

THE NUMBERS GUY DECEMBER 2, 2009

When Combined Data Reveal the Flaw of Averages

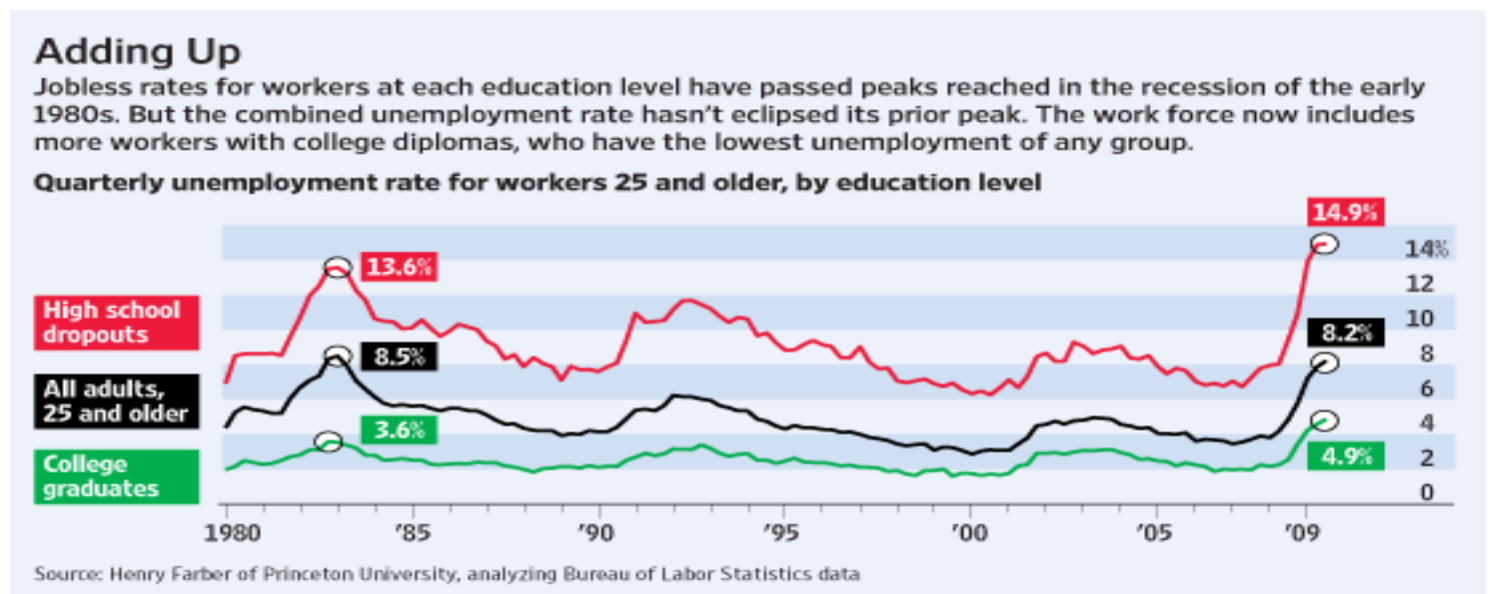
In a Statistical Anomaly Dubbed Simpson's Paradox, Aggregated Numbers Obscure Trends in Job Market, Medicine and Baseball

By CARI TUNA

Is the current economic slump worse than the recession of the early 1980s?

Measured by unemployment, the answer appears to be no, or at least not yet. The jobless rate was 10.2% in October, compared with a peak of 10.8% in November and December of 1982.

But viewed another way, the current recession looks worse, not better. The unemployment rate among college graduates is higher than during the 1980s recession. Ditto for workers with some college, high-school graduates and high-school dropouts.



U.S. Unemployment: A Historical View

Track the national unemployment rate since 1948.

More interactive graphics and photos So how can the overall unemployment rate be lower today but higher among each group? The anomaly is an example of Simpson's Paradox -- a common but misleading statistical phenomenon rooted in the differing sizes of subgroups. Put simply, Simpson's Paradox reveals that aggregated data can appear to reverse important trends in the numbers being combined.

The jobless rates for each educational subgroup are higher today, but the overall rate is lower because workers are more educated. There are more college graduates, who have the lowest

When Combined Data Reveal The Flaw 2

unemployment rate. And there are fewer high-school dropouts, who have the highest unemployment rate.

"It's the magic of weighted averages," says Princeton University economics professor Henry Farber. "We have more skilled workers than we had before, and more-skilled workers are less susceptible to unemployment." Still, he adds, compared with a similarly educated worker in 1983, "the worker today has higher unemployment at every education level."

Simpson's Paradox has fooled many. In the fall of 1973, for instance, the University of California, Berkeley's graduate division admitted about 44% of male applicants and 35% of female applicants. That raised eyebrows among school officials, who feared bias and asked Peter Bickel, now a professor emeritus of statistics at Berkeley, to analyze the data.

