

EconoFact Chats: New Insights on Inflation from a Billion Prices

Alberto Cavallo, Harvard Business School

Published on 25th February, 2024

Michael Klein

I'm Michael Klein, executive editor of EconoFact, a non-partisan, web-based publication of the Fletcher School at Tufts University. At EconoFact, we bring key facts and incisive analysis to the national debate on economic and social policies, publishing work from leading economists across the country. You can learn more about us and see our work at www.econofact.org.

Michael Klein

There is a saying variously attributed to Mark Twain or Will Rogers – ‘there are lies, damn lies, and statistics.’ One of the most closely watched statistics, especially in the past couple of years, is inflation. The problem isn't that the Bureau of Labor Statistics lies about inflation, but that the published inflation rate, like many other aggregate statistics, can mask very different price changes across categories of goods. What new information can we get by taking a more fine-grained look at price changes of more disaggregated categories of goods and services, or even particular products? An expert in this area is my guest today, Harvard Business School's professor Alberto Cavallo. Alberto pioneered the use of online data to measure inflation, co-founding the Billion Price Project in 2008, as well as PriceStats in 2011, the leading private source of inflation statistics in over 20 countries. Alberto, welcome back to EconoFact Chats.

Alberto Cavallo

Thanks Michael, it's great to be here again.

Michael Klein

Alberto, the Bureau of Labor Statistics releases new price and inflation data each month. The most closely watched of these is the headline inflation number, which is the change in prices of a representative basket of goods purchased by urban consumers. How many goods and services are used to construct this index?

Alberto Cavallo

About 80,000 products and services are sampled once a month. Traditionally, by the way, this is done with actually sending price collectors to physical stores, but gradually, in particular post-COVID, the BLS is actually checking some prices online as well. Then they use all these prices to build the price indices and measure headline and core inflation, as you mentioned.

Michael Klein

So 80,000 is a big number, but it can't be the case that even with this many items in the basket, the BLS is actually capturing everything that people buy.

Alberto Cavallo

No, that's true. So it's a large number, but absolutely we miss a few things. Sometimes important, for example, the indices, headline, and core that we use are, like you mentioned, are only for urban consumers. So there are some subpopulations we may not measure. It's also meant to be a

representative consumer. So we look at averages of spending patterns, and we may miss the differences between different groups. And it's also true that the sample size, once you focus on very narrow categories, we may be measuring an entire category like pet services, which includes lots of things, a really wide set of items and services, only with relatively few goods and a few sampled stores. So as you go into more disaggregated data, then sample sizes become smaller and smaller.

Michael Klein

Yeah, I heard that pet services, they just go to a few stores and look at a few things, and that covers parakeets, dogs, cats, goldfish, snakes, alligators, and whatever else people have as pets.

Alberto Cavallo

I mean, they do make a big effort to sample places where they are more representative, and that's a big effort of the BLS and other statistical institutes. But you can sometimes miss things that are important, particularly when you get big shocks and big crises.

Michael Klein

Alberto, you already alluded to one obvious problem with this, that a single basket of goods for a so-called typical consumer does not really reflect what many families and individuals purchase.

Alberto Cavallo

That's correct. And you know, it makes sense for policymakers often to think of a representative consumer and think of the average price level change over time, measure inflation that way, and make policy react to that. But often, we have questions that require us to look at a subset of the population. Perhaps if we want to understand, for example, how income inequality has progressed or changed during a crisis, you actually need more specific indicators. But very few statistical agencies around the world are able to have enough information and produce indicators for subsets of groups, or even in high frequency, they tend to do it only on a monthly basis, in the best of cases. So that's the limitation of the traditional statistics.

Michael Klein

Well, for sure, it's a daunting job. But just as a simple example, if you think about rural families as opposed to urban families, which the CPI is typically measuring, gasoline and energy prices are going to be a much bigger part of their budget than the budget of urban families. And so it's not capturing the real cost of living for those families by looking at the CPI for urban workers or urban people.

