrates the cumulative procedure which, by now, is quite convoluted. Still the results say nothing about real-world disasters. They are entirely a product of our analytic method.

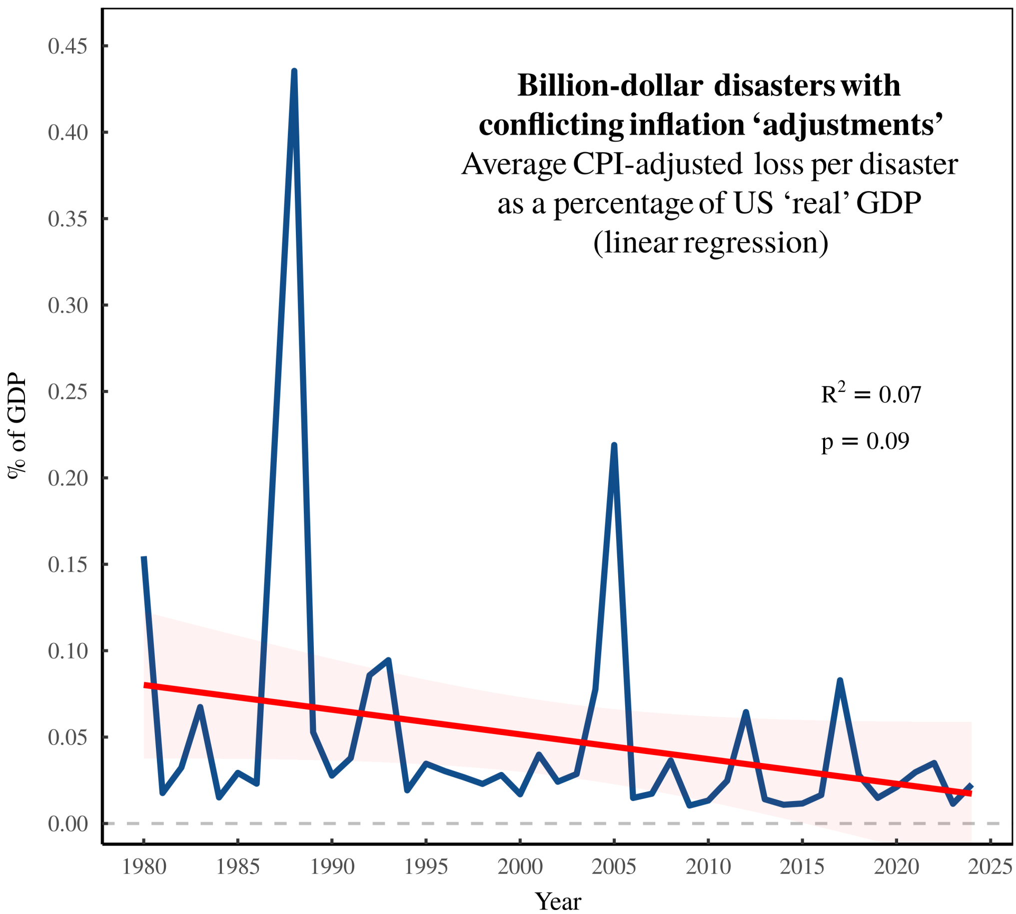


Figure 13: Disaster losses, forced straight

Here, I've reproduced Pielke's decision to plot average billion-dollar disaster costs (as a share of GDP) on a linear scale, and to fit the data with a linear regression. Ignoring the artifacts beneath the data (see Figures 7 to 12), this decision is odd, since the linear fit is demonstrably loose, and the high p-value warrants caution against strong conclusions. Sources and methods

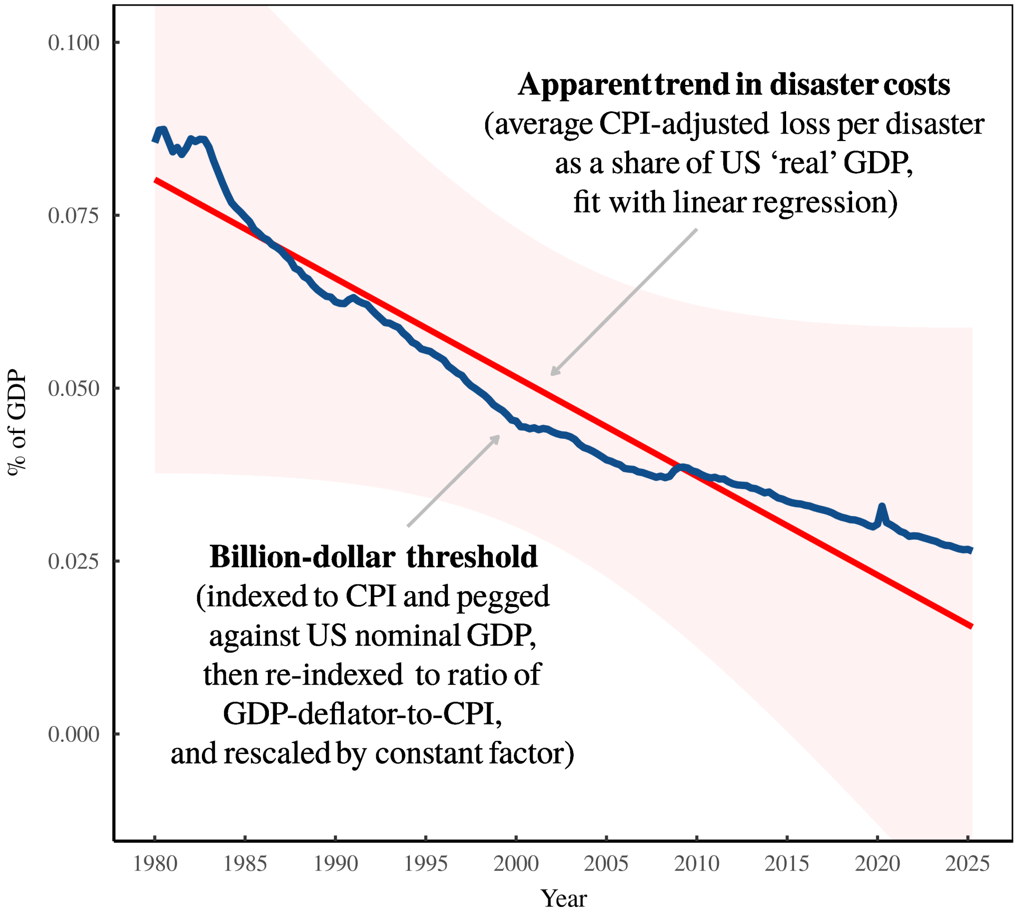
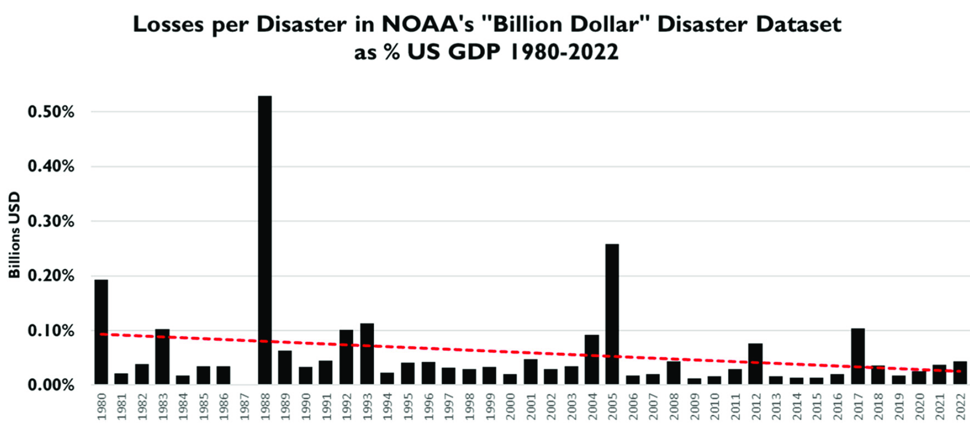


