Economics from the Top Down

new ideas in economics and the social sciences

Masochistic Fun with Plutocratic Murder

Blair Fix

May 20, 2023

There's nothing like waking up to a boatload of Twitter scorn. It's refreshing, in a masochistic sort of way.

Some backstory. After most of my blog posts, I put the charts on Twitter, usually with a provocative caption. (It's more fun that way.) So after last week's review of Cory Doctorow's book *Red Team Blues*, I tweeted the relation between income inequality and the rate of murder:

The next morning my Twitter feed was … interesting. The tweet apparently went viral, helped along by a nice little insult from Nassim Nicholas Taleb:

> No. The statement by that fettuccinibrain is contradicted by his own graph. #fooledbyrandomness.

Apart from him calling me a fettuccini brain, I have a lot of respect for Taleb, who's famous for (among other things) his excellent book *Fooled By Randomness*. It's an engaging romp through the many ways that human's get duped by random patterns.

Back to the tweet. Taleb is implying that the relation between inequality and murder is essentially a random blob of data, and that I'm a fool to plot a trend line through it. I disagree.

Part of the problem with the Twitter chart is that the country labels add visual noise to the plot — noise that's not part of the data. This is a problem with no good solution. If I plot the data without labels, people hound me for not labeling countries. And if I add lots of labels (as I did here), it muddies the plot. It's a Catch-22.
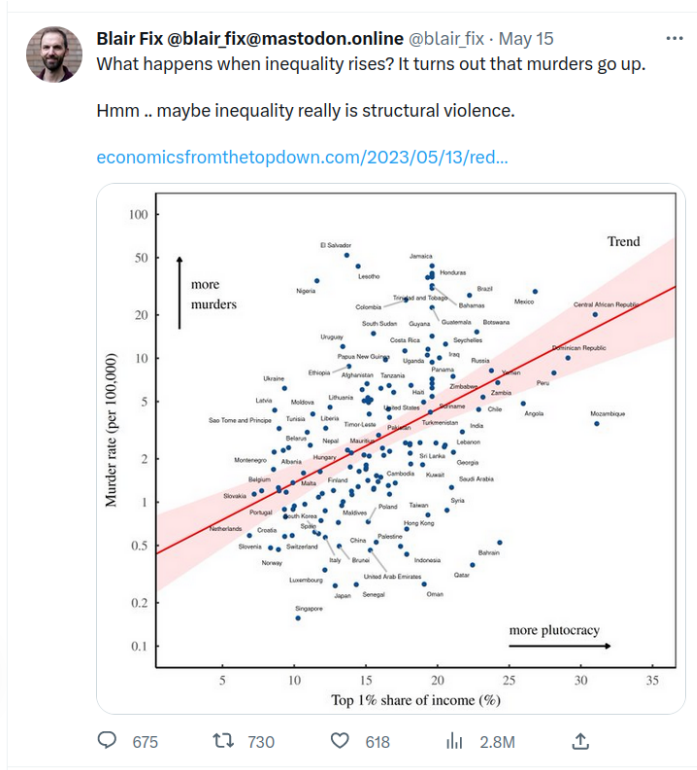
But for the sake of thoroughness, let's go ahead and look at the unlabeled data, plotted in Figure 1.

Now obviously the relation between inequality and murder rates is not super tight. So you wouldn't want to claim that inequality is the sole determinant of murder. (No one claims that.) But then again, if you want to explain homicide rates, the evidence suggests that it would be silly to exclude income inequality as a possible cause.

Now to the fooled-by-randomness claim. Is the data in Figure 1 produced by random noise? Well, the problem is that we can never say for certain. All we can do is run statistics and take an educated guess.

The standard approach is to calculate the p-value of the regression. To do that, we assume that the null hypothesis is true — that the observed data is drawn from two independent normal distributions. Given this null hypothesis, the p-value then tells us the probability of getting an outcome as extreme as the one observed.
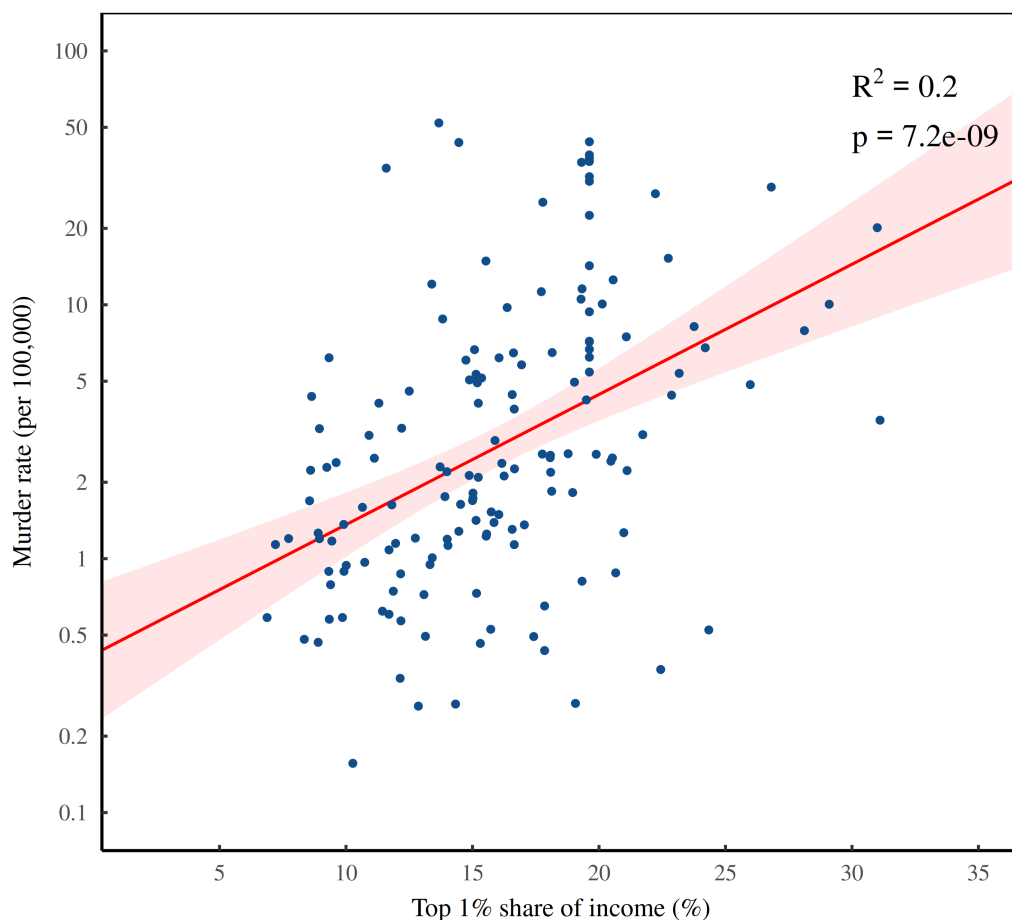
Figure 1: The murder rate vs. the top 1% share of income across countries

