Lukas Püttmann   About   Research   Blog
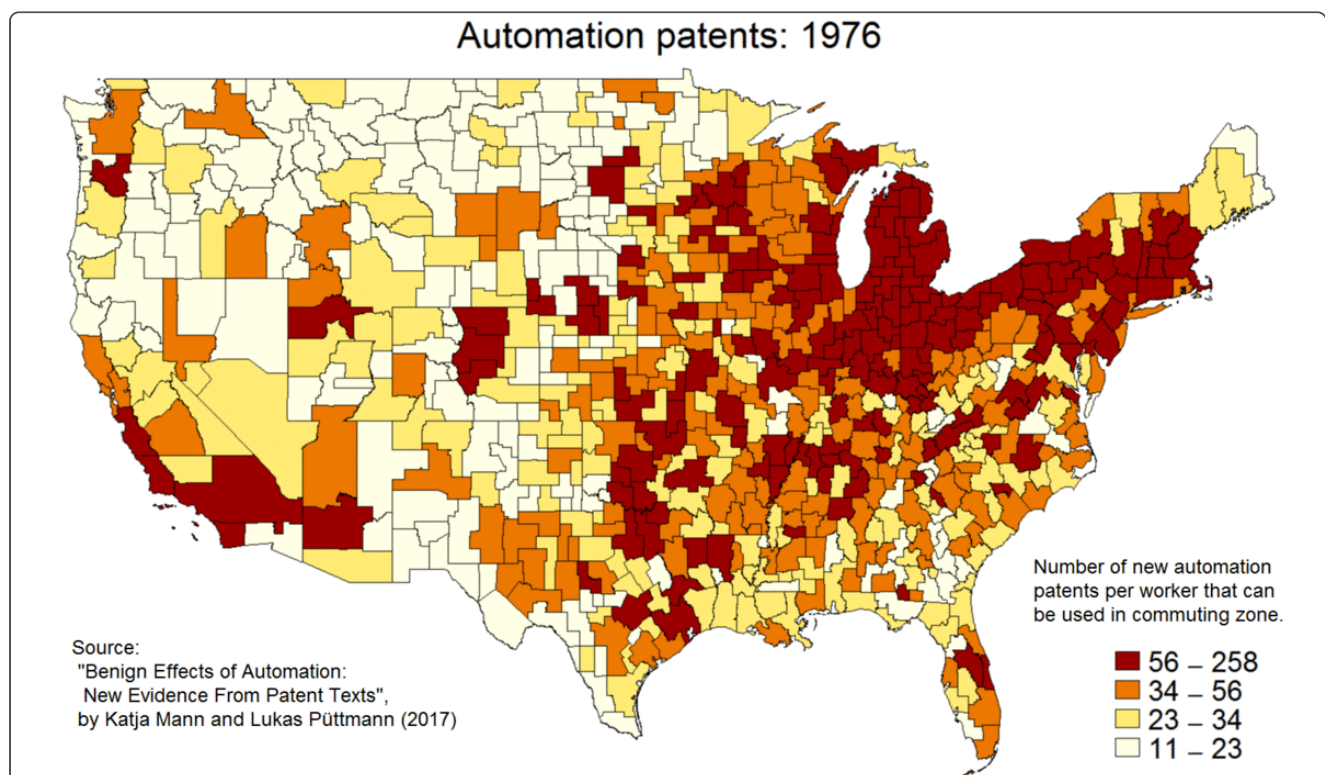
Our paper: "Benign Effects of Automation: New Evidence from Patent Texts"

22 Sep 2017

We've updated our paper on automation patents which you can find here and here (and see here to play around with the data). In short, we identify automation patents based on their patent texts and create this new dataset:

Background

Reading Erik Brynjolfsson and Andrew McAfee's "Second Machine Age" three years ago got us interested in the topic. We were looking for data on how automation had advanced over time and across industries, but none of

the existing proxies quite satisfied us. The idea to use patents came after reading Acemoglu et al. (2014) for a class which made us aware of patents as a data source.

It was striking that (almost) no researchers so far had made use of the actual patent *texts*. Instead, people use patent metadata such as citations (see e.g. here and here).

We were lucky that Google provides a bulk download page for patents (also see these codes). One of the tasks that took us the longest time was to write a parser to extract and clean the text sections for all patents from the titles, abstracts and text bodies of the patents. That was tricky as those are 336 GB of text which makes storing and retrieving the documents an issue. The parsing step took a week on a server where eight cores ran in parallel.

We then trained a naive Bayes algorithm on a sample of patents that we classified ourselves (using these guidelines).

Here's an example of such an automation patent:


So the patent with the name "Automatic Taco Machine" was invented by one Barry Brummet from California and is assigned to (owned by) Taco Bell.

He applied for the patent in 1994 was granted it in 1996. It cites other patents such as this one. Using this data, we can also check who cited the "Automatic Taco Machine". We find that up until 2010 it was cited a total of 11 times (for example by this, this and this patent).

Even in the title of the "Automatic Taco Machine" you have the word "automatic", so this is an easy patent to classify. Other words in the patent text that were important in classifying it as automation were are "removable", "storage", "acceptable", "support arm", "assist", "communicate", "measures", "processor" and 179 others.