Figure 14: A billion-dollar artifact, distorted by conflicting price indexes, and flattened into a Procrustean straight line

This chart summarizes Pielke's stack of mathematical artifacts. The apparent trend in disaster costs is generated by the billion-dollar threshold effect, which is then distorted by conflicting price indexes. The final step, illustrated here, is to shove the curved pattern though a linear regression, which forces it into an ill-fitting straight line. Sources and methods

Pielke's artifact stack

We're now ready to return to Pielke's published analysis of the billion-dollar-disasters data. As shown in Figure 1 (reproduced below), Pielke reports that relative to US GDP, average billion-dollar-disaster costs have declined by about 80% over the last four decades.



Roger Pielke Jr.'s analysis of the billion-dollar-disasters dataset.

Having painstakingly tracked Pielke's method, we know that such a trend *can* be extracted from the billion-dollar-disasters data. But we also know that this pattern has *nothing* to do with real-world disasters. The apparent downward trend in disaster costs is created entirely by a stack of mathematical artifacts — the billion-dollar threshold effect, distorted by conflicting price indexes, and then straightened by an inappropriate linear regression.

Atop this artifact stack, Pielke's chart rounds things out with a few more errors. For example, it seems that Pielke mis-indexes his 'real' GDP data. (He uses data with the wrong reference year, thereby slightly inflating all the values in Figure 1.) Pielke also reports GDP source data that is either erroneous or incomplete. (His chart extends to the year 2022, but his reported GDP data ends in 2019.) Finally, Pielke's chart mislabels the vertical axis. (It should read '% of GDP', not 'Billion USD'.)

(For my exhaustive replication of Pielke's work, see the appendix.)

Summarizing the whole affair, if Pielke's analysis of billion-dollar disasters was meant to parody his own appeal to 'scientific integrity', it certainly succeeded.

Selecting for bad science

In my mind, Pielke's analysis of billion-dollar disasters is a good great example of what Paul Smaldino and Richard McElreath call the 'natural selection of bad science'. The idea is that doing junk science is not a conscious goal, but is instead a selection effect created by the pressures of the academic environment.

Put simply, good science is marked mostly by ideas that don't pan out. However, the modern academic environment rewards the incessant production of hypotheses that (apparently) *do* pan out. Given this environment, scientists have two choices for survival: they can either get better at generating true hypotheses, or they can develop unwitting ways for transforming null results into something 'significant'. Smaldino and McElreath argue that many scientists have opted for the second approach, which they achieve by unknowingly using bad statistics.

Turning to Pielke's analysis of the billion-dollar-disasters data, I think a similar effect is at play. Which is to say that Pielke's long list of errors was almost certainly unintentional. That said, I'd guess that the *direction* of these errors was not random.

Let's put it this way; suppose that a scientist is both error prone *and* biased towards results which downplay the financial costs of climate-related disasters. In this situation, random errors get filtered by the storytelling bias. The result is that *published* errors tend to cut in a direction that bolsters the storytelling goals. In other words, had Pielke's cascade of errors made it appear like average disaster costs were *increasing* against income, I'd guess that the results would not have been published.

The lesson here is that when we mix error-prone analysis with a storytelling bias, we create a surprisingly powerful method for identifying quirks and loopholes in a set of data. On that front, the billion-dollar-disasters dataset contains an unfortunate liability, which is the billion-dollar threshold itself. Simply put, this threshold makes unbiased analysis of the disasters data more difficult, and it creates unintentional artifacts that can be used to distort the actual evidence. In short, if the NOAA one day resumes tracking weather and climate-related disaster costs, it should abandon the billion-dollar threshold. At least then, there will be one less artifact to exploit.

Support this blog

Hi folks. I'm a crowdfunded scientist who shares all of his (painstaking) research for free. If you think my work has value, consider becoming a supporter.

Become an ETD supporter

Replicating Pielke's analysis

Here are the steps I took to replicate Pielke's analysis of the billion-dollar-disaster (BDD) data.

Since Pielke did not publish his chart data, my first step was to digitize the data plotted in his Figure 3 (which is my Figure 1). For that, I used a program called engauge-digitizer.

The next step was to figure out which BDD version Pielke worked with. To do so, I headed to the NOAA archive page and downloaded all of the BDD datasets published since 2019. (There are 21 versions in total.)

Next, I did some sleuthing. According to Pielke, his chart was generated using a BDD version "downloaded in July 2023". This date eliminates BDD versions published *after* July 2023. Meanwhile, Pielke's chart contains data up to 2022, which eliminates BDD versions which *don't* contain data from that year. Now, because the BDD database was published quarterly, this elimination process still leaves several possible BDD versions which Pielke might have used. To boil the options down to one, I then used each remaining BDD version to try to replicate Pielke's published results.