In the case of the data in Figure 1, the *p*-value is about 1 in 100 million. In other words, it's not particularly likely that chance alone would produce the observed results.

The wrong model?

Another criticism levied in the Twitter melee is that I'm using the wrong model. You see, the thing about regressions is that you can fit any model to any data. But that doesn't mean you should.

Fortunately, there are alternatives to models. A simple option is to put the data into bins, and then see what kind of trend emerges.
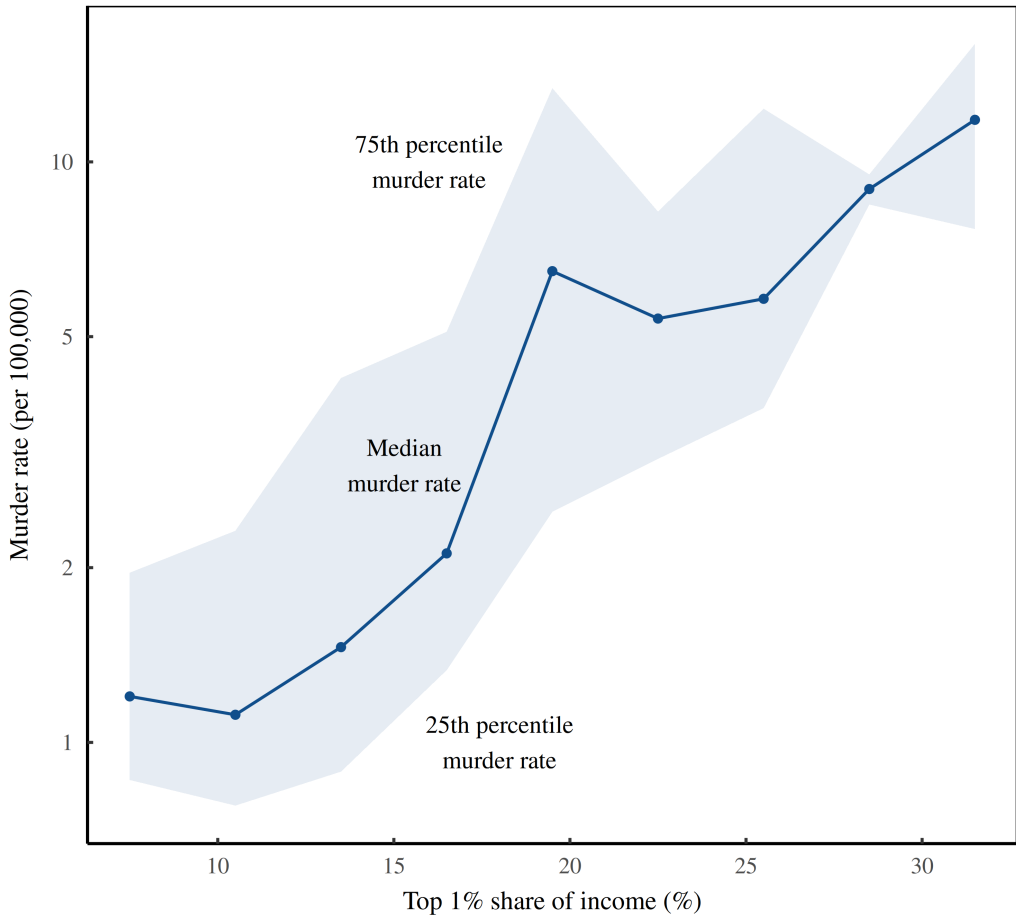
Figure 2: A model-free analysis of murder rates and inequality
This figure shows the cross-country trend between murder rates and the top 1% share of income. Each blue point indicates the midpoint of an income-share bin. The line shows the median murder rate within each bin. The shaded region indicates the middle 50% of data.
Sources and methods

I've done that in Figure 2. Here, I've binned the international data by levels of income inequality. Each blue point shows the midpoint of an inequality bin. The line and shaded region then show the murder-rate trend across countries. The pattern seems to be ... drum roll ... a straight line. In short, it seems entirely reasonable to fit this data with a linear regression (on the log of the murder rate).

Log scales ... how dare you!

Speaking of logs, another Twitter accusation is that I'm a fool for plotting murder rates on a log scale. (In Figures 1 and 2, the vertical axis has tick marks that correspond to powers of 10.) The idea is that the log scale — and the trend plotted on it — somehow misleads.

Nonsense.

A feature of murder rates is that they vary over an enormous range — from a low of 0.16 murders per 100,000 (in Singapore) to a high of 52 murders per 100,000 (in El Salvador). That's a 330-fold difference.

