

Are Recession-Proof jobs real?

An attempt to discern what industries recovered from the *last* recession the quickest, and to analyze the number of jobs and median wages from each industry, with the goal of figuring out just what jobs are "recession proof".

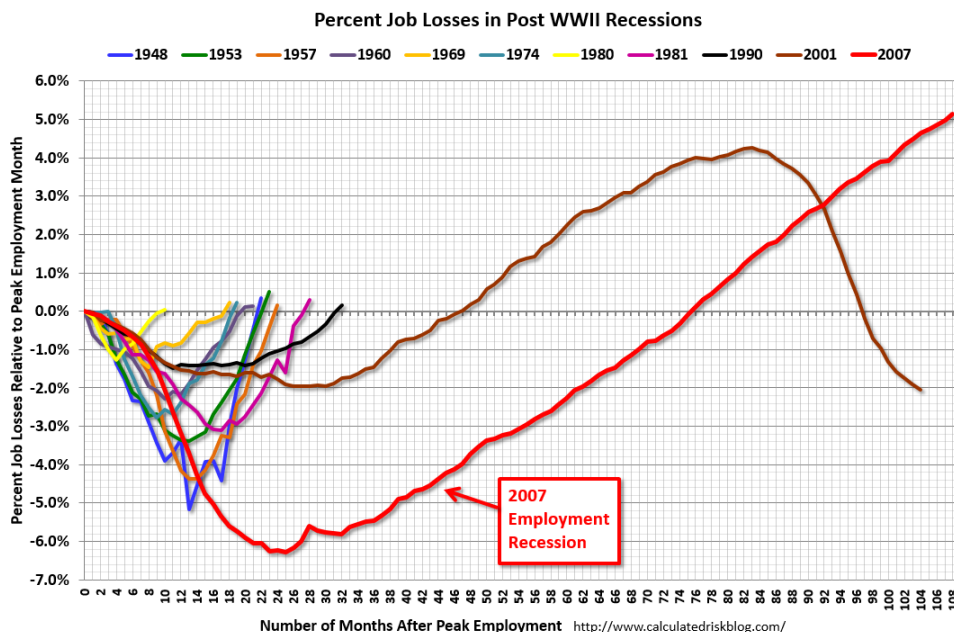
Overview

We are currently in one of the worst recessions in living memory. The COVID-19 pandemic has hobbled the American economy, with layoffs putting millions of Americans in precarious financial positions. Firms have collapsed or gone dark, education is being (temporarily?) reworked to be online-only, jobless Americans, myself included, are currently using their idle time to gain new marketable skills so they can reenter an extremely competitive workforce.

Job training is the solution to skills mismatch in the workforce most favored by American policymakers, and there are government-funded programs to help workers learn new skills. The decision of *what* new skills to learn, however, is left to individual workers.

This project is an attempt to describe the employment numbers from the last recession. I intend to compare the employment numbers across industries, to see which industries are the "safest" bets for workers to invest their resources into joining.

Dataset



I will be using data from the U.S. Bureau of Labor Statistics' Quarterly Census of Employment and Wages. Data exists as far back as 1990 (for the current questionnaire), however I will only concern myself with the months between February 2007 (the "official" start of the recession) to May 2014 (The month in which the American economy achieved the same overall job numbers as when the recession began (see image)).

Given this is a government dataset, it is exceedingly clean. There are a few columns that seem to be of the incorrect type (some floats where they should be integers) but cleaning the data is not going to be much of a challenge.

What **will** be a challenge is working with a dataset this large. Plotting on just a sample of 30k records or so takes a while, and there's no way it can run off my own machine. Overall I'm looking at over 14 gigs of data. (I would like to discuss the limits of spark/AWS and what can be reasonably done with those resources.)

Trimming the dataset will involve several steps. While the initial analysis are by month/quarter, it can easily be converted to yearly data. There's quite a few columns that are not useful for my initial questions below, and they can be dropped. I'm not familiar enough with the processing power, and whether dropping columns vs. rows will make the job either. See data/csew_schema for a breakdown of columns (green I want, yellow might be useful for stretch goals)

Assumptions

1. The BLS data in question is accurate (duh).
2. Inflation will be ignored (for now).
3. I will concern myself only with job counts, not population counts or proportions of jobs to working-age/capable Americans.
4. Average wage in each industry (the only available data) is generalizable to the working population.

Questions (MVP)

1. What were the most "resilient" industries? Which industries had the shortest period between job losses and recovery to their pre-recession numbers?
2. Were there any industries that remained constant in their job numbers throughout the recession?
3. Which industries had the most dramatic shifts in wages across the recession? Which were devalued (lower average wage) and which increased in value?
4. How did the job numbers change for each industry as a percentage of total American jobs?

Stretch Goals (MVP+)

1. How did wages fare when adjusted for inflation?
2. Which region(s) were the highest recovery/growth areas for the most resilient industries? (I know you want maps, Land)
3. Are there any familiar trends in the *current* (Q1-Q2 2020) recession?