For this replication, I used GDP data from the Bureau of Economic Analysis (downloaded via FRED). Now, ideally, I'd have used the same GDP data as Pielke; however, along the way, I discovered that doing so was impossible. According to his paper, Pielke used 'real' GDP data from FRED series RGDPNAUSA666NRUG. But the problem is that this dataset only goes up to the year 2019, while Pielke's chart goes to 2022. So something is amiss. At any rate, the source for 'real' GDP data doesn't particularly, because most GDP time series are quite similar.

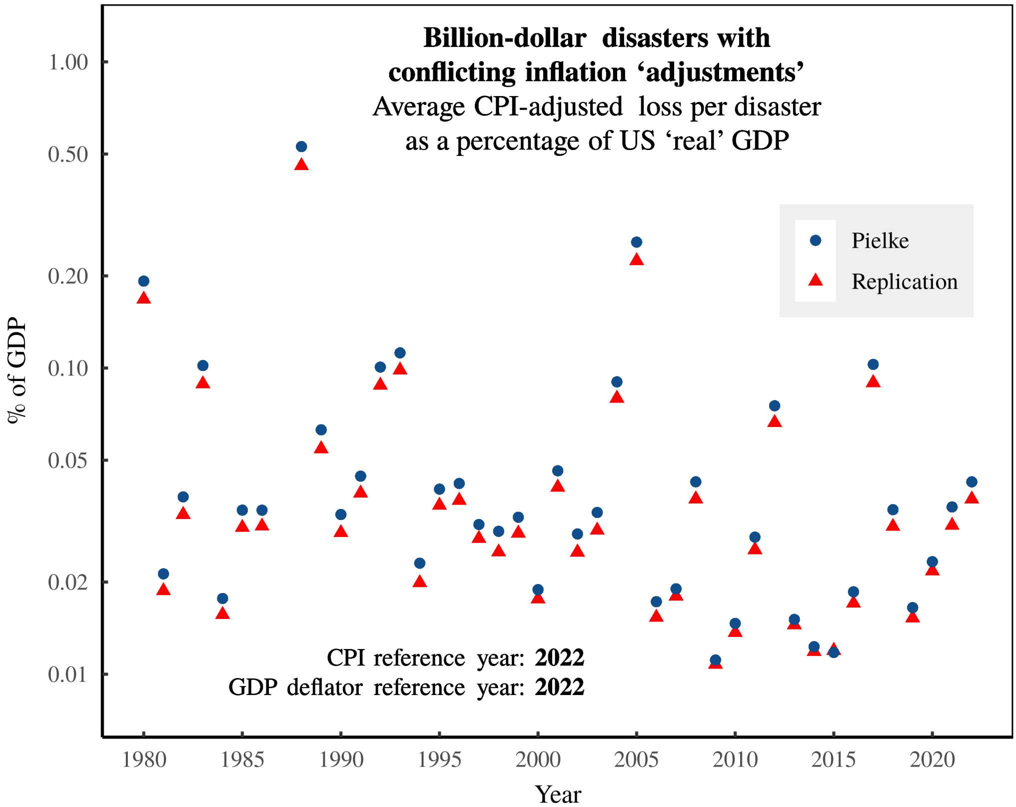


Figure 15: A first attempt at replicating Pielke's results

Blue circles show Pielke's published results (Figure 1). Red triangles show my replication, which uses the billion-dollar-disasters dataset version 209268.13.13. Notice the systematic error — my results are on average about 10% below Pielke's values. Sources and methods

---

With my canonical GDP data, I set about trying to replicate Pielke's work with several relevant BDD versions. The best-fit data comes from a BDD version numbered `209268.13.13`, which was published in early 2023.

Figure 15 shows how my attempted replication compares to Pielke's published data. (Blue circles show Pielke's data. Red triangles show my replication.) Oddly, the replication suffers from a systematic error: on average, my replication data sits about 10% below Pielke's values. I wonder why?

Pielke's paper gives a clue about what went wrong. Pielke claims to have used FRED series RGDPNAUSA666NRUG as the source of his 'real' GDP data. Now, in addition to ending prematurely (in 2019, rather than in 2022), this GDP time series uses a reference year of 2017. (The reference year is the point when 'real' GDP equals nominal GDP.) Now, as far as I can tell, the CPI-adjusted BDD data that Pielke downloaded uses a reference year of 2022, which is incompatible with the GDP reference year of 2017.

Taking a clue from this apparent error, if I *mismatch* the GDP reference year in my replication of Pielke's work, I get a much better fit to his data. Figure 16 illustrates. Here, I've adjusted my GDP data to have a reference year of 2018 — a reference year which is incompatible with the CPI reference year of 2022 used for the disaster data. By including this error, the fit with Pielke's data is greatly improved.

Looking at Figure 16, I'd guess that the remaining replication discrepancy owes to Pielke's different choice of GDP data. But since I'm not sure which data he actually used, I can't verify this suspicion.

Missing p-values and odd regressions

Looking at Pielke's published chart (my Figure 1), a simple eyeball test suggests that a linear trend is a poor fit to the data. More rigorous statistics confirm this suspicion.

Figure 17 illustrates. With my digitized version of Pielke's data, I find that the linear regression has an R^2 of 0.07 and a p-value of 0.1. For his part, Pielke did not report these statistics in his published work. This omission is both perplexing and inexcusable. If Pielke *had* reported these statistics, competent peer reviewers should have balked at his interpretation of the evidence. And if these statistics were *absent* from Pielke's manuscript, competent peer reviewer should have looked at the sloppy regression and requested summary statistics. Either way, Pielke's interpretation of his chart should *not* have survived peer review.

Now, the odd thing is that this situation could have been avoided had Pielke simply run a log-linear regression. As Figure 18 illustrates, running a log-linear regression on Pielke's data gives a much more legitimate trend line, with summary statistics that could survive peer review.

Importantly, this switch to a log-linear regression is not just p-hacking. It's an objectively better choice of regression. For starters, the linear regression implies that future disasters could have losses which are *negative*, meaning that on average, disasters *create* value. That's obviously nonsensical. (By definition, a log-linear regression avoids this problem, since its values cannot be negative.)