When you're dealing with such an enormous range, a linear scale is misleading because it equates constant changes. For example, suppose that across all countries, homicides rates increased by 1 murder per 100,000 people. In El Salvador, that corresponds to a 2% increase in the murder rate. But in Singapore, the same change corresponds to a 640% jump in murders.

In contrast, a log scale puts percentage change on the same footing. So if murder rates grow by 5% across all countries, a log scale will show constant change. Given that we tend to think about murder rates in terms of percentage change, a log scale seems appropriate.

And when it comes to regressions on the murder-rate data, a log scale is *non-optional*. You see, a key assumption of regression analysis is that the data should be normally distributed. When we look across countries, murder rates do follow a normal distribution ... but only when we observe them on a *log scale*, as shown in Figure 3. (This pattern means that across countries, murder rates have a roughly log-normal distribution.) In short, running a regression on the log of murder rates is the right thing to do.

The dance of the *p*-values

Returning to Taleb's fooled-by-randomness quip, lets take a deeper dive into *p*-values.

A big problem with *p*-values is that they're not particularly reliable. A slight change in the data can yield a drastic change in the *p*-value — a fact that's nicely demonstrated by Geoff Cumming in his video 'The dance of the *p* values'.
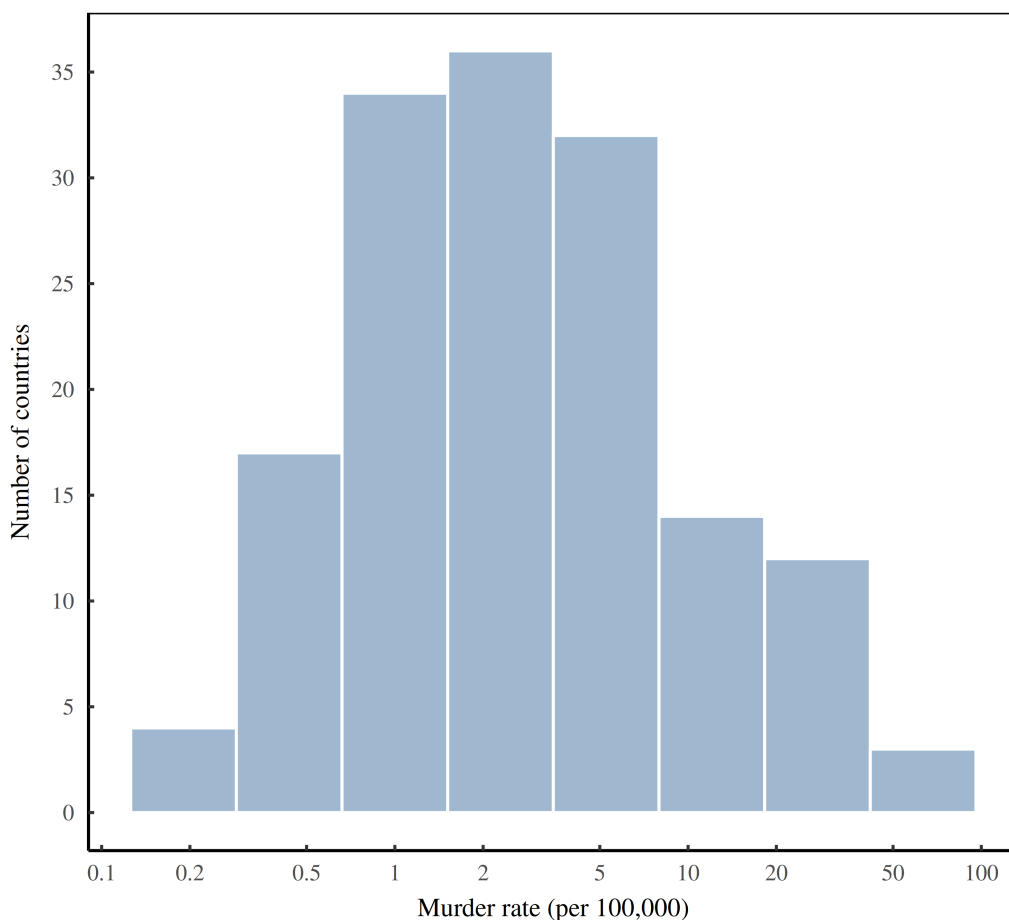
Figure 3: Across countries, murder rates are log-normally distributed
This figure shows the distribution of country-level murder rates. The distribution has a bell curve when plotted on a log scale, meaning cross-country murder rates roughly follow a log-normal distribution. Sources and methods

Looking at the trend between murder rates and inequality, the regression *p*-value is pretty low (which suggests that randomness is unlikely to produce the observed result). But perhaps our low *p*-value is an accident, driven by a few outliers.