"The associate dean of the graduate school thought that the university might be sued," Mr. Bickel says.

When Mr. Bickel and his colleagues scrutinized the data, they found little evidence of gender bias. Instead, they discovered that more women had applied to departments that admitted a small percentage of applicants, like English, than to departments that admitted a large percentage of applicants, like mechanical engineering.

Simpson's Paradox is "responsible for a vast quantity of misinformation," says Xiao-Li Meng, chairman of Harvard University's statistics department. "You can easily be fooled."

Janet Francis, 36, has been unemployed for two years. She is working toward a college degree in social work.

Mr. Meng says medical studies are replete with examples of Simpson's Paradox, often identified because such research attracts great interest. He cites a study of kidney-stone treatments published in 1986 in the *British Medical Journal*.

In the summary, the authors note that traditional surgery successfully removed or substantially shrank the stone in 78% of cases. A less-invasive form of surgery succeeded in 83% of cases, suggesting it is more effective.

"But if you read the paper carefully, you see it's completely the reverse," Mr. Meng says.

That is because the treatments were applied to two different

When Combined Data Reveal The Flaw 3

groups of kidney stones. The traditional treatment was more successful on both small stones and large stones. The paradox rears its head because the less-invasive method was used much more often on small stones, and treatments of small stones are generally more successful than treatments of large stones, pulling the aggregate success rate in favor of the less-invasive method.

Less-invasive surgery appears more successful, but traditional surgery is always preferential, Mr. Meng explains. "The aggregate number really is meaningless."

Mr. Meng says he thinks many people who wield similarly misleading data do so unintentionally. "When you find data that go with your theory, then you don't dig deeper."

Other examples of Simpson's Paradox have surfaced in fields as varied as baseball, education and air travel. Ken Ross, a professor emeritus at the University of Oregon and baseball enthusiast, notes in a 2004 book on baseball statistics that in both 1995 and 1996, Derek Jeter of the New York Yankees had a lower batting average for each season than David Justice, then of the Atlanta Braves.

Combining the two years, however, Mr. Jeter had a better average. The paradox resulted from the fact that in 1995 Mr. Jeter had only 48 at-bats with a .250 average while Mr. Justice had more at-bats (411) with a .253 average. The following year, Mr. Jeter had 582 at-bats with a .314 average while Mr. Justice had only 140 at-bats with a higher average of .321, pushing the two-year average in Mr. Jeter's favor.

Jeff Witmer, a statistician at Oberlin College, uncovered another example of the paradox by analyzing flight-delay data from the Bureau of Transportation Statistics. In January 2009, he says, UAL Corp.'s United Airlines had a higher percentage of flight delays than Continental Airlines at each of four major metropolitan airports. Combining data for all four cities, Continental appeared to have a worse record because it flies many more planes from one city where weather delays are more frequent.

The overall delay data could mislead travelers. "You can't just look at the overall rate," Mr. Witmer says. "It matters where you are flying to and from."

United spokesman Michael Trevino says "this is four of some 200"

airports to which United flies, adding that United's November on-time rate ranks the company first among major airlines.

The Simpson's Paradox in unemployment rates by education level is but the latest example. At a glance, the unemployment rate suggests that U.S. workers are faring better in this recession than during the recession of the early 1980s. But workers at each education level are worse off, says Mr. Farber, the Princeton professor.

His analysis of Labor Department data found unemployment rates for workers 25 and older at each of four education levels were higher in the third quarter of 2009 than at any point since data series began, in 1976.

Among workers with college diplomas, unemployment was 4.9%, compared with 3.6% in fourth quarter of 1982, the prior peak. Likewise, unemployment among high-school dropouts surged to 14.9%, from 13.6% in the first quarter of 1983.

But the 8.2% overall unemployment rate for workers 25 and older in the third quarter of 2009 trailed the 8.5% rate in the first quarter of 1983. (The unemployment rate for workers of all ages was 10.2% in October 2009. Continuous historical education-level data are available only for workers 25 and older on a quarterly basis.)

The reason the current overall rate looks better: College graduates, who have the lowest unemployment rate, are now more than a third of the work force, compared with roughly 25% in 1983, says the Labor Department. Meanwhile, the share of high-school dropouts has shrunk to roughly 10% of the work force, from nearly 20% in 1983.

That means the paradox will persist until the total current unemployment rate surpasses the high watermark of the early 1980s. Economists don't expect the November unemployment rate -- due out Friday -- to reach those heights.

—Carl Bialik is on vacation.

Write to Cari Tuna at cari.tuna@wsj.com

Printed in The Wall Street Journal, page A21

Copyright 2009 Dow Jones & Company, Inc. All Rights Reserved