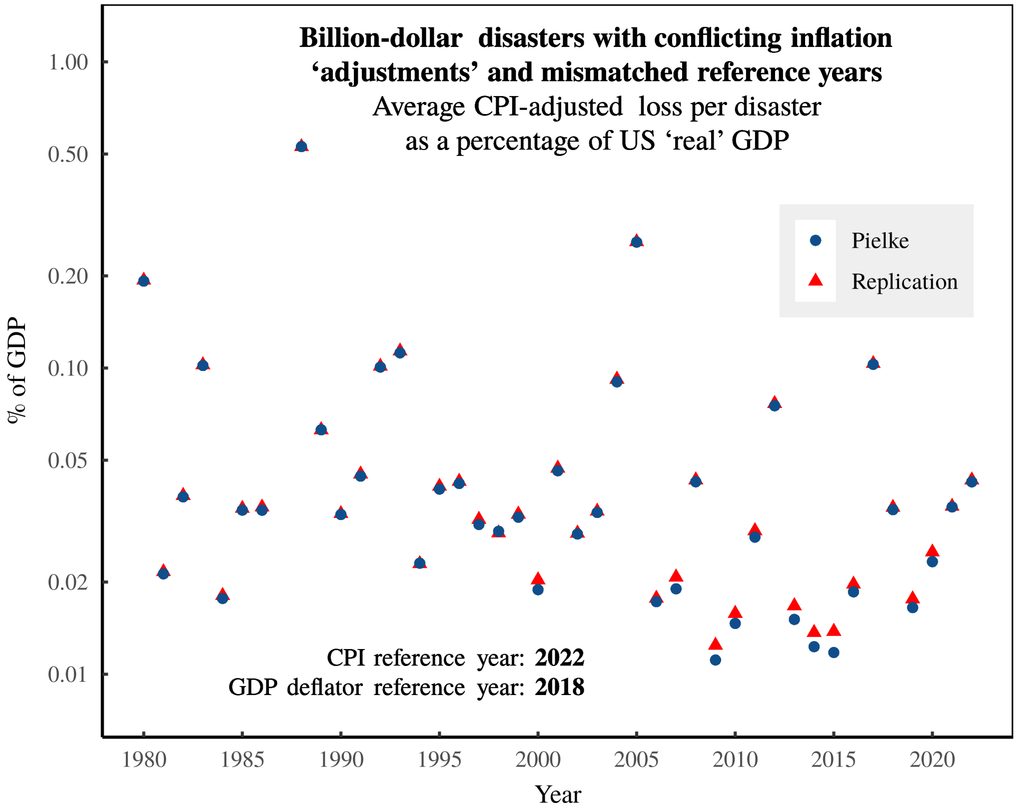


Figure 16: A second attempt at replicating Pielke's results

Blue circles show Pielke's published results (Figure 1). Red triangles show my replication, which uses the billion-dollar-disasters dataset version 209268.13.13. Unlike in Figure 15, this replication purposefully mismatches reference years used for inflation adjustment. Disaster costs are indexes with a CPI reference year of 2022. But GDP is indexed with a reference year of 2018. This error significantly improves the fit with Pielke's data. Sources and methods

---

Second, the log-linear regression has better behaved residuals. (Residuals are the vertical distance between the data and the regression line). Consider that a good regression should have residuals which are normally distributed. Well, a log-linear regression (of Pielke's data) meets this criteria. A linear regression does not.

Figure 19 illustrates, using the Kolmogorov–Smirnov test of distribution similarity. Here, each panel shows the cumulative distribution of regression residuals. (The top panel shows a linear regression, the bottom, a log-linear regression.) The blue curve shows the *empirical* residuals. And the red curve shows the best-fit *normal distribution* (where μ is the mean of the empirical residuals, and σ is the standard deviation of empirical residuals).

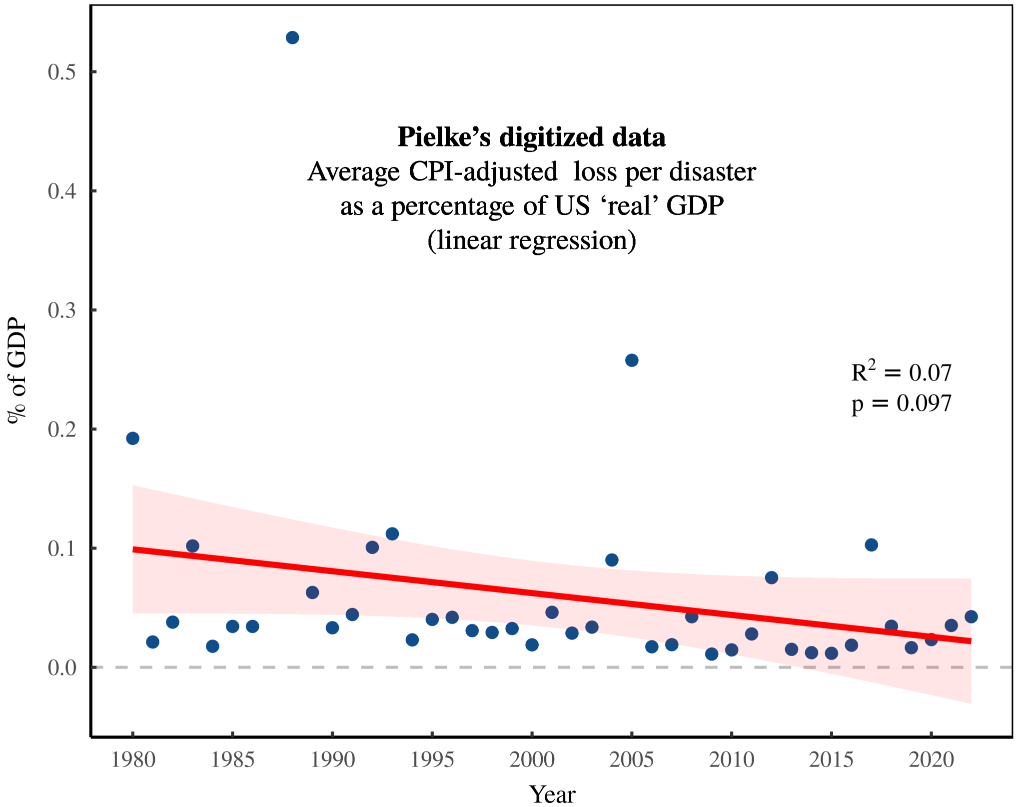


Figure 17: Retroactively insignificant

Here, I run a linear regression on my digitized version of Pielke's data. The results confirm that the linear regression is exceedingly loose. It has an R-squared value of 0.07 and a p-value of 0.1. Had Pielke published these summary statistics, competent peer reviewer's should have balked at his interpretation of the evidence. Actually, competent peer reviewer should have eyeballed Pielke's loose regression and requested summary statistics. Either way, Pielke's existing chart should not have survived peer review. Sources and methods

---

The Kolmogorov–Smirnov test consists of finding the greatest distance (D) between the empirical and theoretical curves. The larger this distance, the *less* the regression residuals follow a normal distribution. For our linear regression, we get $D = 0.24$. For a log-linear regression, we get $D = 0.15$. Clearly, the latter regression is superior.