To quantify how accidental our *p*-value might be, we can do a bootstrap analysis. This means we sample (with replacement) our murder-inequality data and see what kind of relation we get in our sample. In this case, I take the sampled data and regress the log of murder rates onto income inequality. Then I measure the *p*-value of the regression. Finally, I repeat the whole operation with a different sample.
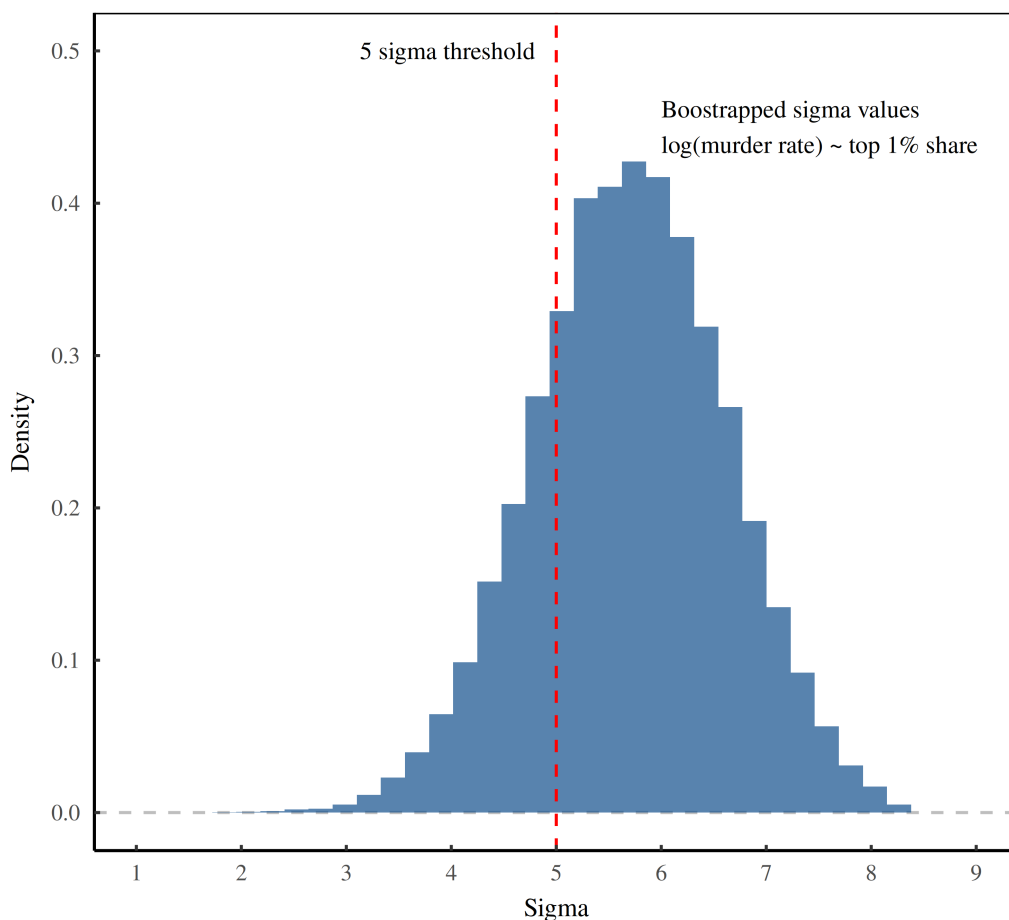
Figure 4: Bootstrapping statistical significance

This figure shows the results of a bootstrap analysis — 50,000 random samples from the murder-inequality data plotted in Figure 1. The histogram shows the distribution of statistical significance (in units of sigma) for a regression between the log of the murder rate and the top 1% share of income. About 78% of the time, the results pass the 5 sigma threshold — the bar for 'discovery' in particle physics. Sources and methods

The histogram in Figure 4 shows the results of 50,000 bootstraps. Note that I've expressed the *p*-values in terms of *sigma* — something that's convenient when looking at extreme values. Each sigma corresponds to one standard deviation in the null hypothesis. So a 5 sigma result is an event that is five standard deviations from the null hypothesis's expected value.

Speaking of 5 sigma, this is the standard of statistical significance used in physics. In other words, if an experimental result passes the 5 sigma mark, physicists take it seriously. Of course, thresholds for statistical significance

are always arbitrary, and the 5 sigma mark is no different. Still, if it's good enough for physics, it should be good enough for the social sciences, where data is far poorer.

With the 5 sigma mark in mind, let's look at the histogram in Figure 4. As expected, our sigma levels bounce around as we resample the cross-country data. Still, about 78% of the time, we get a value that exceeds 5 sigma. And 99.8% of the time, we get a result that exceeds the less stringent 3 sigma threshold — a level that will usually get your social-science research published.

Now in general, I don't like appealing to statistical significance to defend scientific results. (Statistical significance says nothing about scientific significance.) Still, a bootstrap analysis can tell us if we're torturing the data.

For example, if the relation between murder and inequality was driven by a few outlier countries, then removing these countries would turn our results to mud. But the flip side is that if we're going to play the 'outlier' game, we have to be impartial about it. In other words, we also have to consider 'outliers' that, when removed, *strengthen* the empirical trend.

Our bootstrap analysis shows how this outlier game cuts both ways. Sure, we can ruin the murder-inequality relation by randomly removing some countries. But we can also greatly improve the relation — turning a 5 sigma result into an 8 sigma one.

So if the Twitter hoards want to call 'outlier', they've got to do it in both directions. But then that would ruin their argument.

When the 'fluke' doesn't disappear

Enough reanalysis. If an observational result is actually a random fluke, the only fool-proof way to find out is to gather more data.

Let's do that now.