Alberto Cavallo

Yeah. And even within urban consumers, you have high-income households that they may spend a bigger share on travel, transportation, and lower share on food. And you have the opposite with low-income households. So I guess in normal times, when many of these different categories have similar rates of inflation, we wouldn't worry so much. But there are times when this actually creates big differences in the experienced inflation.

Michael Klein

Alberto, do you have any sense of how much this matters quantitatively? Qualitatively, it's easy to come up with examples of how baskets are different. What about quantitatively? Do you have any sense of that?

Alberto Cavallo

Yeah. So actually one of the first papers I wrote at the beginning of COVID tried to acknowledge this and give us a number, a quantitative number. So the point we made was that the COVID crisis was changing the spending patterns across these really large categories like food and fuel. And we did find that, you know, overall everyone was consuming more food, where we had rising prices, and we were all consuming less fuel because we couldn't go anywhere and prices were falling there. So across the board, everyone was experiencing higher inflation than what the official statistics could actually measure with these fixed weights that they tend to have. But these pre-existing differences between high and low-income households, and the fact that low-income households tend to spend so much more on food, meant that they were actually particularly affected by the rise of food inflation. How much was it? It was actually twice as high at the beginning of the pandemic. The levels weren't that different because we were talking about one or two percent inflation. So it doesn't sound like a big difference at that moment in time, but certainly, over time, that can accumulate and create some differences. And what's striking to me as I progressed with this analysis was that things got even worse if you managed to look within categories of goods. Those large spending patterns like food and fuel, they tended to dissipate after a few months. And we kept on working and looking at different narrow categories, like think of milk, for example. Even though it's a very narrow category, there are premium products that high-income families tend to buy, and there are very more basic, cheaper varieties of milk that low-income households can buy. And with the online data, we were able to actually identify the premium and cheap varieties, and follow their price changes over time. And we found that cheaper varieties during the two or three years of the pandemic, they actually experienced twice as high inflation again than more premium varieties. And in terms of accumulated inflation, this was quite significant. These were levels of 15, 20 percent inflation, compared to only 7 or 8 percent for premium products. So within categories, there's a lot of unmeasured, hidden inflation inequality that matters a lot, particularly at times like this.

Michael Klein

So this goes back to what I was saying in the introduction, that when you aggregate up, you miss things. And you're pointing out that you can miss some really important and big changes. So you're referring to your research, and that's the Billion Price Project. Alberto, can you describe this work, and do you, in fact, have a billion prices?

Alberto Cavallo

Yes, that's a good question. So this is based on an academic project, like you mentioned, starting in 2008 called the Billion Prices Project. In 2010, it had grown enough that we actually started a company that now collects the data, and produces daily inflation series in about 30 countries. That's PriceStats. And then we use all this data for research now, what we call the HBS Pricing Lab. Now, PriceStats collects daily prices for all the goods sold by about 1500 retailers, which on average sell about 10,000 products, each one of them. And that's roughly 15 million prices a day that get collected. So that gives you a billion prices every 66 days on average. So there is data

over time and has been accumulated. We've done this for over 12 years now. So there's a ton of information. But the key thing to remember here is that you don't need all that information or to use all that information. In fact, it would be a mistake to just put all those prices there into the indexes. But what big data in this sense does, it expands the options we have. So we can then go and identify the subsets of the data that really provide valuable information for us to be able to measure inflation in a certain way for different groups in high frequency. So that's, I think, the biggest advantage here. There's a revolution in data collection technology that enlarges the possibilities of the statisticians and economists trying to measure these things.

Michael Klein

So we already talked, Alberto, about the advantage of being able to track differences in inflation for different baskets of goods and services, and we discussed how typical purchases are different across households by income groups or regions. But in some other research, you also show that these disaggregated numbers help us understand when inflation turns either higher or lower. Can you describe what you found with that?