The math behind the billion-dollar threshold effect

Here's the math behind the billion-dollar threshold effect. First, we define the default loss threshold, T_0, as \$1 billion:

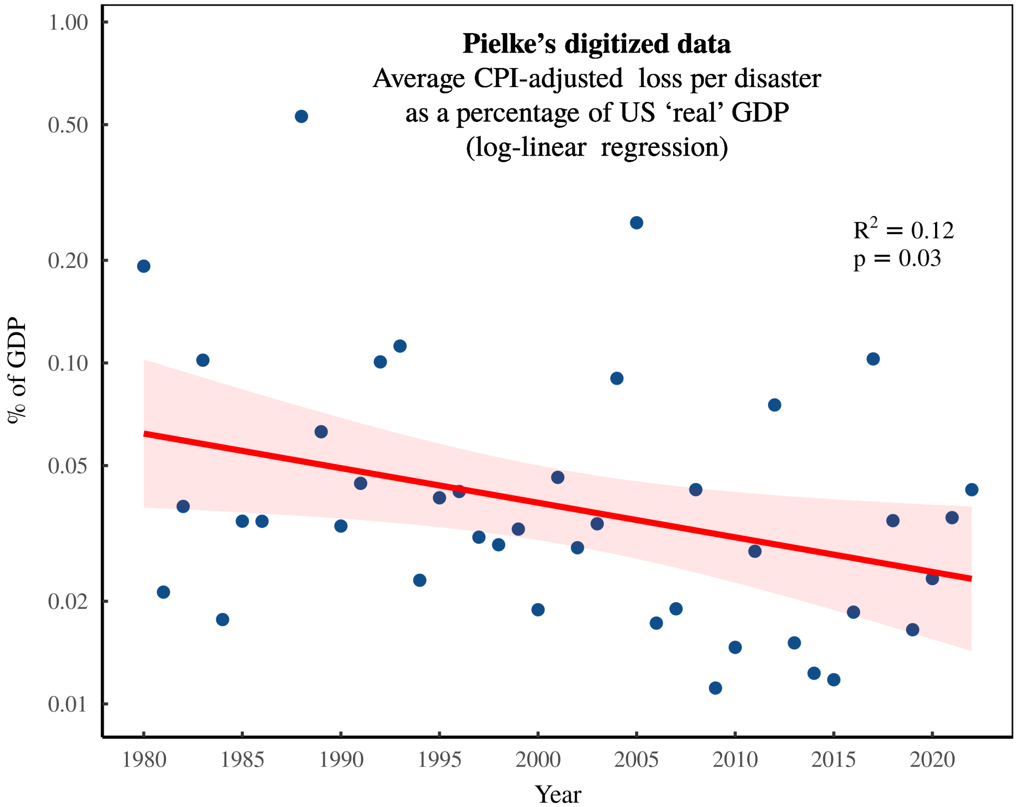


Figure 18: Pielke's data fit with a more appropriate log-linear regression

This chart has identical data to Figure 17, but uses a log scale on the vertical axis, and a log-linear regression to measure the trend Pielke's data. The summary statistics indicate that the log-linear regression is a much better fit to the data. Sources and methods

---

$$T_0 = \$1 \text{ billion}$$

Next, we calculate the inflation-adjusted cost threshold in year y by indexing the default threshold to the consumer price index:

$$T_y = T_0 \cdot CPI_y$$

To calculate the billion-dollar threshold effect, we then divide the indexed threshold, T_y, by nominal GDP:

$$\text{threshold effect} = T_0 \cdot \frac{CPI_y}{GDP_y}$$

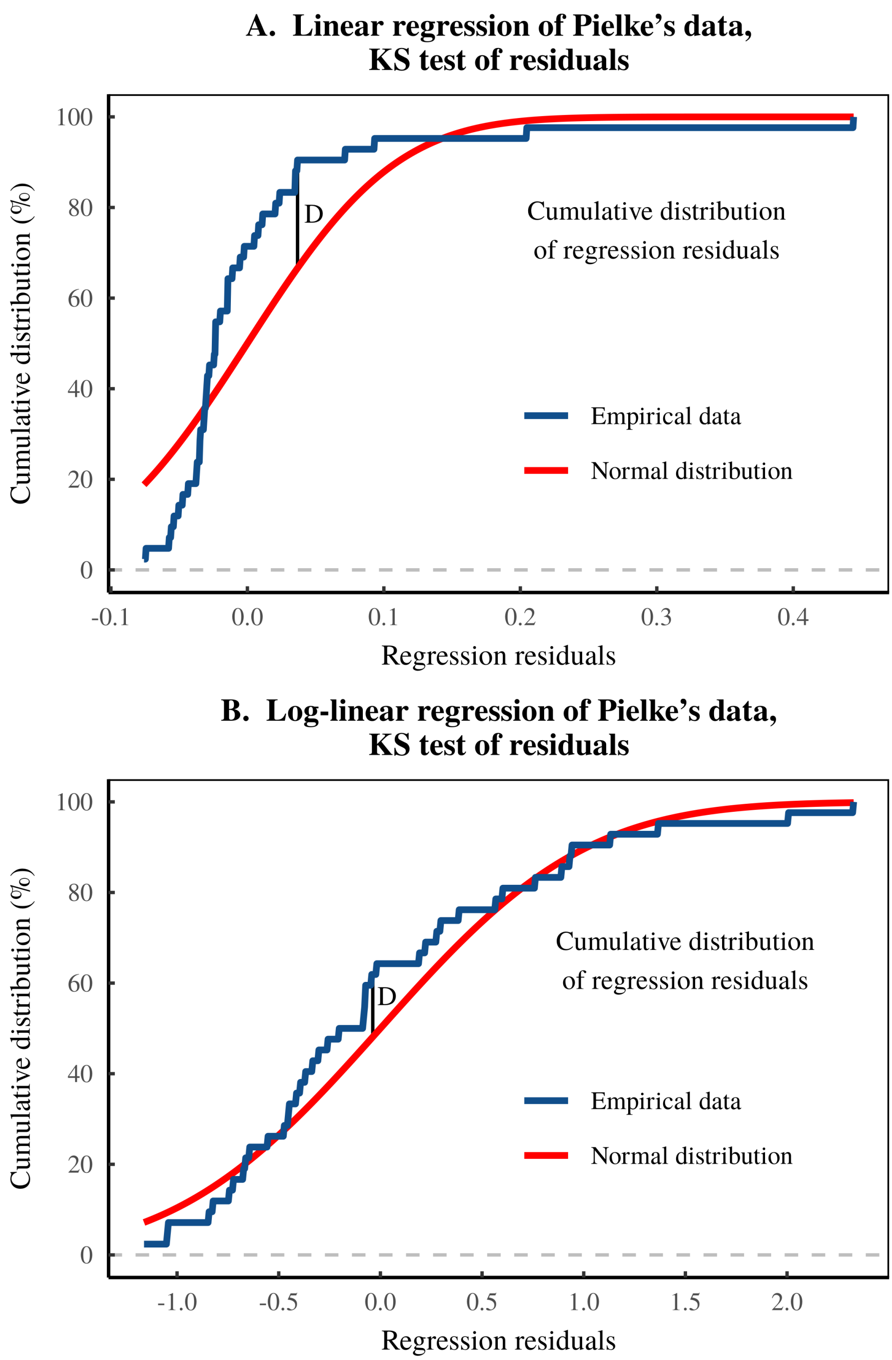


Figure 19: Testing for normality — a KS test of regression residuals

This chart shows how we can use the KS test to determine whether Pielke's data is better fit by a linear regression (panel A) or a log-linear regression (panel B). The KS works by first constructing the cumulative distribution of the regression residuals. Next, we compare the empirical distribution to a best-fit normal distribution. The KS test then consists of measuring the greatest distance (D) between the empirical and theoretical curve. The smaller this distance, the closer the residuals are to a normal distribution. In this case, we find that the log-linear regression is the better model since it delivers residuals that are closer to the normal distribution. Sources and methods

Looking at this math, the size of the threshold effect is determined by the changing ratio between the CPI and nominal GDP.

To distort this threshold affect with conflicting price indexes, we multiply the equation above by the ratio of the GDP deflator (GDPD) to the CPI:

$$\text{distorted threshold effect} = T_0 \cdot \frac{CPI_y}{GDP_y} \cdot \frac{GDPD_y}{CPI_y}$$

Interestingly, the CPI cancels out, leaving:

$$\text{distorted threshold effect} = T_0 \cdot \frac{GDPD_y}{GDP_y}$$

The message here is that Pielke's use of mismatched price indexes effectively rewrites the billion-dollar-disaster data to behave as though its original cost threshold was tied to the GDP deflator (instead of the CPI). And since the GDP deflator moves less steeply than the CPI (see Figure 10), the effect is to bolster the original threshold effect.

Disaster roulette

An important feature of natural disasters is that the financial losses are concentrated in events which are rare but severe. Or put another way, the distribution of disaster losses has an extremely fat tail.

Figures 20 and 21 illustrate this principle using the Storm Events Database, maintained by the NOAA. Figure 20 shows the distribution of storm losses over the last quarter century (with the damage of each storm pegged against US GDP in the appropriate year). Note the log scale on both axes. The shape of this distribution indicates that the vast majority of storms are small, typically causing damage that's less than a 100-millionth of GDP. And yet there is a long tail of rare but destructive storms.

The shape of this distribution leads to counter-intuitive effects. For example, although most storms are small, almost all of the damage is created by a few large storms. Figure 21 illustrates this principle using a stylized Lorenz curve.

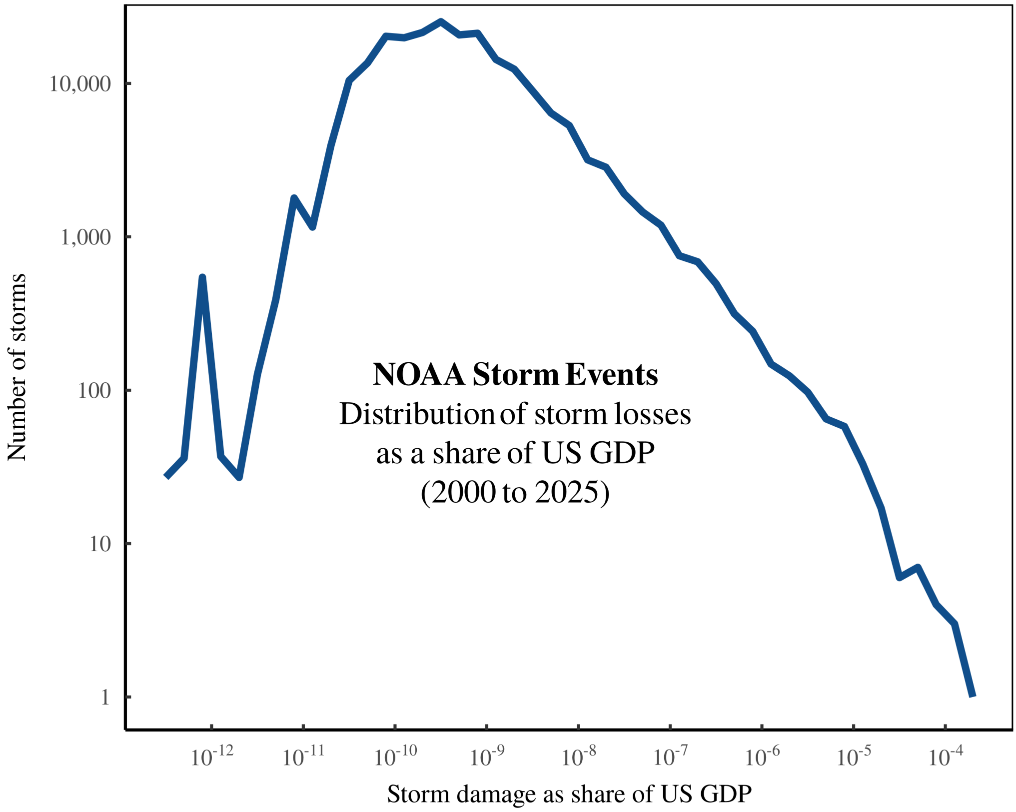


Figure 20: The size distribution of US storm losses

This chart measures the size distribution of storm losses in the NOAA Storm Events Database. The horizontal axis shows the scale of individual storm losses, measured relative to US GDP. The vertical axis shows the frequency of such storms. Note the log scale on both axes. Sources and methods

---

Typically, a Lorenz curve is used to measure income concentration, and plots income share against income percentile. Translating this thinking to storm losses, Figure 21 plots the cumulative share of storm losses against storm percentile. (Storm losses are again measured against US GDP.) Note that because storm losses are so concentrated, I've used a horizontal axis with a stylized log scale that gradually zooms into the top percentiles. What we see here is that the vast majority of damage is created by the top 1% of storms. Stranger still, about 50% of damage is cause by the top 0.05% of storms.