The World Bank has an extensive dataset on what it calls 'intentional homicides'. (The data is compiled by the UN Office on Drugs and Crime.) Merging this World Bank data with the World Inequality database, I get cross-country murder-inequality data that goes back to 1990.
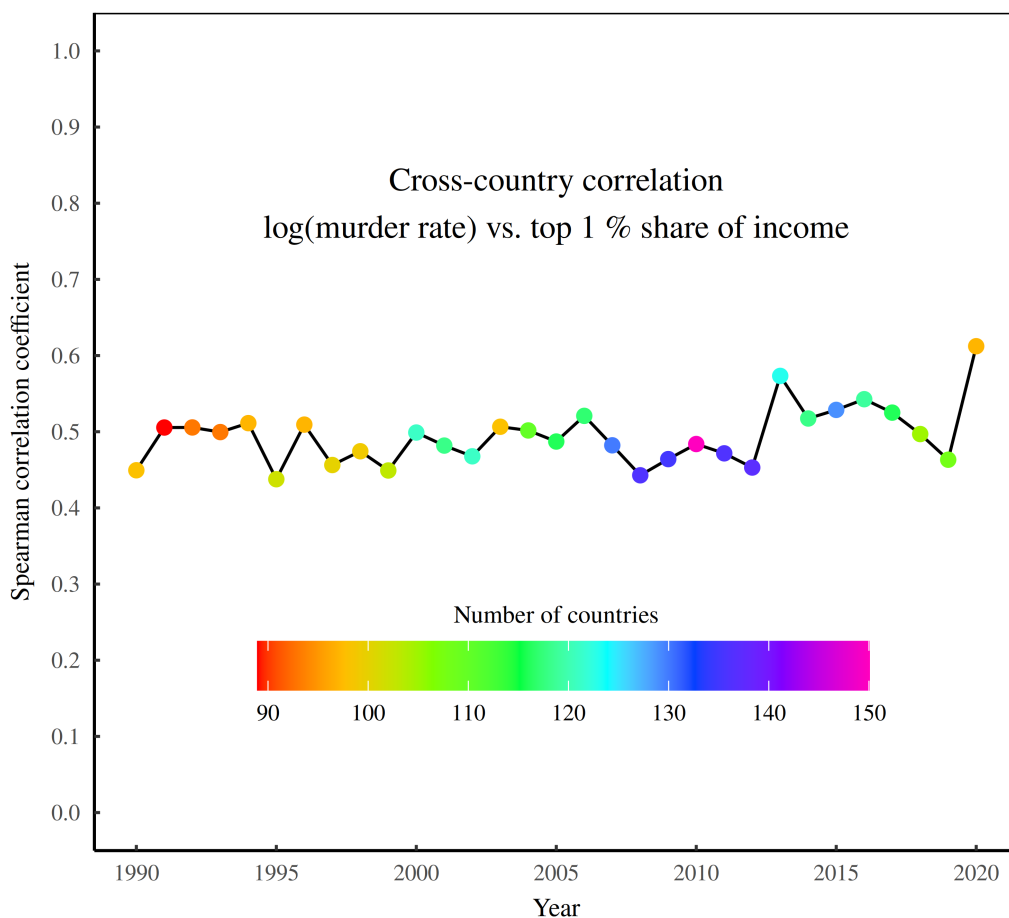
Figure 5: Not a fluke correlation

This figure analyzes the robustness of the cross-country correlation between murder rates and income inequality, as measured by the top 1% income share. Each point shows the Spearman correlation coefficient (measured across countries) between the log of the murder rate and the top 1% share of income. The horizontal axis shows the year of the observation. Color indicates the number of countries in the dataset. Sources and methods

Looking at this data, there's no evidence that the murder-inequality trend is a statistical fluke. Instead, I find that the relation is remarkably consistent across time

Figure 5 tells the story. Here, each point shows the Spearman correlation coefficient (measured across countries) between the log of murder rates and the top 1% income share. The horizontal axis shows the observation year. Color indicates the number of countries observed. Over three decades of data, the Spearman correlation oscillates around a value of 0.5.

Let's take stock. The Twitter hoards piled on me for posting a correlation between murder rates and inequality that they deemed to be little more than random mud. Nassim Nicholas Taleb helped things along by suggesting that I'd be fooled by randomness.

Maybe I have. But if so, why does the 'random' correlation replicate consistently for 30 years? It's almost like the murder-inequality relation is *not* random. It's almost like the data *is* telling us something.

Wait, there's more

At this point, I could probably stop. But let's keep going. When I decided to correlate murder rates with income inequality, it's not like I was pulling the idea out of nowhere.

Back in 2009, Richard Wilkinson and Kate Pickett published a stunning book called *The Spirit Level*, in which they documented the many social ills that correlate with income inequality. The murder rate was one of the conspicuous social bads.

Speaking of Wilkinson and Pickett, their book set off a cottage industry of criticism — some of it coherent, much of it silly. Sure, Wilkinson and Pickett didn't publish correlation coefficients or *p*-values, obviously because the book was meant for a lay audience. (Kind of like how I didn't put R^2 values in a chart designed for a book review.) But that doesn't mean their evidence lacked rigor.

Take the relation between income inequality and the murder rate. I've already shown you that across countries, there's a robust relation between the murder rate and the top 1% share of income. But the top 1% share is just one way to measure inequality. There are many others metrics, each of which highlights a different feature of the distribution of income.

Let's look at the most common measure of income inequality: the Gini index. Returning to the World Bank database, Figure 6 shows what happens when we plot the cross-country pattern between the Gini index and the murder rate. Let me spell it out for the Twitter trolls: the line goes up.

But maybe we're still getting fooled by randomness? Let's check by looking at how the gini-murder relation changes over time. Oh wait, it *doesn't* change.