Alberto Cavallo

Absolutely. So that is both the disaggregated data and the ability to measure this on a daily basis that gives us an advantage. So for example, you might think if we measure inflation on a daily basis, we may get a lot of noise. But the truth is the large sample size actually controls some of that noise. And when you look at these daily indices, one thing that we have always found striking is that you can identify relatively stable trends over time. And by measuring it on a daily basis, you see exactly when those trends in the price level actually changed. So this gives you a very quick indicator of a change in the price level trend. Or if you think about it, it's essentially the inflation level. So in a paper with Gaston Garcia Zavaleta, my co-author, we used this feature in daily price data and daily indices to statistically identify when we can detect a structural break in the slope of the price index. And we do this not at the aggregate level, but rather with the most disaggregated indices we can construct, and get official weights for the CPI basket. And the advantage here is, imagine the Federal Reserve is raising rates and they do see that the aggregate index is slowing down, but they realize, well, maybe that's only energy, we don't know how widespread this is. Well, with these daily indicators and the structural breaks, we can actually build a measure that tells us when more than 50%, for example, of the CPI basket has actually experienced a turn in the slope of the price index, has experienced a slowdown in this case and allows the Federal Reserve to make a far more informed decision knowing that this is widespread, that this is affecting many subsectors. So that's the ability both of being able to see this on a daily basis, but also looking into these disaggregated sectors and building that up.

Michael Klein

Did you ever look at what would happen if you only took like every 30th daily observation, so you sort of mimicked what the Bureau of Labor Statistics does? How much information do you lose about, for example, the turning points that you're describing?

Alberto Cavallo

So that's a great question. If you have monthly data, you can sometimes detect the same break, but only after you have observed four or five months of data. So you can actually measure the new trend, which is not so obvious with monthly data. It takes about four or five months to

actually see it. Whereas if you're following this on a daily basis, after a few weeks, it can become clear and it becomes statistically detectable that you're in a new trajectory. So that's the advantage of going into high frequency for this particular question.

Michael Klein

I mean, this is really interesting. It's like another example in economics where people are using very, very big data sets to have new insights. And I really commend you for doing that. But you have data not just from the United States, but a range of countries. Do you find that inflation dynamics follow similar patterns across countries over the past few years, or are they really distinct across different countries?

Alberto Cavallo

Well, in the first two years of COVID, all these inflation dynamics were driven, essentially, by the same shocks. So they were very similar. Everyone experienced these spikes, different levels, of course, but the patterns over time look similar. Now, around the middle of 2022, you start seeing divergences, and more localized shocks and events seem to be driving many of those differences. So in that same paper I was referring to a minute ago that we take these trends, we can actually show that there's been a lot of heterogeneity in the moment in time when countries started to slow down. The first countries that I found with this data to slow down that slowed down were China, the US, Canada, they were relatively early, towards the end of 2022. And then the European countries took much longer. Latin American countries were somewhere in between. And there's a lot of research and understanding that has to go into why these differences happen. But you always have to start with being able to actually observe that difference. And certainly, something that stands out when you look at the data is how initially we were all driven by COVID factors, and today we again see a lot of dispersion in inflation.

Michael Klein

Alberto, as you well know, there's a lot of debate in this country about the sources of inflation, whether it's caused by the kind of supply shocks like COVID and energy prices after the invasion of Ukraine, or caused by very high demand given the generous government support during that period. Do your data say anything about the distinction between supply and demand shocks, especially, I guess if you're looking across countries?