Thinking about this extremely skewed distribution, it has important consequences for how scientists estimate disaster losses. Suppose, for example, that budget constraints lead to a statistical trade off between database *scale* and estimate *accuracy*. If scientists opt for database scale (by studying all events), then their loss estimates per disaster are fairly loose. But if scien-

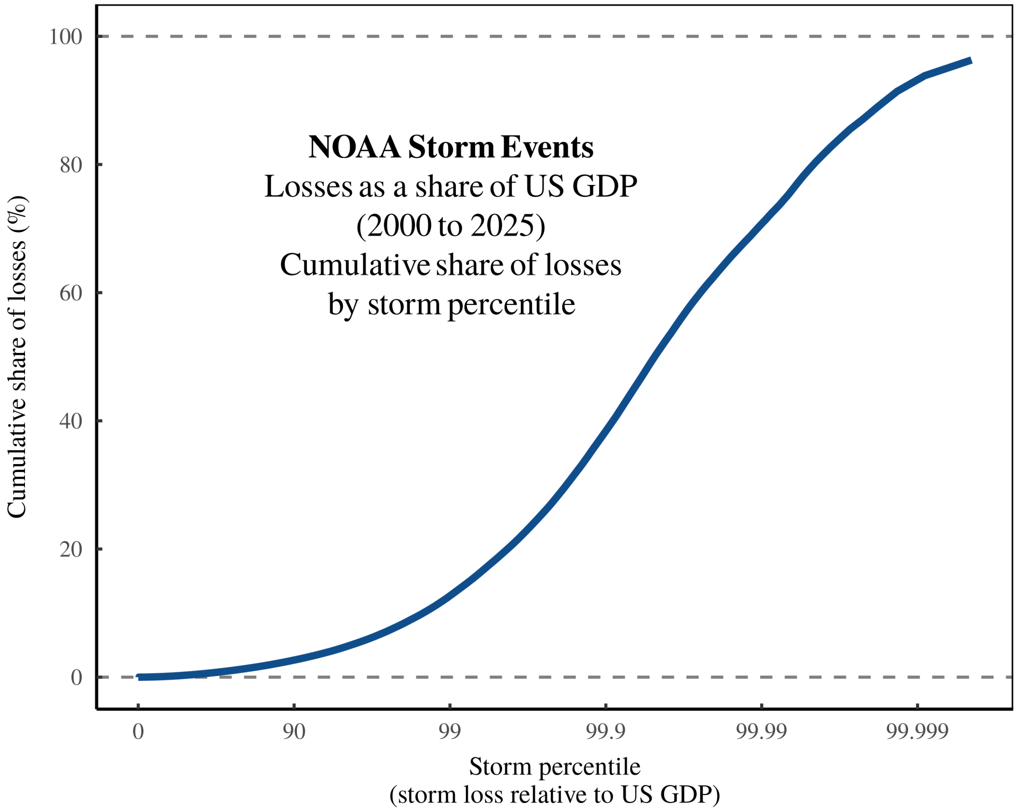


Figure 21: The concentration of storm losses by storm percentile

This chart shows a modified Lorenz curve for storm damage. The horizontal axis shows storm percentile, ranked by storm loss as a share of US GDP. The catch is that because storm losses are so skewed, I've used a stylized log scale that zooms into the top percentiles. The vertical axis shows the cumulative distribution of storm losses. Sources and methods

---

tists study a small sample of events, then their accuracy per disaster is much better. Given this trade off, which method is most accurate for estimating total damages?

Well, the answer depends on the exact trade off between scale and per-event accuracy. But let's imagine the following numbers. Suppose that if scientists attempt to quantify all natural disasters, their average per-event error is 30%. If, however, they sample only 5% of all disasters, then their average per-event error improves to 3%.

Next, let's suppose that the NOAA's storm events database is the hypothetical source of truth, revealing the actual cost of natural disasters. As such, let's see how our two methods perform.

Figure 22 illustrates several scenarios. In the top panel, we imagine that our scientists compile two different estimates. The first estimate (red violin) tracks the damage for *all* storms with an average estimate error of 30% per storm. The second estimate (blue violin) tracks damages for a random sample of 5% of storms with an average estimate error of 3% per storm. Each 'violin' then shows the distribution of error in the resulting estimate for total storm damage.

What we find here is that despite the improved accuracy per storm, the random sampling method performs *horribly*, giving estimates for total damage that are often wrong by an order of magnitude. This horrendous error owes to the skewed nature of storm losses. Since most of the damage is caused by a few large storms, a small random sample has a low probability of capturing the most important events. Hence inferring total storm damage from such a sample is inadvisable.

Realizing this problem, our scientists try a second method, shown in the bottom panel. Instead of sampling randomly, our scientists select the top 5% of storms, and measure their losses with an average error of 3%. The result is a much more accurate estimate. Indeed, this top-sampling method is even more accurate than the scale method of measuring the losses from all storms. (True, the top sampling method tends to undercount total losses. But since this underestimate is fairly consistent, results could be adjusted upwards to compensate.)

This thought experiment illustrates why the billion-dollar-disasters database is wise to focus on the accurate measurement of large disasters. The trick, though, is to apply this large disaster focus without generating unwitting threshold effects. It's a difficult task, with no easy solutions.

Details of the Figure 22 model

- Hypothetical 'true costs' consist of losses per storm as a share of US GDP, as measured by the NOAA Storm Events Database covering the years 2000 to 2025.
- Loss 'estimates' are generated by multiplying 'true costs' by a random error, generated from a truncated normal distribution with a mean of 1 and a lower bound of 0.
- When studying all storms, the error distribution has a standard deviation of 0.3; when studying the sample of 5% of storms, the error distribution has a standard deviation of 0.03.