Figure 7 runs the numbers, using the same method as in Figure 5. For being 'random', the inequality-murder relation sure seems to be consistent.
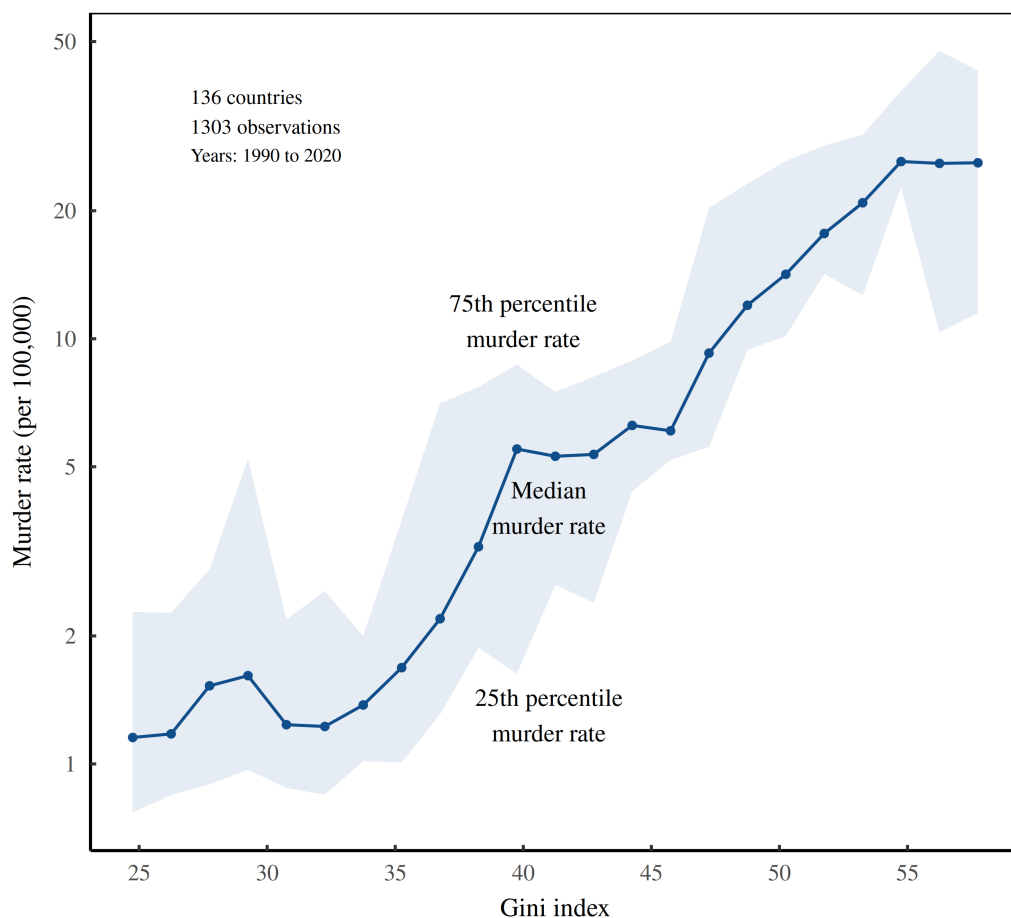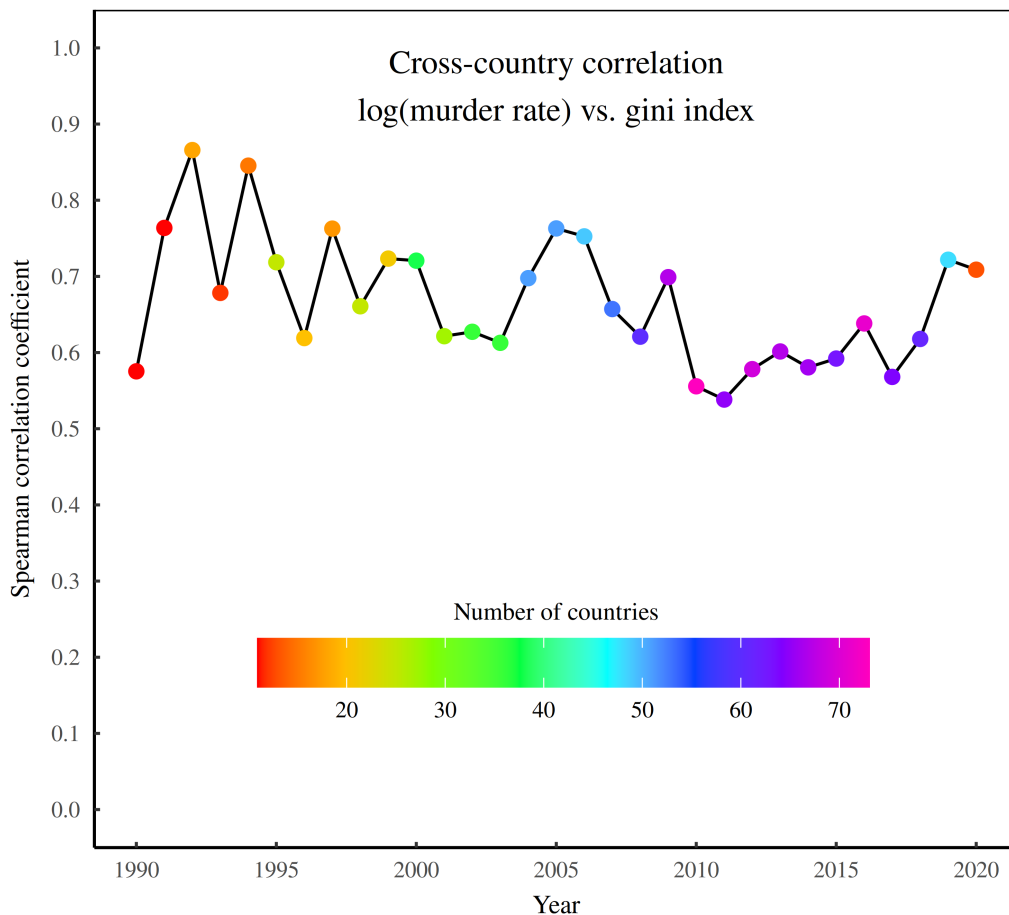
Figure 6: The cross country pattern between murder rates and the Gini index of income inequality

This figure is similar to Figure 2, but measures income inequality using the Gini index. Blue points show income inequality bins. The blue line shows the median murder rate within each bin, measured across countries. The shaded region shows the middle 50% of data. Sources and methods

Still more evidence

I'd like to stop now, but I keep thinking of different ways to measure income inequality. Here's one last go. Let's see how the bottom 10% share of income relates to murder rates.

First, let's set the stage. Unlike their counterparts in the top 1%, people in the bottom 10% generally take home a minuscule share of the income pie. They are the people that society has left behind — the ones most likely to resort to violence for survival. So it makes sense that if we improve the welfare of society's bottom strata, murder rates should go down.