Alberto Cavallo

Yeah, so that's a great point. We do find that in the US, some of the spike in the levels were higher than in some countries that did not do so much fiscal policy. But I guess the problem is usually, how do we measure the supply side? We don't have really good indicators of that. If you remember the discussion at the beginning of COVID, everybody knew that there was a supply shock, but the evidence mostly relied on ships at the ports and some backlogs there. We actually wrote a paper trying to leverage the fact that we could see how many goods were actually out of stock. When we go to these stores, we see their prices, but we saw how many were either out of stock in the store or they had completely disappeared from the website. We used that to build a measure of stockouts or, if you will, supply disruptions. Now, just to be clear, that can come because there is excess demand and limited supply. Both factors matter, but it gave us a measure to how much discrepancy there is, how much shortage, or excess demand, or insufficient supply, however, you want to think about it. We did track it over time, and what we found is that,

certainly there was a spike in stockouts at the beginning of the pandemic, and then they started to improve, but then they got worse in 2022. There was a sense in which the supply-side problems continued. They persisted for a long time. Interestingly, towards the middle of 2022, they started to improve quite dramatically. We sort of went back to normal in terms of stockouts by early [2023]. That actually gave me information in terms of this discussion. My sense at the time was that people were very concerned that the policies of the Fed were going to create a recession and that's the way to bring prices down. But I think, certainly on the supply side, we were seeing improvements that allowed the economy to achieve disinflation without actually experiencing a recession. Moving forward, though, I guess the same data suggests that that buffer on the supply side is no longer there. If more disinflation were to be required, then that could lead to more recession. But I don't think, to be honest, that is going to happen because we have seen in our data a lot of slowdown and going back to normal across most of the categories we follow. There's been this good disinflation period that has lasted quite a long time, and I think we're getting back to pre-pandemic levels very soon.

Michael Klein

The stockout question is really interesting. Did you see differences across somewhat broader categories of goods, what was most likely to be stocked out, what wouldn't appear on the shelves, and what was more likely to actually continue to be available?

Alberto Cavallo

Yeah, so we actually saw some categories recover relatively quickly, like electronics, household goods were also recovering relatively quickly and most of the persistent stockouts remained in food until the middle of 2022. Not surprisingly, if you look at food inflation, it was very high, it got worse with the Ukraine war, with the stockouts also rising around that time. So food was certainly one of those sectors, but nearly across all goods, we have seen them go back to normal already. So I'm not surprised that we have seen disinflation, both driven by the supply side but also by the higher rates that central banks have pushed.

Michael Klein

Alberto, we're recording this a couple days after what was predicted to be a big snowstorm, and in fact, there wasn't one at all. So that kind of forecasting makes economists look good, but without holding you to it too much, given that you have daily data, what's your sense of what's happening with inflation most recently?

Alberto Cavallo

Well, with our daily data, we tend to see mostly goods, and we've seen this disinflation. What has happened in the U.S., in particular recently, is that service inflation, in particular shelter inflation, seems to be more persistent, and that's pushing up the annual inflation rate numbers that we see published in the media all the time. But there are some indications in the service sectors, and from other people who also have high-frequency data from the shelter component, there's a sense in which some of the same dynamics we have seen in goods are also happening in shelter, and the CPI or official statistics we tend to have a lag given the way the data is collected and constructed will naturally tend to reflect, possibly in the next few months, a slow down. But as you said, forecast [inaudible] we are bad, economists in general, are bad at forecasting, we may get shocks that we don't know about, it's always hard to do. I do always argue that the best I can

do is get better data to know what's going on exactly today, and I can tell you from what we see in the goods side, there's actually significant and very broad-based disinflation in the U.S. so it appears to be good news.

Michael Klein

And coming full circle to how we started out, the fact that shelter prices are falling more slowly has differential impacts on people. If you're a homeowner and you aren't really facing that, especially if your mortgage is paid off, that's one thing. If you're a young person who's renting, that's another. So there's again a lot of value in being able to look at this very disaggregated price series that you've put together both for understanding how inflation affects different people differently, but also in your very interesting research on the dynamics of inflation. So Alberto, thank you very much for joining me today. I'm always fascinated by the research that you do, and I appreciate you coming on to this podcast.

Alberto Cavallo

Thank you Michael, It's always a pleasure to be here.

Michael Klein

This has been EconoFact Chats. To learn more about EconoFact and to see the work on our site, you can log into www.econofact.org. EconoFact is a publication of the Fletcher School at Tufts University. Thanks for listening.