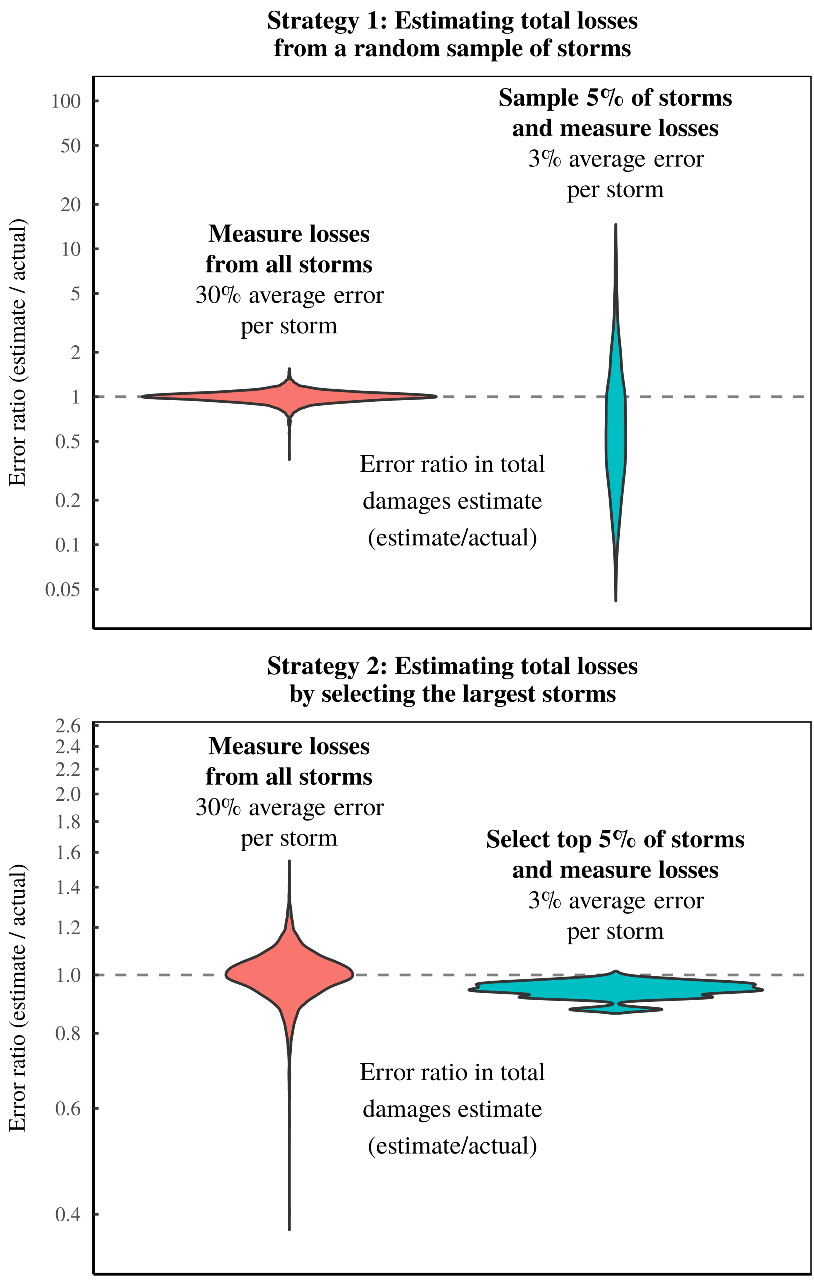


Figure 22: Focusing on the largest disasters is potentially the most accurate method for estimating total losses

This chart illustrates a thought experiment in which we investigate different methods for measuring total storm losses. We begin by assuming that the NOAA storm events database is the source of 'truth'. Then we imagine a trade off between database scale and estimate accuracy per storm. Studying all storms leads to a 30% average error per storm. In contrast, studying just 5% of storms improves the average error per storm to 3%. In the top panel, we compare the scale approach to a random sample of storms. In the bottom panel, we compare the scale approach to a sample of the largest storms. The latter approach is far more accurate, better even than the scale method. The message is that the billion-dollar-disasters dataset is justified in focusing on the losses from the largest natural disasters. Sources and methods

- For the random sample, we assume that total storm damage is 25 times the sample estimate; for the top storm sample, we do not adjust for the sample size.
- In our simulated storm sample, we estimate total damage in each year between 2000 and 2025. Next we measure the error in the annual estimates. Finally, we repeat this whole operation 200 times and measure the resulting distribution of error.

Sources and methods

All data and code for my analysis are available at the Open Science Framework: https://osf.io/jqu8v

Billion-dollar disasters

Archived versions of the billion-dollar-disasters (BDD) dataset are available here:

https://www.ncei.noaa.gov/archive/archive-management-system/OAS/bin/prd/jquery/accession/download/209268

For my exposition of Pielke's method (Figures 2 to 14), I use the most recent published version of the dataset, labelled `209268.21.21`. Figures 2 to 7 use data for 'unadjusted costs'. Figures 11 to 14 use data for 'CPI-adjusted cost'.

For my direct replication of Pielke's results, I've determined that he likely used the BDD version labelled `209268.13.13`.

US GDP

US GDP data comes from the following series:

- US nominal GDP: FRED series GDP
- US GDP deflator: annual data from FRED series A191RD3A086NBEA; quarterly data from FRED series A191RI1Q225SBEA

Most of the charts match disaster loss data with annual GDP (since this is the method used by Pielke). The exceptions are Figures 2 and 3, which, for added fidelity, matches disaster losses with quarterly GDP. (In these charts, the billion-dollar threshold is also calculated using quarterly data for the GDP deflator and the CPI.)

I calculate 'real' GDP by indexing nominal GDP to the GDP deflator. Note: the reason I don't use 'real' GDP data directly is because I need to adjust the GDP reference year to match the CPI reference year used in the natural disaster data.

Consumer price index

Data for the US consumer price index is from the following series:

- 1947 to present: FRED series CPIAUCSL
- 1929 to 1947: Historical Statistics of the United States, series Cc1

Note: I use this CPI data to calculate the billion-dollar threshold effect (Figures 2, 3, 8, 12, and 14), and to illustrate the mismatch between the CPI and the GDP deflator (Figures 9 and 10). For CPI-adjusted disaster costs, I use the NOAA's own calculations.

NOAA storm events

Historical data for NOAA storm events is available for bulk download here: https://www.ncei.noaa.gov/pub/data/swdi/stormevents/csvfiles/

My loss estimates (Figures 20 and 21) use the sum of crop and property damage.

Further reading

Keen, S. (2021). The appallingly bad neoclassical economics of climate change. *Globalizations*, *18*(7), 1149–1177.

Pielke Jr, R. (2024). Scientific integrity and US "billion dollar disasters." *NPJ Natural Hazards*, *1*(1), 12.

Smaldino, P. E., & McElreath, R. (2016). The natural selection of bad science. *Royal Society Open Science*, *3*(9), 160384.