Figure 7: Still not a fluke correlation

This figure analyzes the robustness of the cross-country correlation between murder rates and income inequality, as measured by the Gini index. Each point shows the Spearman correlation coefficient (measured across countries) between the log of the murder rate and the gini index. The horizontal axis shows the year of the observation. Color indicates the number of countries in the dataset. Sources and methods

And that's exactly what we find. Figure 8 shows the pattern across countries. As the income share of the bottom 10% increases, murder rates decrease.

And as before, the inequality-murder pattern is consistent over time. Figure 9 shows how the Spearman correlation coefficient (measured across countries) varied over the last three decades. It oscillates around a value of -0.6. (The correlation is negative because the relation between murders and the bottom 10% share is negative.)

If we're dealing with random noise, the evidence sure has a funny way of showing it.
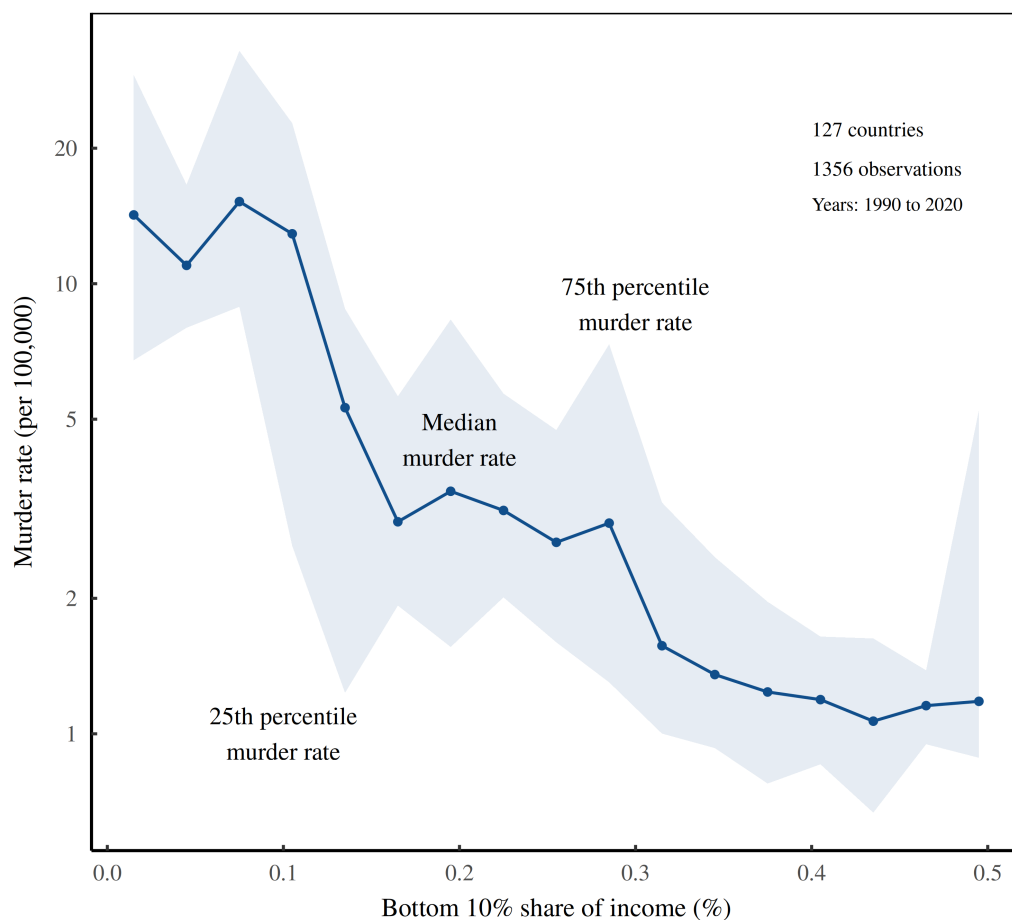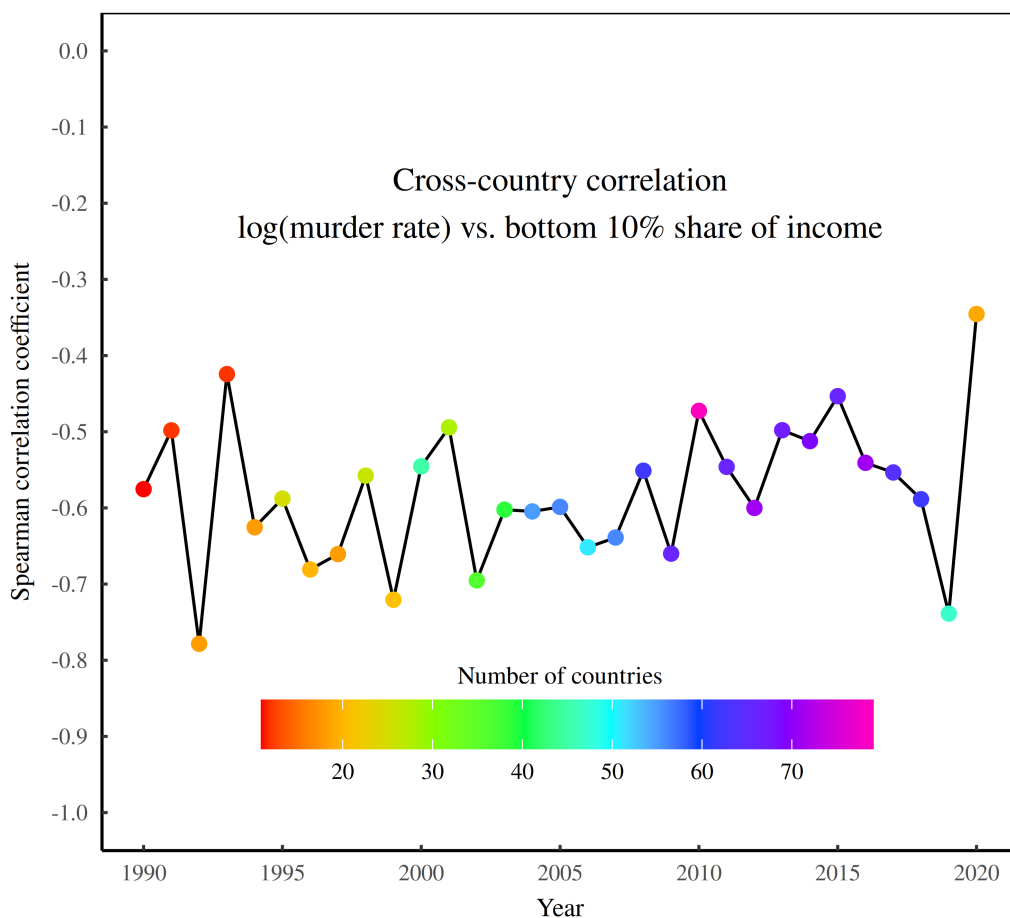
Figure 8: The cross country pattern between murder rates and the bottom 10% share of income