Next, we checked where every patent is likely to be *used* (not invented). For this, we used Brian Silverman's concordance tables. For our example, we find:
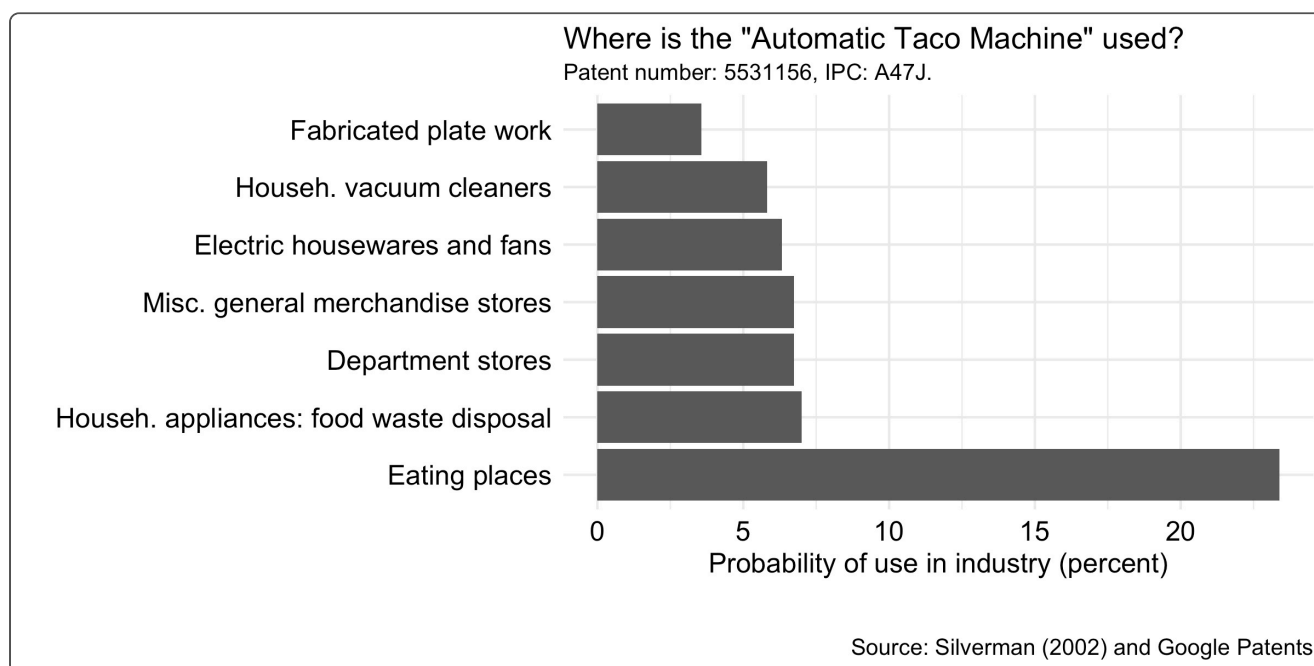
So with 23% probability it's used in "Eating places", "7% in "Household appliances …"" and 6% in "Department stores". Looks plausible.

We then matched our industries to US commuting zones (with the CBP) and get the picture at the top of this blog post. It shows the number of automation patents that can be used by a single worker in each of these commuting zones. In the earlier years, the rust belt saw a lot of automation patents, but this has become much more spread out and diffuse over the years.

Our empirical results are the following:

- Between 1976 and 2014, about 2 million automation patents (out of 5 million patents in total) were granted.

- The share of innovation concerned with automation rose from 25% in 1976 to 67% in 2014.

- There was more investment in robots and computers in industries with more automation patents. Those were also industries where in 1960 more people worked in routine tasks.

- Local labor markets (commuting zones) in the US where more new automation patents could be used experienced *increases* in employment.

- Automation led to a loss in manufacturing employment, but this was more than compensated by a rise in service sector employment.

If you've become curious about the paper, you can find it here.

References

Acemoglu, D., U. Akcigit and M. A. Celik (2015). "Young, Restless and Creative: Openness to Disruption and Creative Innovations". NBER Working Paper No. 19894.

Autor, D., D. Dorn, G. H. Hanson, G. Pisano and P. Shu (2016). "Foreign Competition and Domestic Innovation: Evidence from U.S. Patents". NBER Working Paper No. 22879.

Bell, A., R. Chetty, X. Jaravel, N. Petkova and J. V. Reenen (2017). "Who Becomes an Inventor in America? The Importance of Exposure to Innovation". NBER Working Paper No. 24062.

Bessen, J. and R. Hunt (2007). "An Empirical Look at Software Patents". *Journal of Economics and Management Strategy*, 16(1): 157–189.

Brynjolfsson, E. and A. McAfee (2014). *"Second Machine Age"*. Norton & Company.

Katja Mann and Lukas Püttmann (2023). "Benign Effects of Automation: New Evidence From Patent Texts". *The Review of Economics and Statistics*

Silverman, B. S. (2002). *Technological Resources and the Logic of Corporate Diversification*. Routledge.

Further reading:

- Our VoxEU column
- Noah Smith's take on Bloomberg View
- Play around with the data from the paper.