This figure shows the cross-country trend between the bottom 10% share of income and the murder rate. Blue points show income inequality bins. The blue line shows the median murder rate within each bin, measured across countries. The shaded region shows the middle 50% of data. Sources and methods

Interesting questions

Let's summarize this foray into inequality and murder. Given the evidence presented here, it defies reason to conclude that the relation between income inequality and murder rates is driven by 'random chance'. So let's throw that idea in the garbage.

Figure 9: Really still not a fluke correlation

This figure analyzes the robustness of the cross-country correlation between murder rates and income inequality, as measured by the bottom 10% of income. Each point shows the Spearman correlation coefficient (measured across countries) between the log of the murder rate and the bottom 10% share. The horizontal axis shows the year of the observation. Color indicates the number of countries in the dataset. Sources and methods

---

The real question is why the inequality-murder relation exists. And that's not easy to settle. It's not like inequality is some knob that we can turn, leaving other areas of society unchanged. No, inequality is a complex outcome of many changing social dynamics. In statistical parlance, that means the real world is a mess of confounders.

Still, there are some interesting questions to ask. For example, does the *type* of inequality matter? You see, there are different ways that societies can be unequal. In simple terms, you can either have inequality at the *top* of the distribution or at the *bottom*. (See this post for details.)

I suspect that inequality at the bottom is more socially corrosive. In other words, if there is zero social safety net, people will plumb the depths of deprivation, making murder a legitimate tool for survival. (Hat tip to Daniel Lakeland for raising this issue on Mastodon. Yes, over there, the discussion was quite civil.)

Still, human misery is a curiously relative beast. Even the poorest US citizens are, in material terms, among the richest humans ever to have lived. But that doesn't make their low status any less indignifying. Such is the nature of social life; we look to those around us to judge our wellbeing.

Toxic competition

Zooming out to the big picture, a good way to frame the relation between inequality and murder is as a break-down of human sociality. You see, from an evolutionary perspective, the key to sociality (human or otherwise) is that competition within groups gets *suppressed*. In other words, to be 'social' is to cooperate with others in your group.

Now, the problem with inequality is that it creates a reward structure that *stimulates* within-group competition. So if I have $10,000 in my pocket and you've got nothing, there's a significant payoff to killing me and taking my money. But the same is not true if we've both got $5,000.

Of course, I'm not suggesting that all murders are based on cash calculations . . . far from it. But the point is that the social environment imprints on the human psyche. In a dog-eat-dog world, everyone is your competitor. And so individuals are on hair-trigger alert. But in a pro-social world, your fellow humans are your comrades. So your whole social demeanor changes.

To wrap things up, think of income inequality as an index of *anti-sociality*. Ramp up this index, and you tear the fabric of social life, leading to many forms of toxic competition — murder being the most extreme. Reduce inequality, and human behavior becomes more prosocial. Competition gets suppressed, and cooperation becomes the norm. Murders become rare.

Of course, this thinking makes no sense if you believe economists' mantra that competition stimulates social welfare. Fortunately, the facts speak for themselves: inequality and murder go hand in hand.

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 E**T**D supporter

Sources and methods

Data for Figures 1 to 4

- Murder rates by country: data is from the World Population Review

- Top 1% income share: data is from the World Inequality Database, series sptincj992 (using the most recently available data in each country)

Download the data

Data for Figure 5

- Murder rates by country: data is from the World Bank, series VC.IHR.PSRC.P5, Intentional homicides (per 100,000 people)

- Top 1% income share: data is from the World Inequality Database, series sptincj992

Note: I exclude countries with a murder rate of zero. In my merged World-Bank-World-Inequality database, that amounts to removing seven country-year observations, five of which are in Iceland.

Why remove these zero-murder-rate observations? Well, when we work with the logarithm of the murder rate, we have no choice, since the log of 0 is undefined. Also, it's not clear whether a murder rate of zero indicates *no murders*, or if it is an artifact of rounding. At any rate, these zero murder-rate observations constitute 0.2% of the dataset.

Download the data

Data for Figure 6 and 7

- Murder rates by country: data is from the World Bank, series VC.IHR.PSRC.P5, Intentional homicides (per 100,000 people)

- Data for the Gini index of income inequality is from the World Bank, series SI.POV.GINI

As above, I remove observations where the murder rate is zero. There are four of them in a dataset of 1307 observations.

Download the data

Data for Figure 8 and 9

- Murder rates by country: data is from the World Bank, series VC.IHR.PSRC.P5, Intentional homicides (per 100,000 people)

- Bottom 10% income share: data is from the World Inequality Database, series sptincj992

As above, I remove observations where the murder rate is zero.

Download the data

Further reading

Wilkinson, R. G., & Pickett, K. (2009). *The spirit level: Why more equal societies almost always do better*. New York: Penguin Books.