**Part 1: Assembling a textual corpus**

I will give you three files. These correspond to the compilations of sections affected by documents published in the Federal Register from 1964-1972, 1973-1985, and 1986-2000. In particular, I have already focused us on Title 12, and I want you to further focus on Chapter 2. Chapter 2 is the regulations of the Federal Reserve. You will see these documents appears to be a very boring table or index, but it is a remarkably powerful document once we organize it correctly. This is basically telling you what is happening where in regulatory space.

For the first part of our analysis, we want to digitize these in an excel format. So you can do this in two ways. For the earliest document, OCR looks bad because it doesn’t appear to recognize that the document has a two column structure, so it thinks that sentence continues on the right column before going down. This will eventually lead to junk. So I would suggest screenshot the relevant section and have ChatGPT OCR from scratch. I believe it is the case that the bigger the screenshot the better the OCR, so zoom in. In other cases the OCR doesn’t look so bad and you can copy paste the content you need without OCR. Save the text file with the year that these changes come from, that will be important for staying organized.

A close-up of a document

Description automatically generated

Once you have assembled the corpus, proceed to step 1(b). It’s basically the same, but I don’t want to confuse you about the overall plan. You can proceed to Step 2 if you just want to understand the whole task.

**Step 2**: digitization

Using chatgpt, turn these raw text tables into csv files. The batch API is appropriate for this. Some prompt engineering may be desirable, but I found good success with the following:

Format this for me as a 3 column. The first number refers to the part of the code fo federal regulations. Column 2 is what was changed. Column 3 is the Page of the Federal Register where it happened.

**Step 3:** cleaning tables. Make sure the output looks reasonable. Combine the CSVs you got by adding in the proper year and Federal Register Volume number into one master CSV for years 1964-2000.

**Future Steps:** Where is this headed? Well, you will notice that there are parts like 204.12 or 203 etc. If there is no dot, that means a whole regulation of the federal reserve was affected. Differences at the part level, between 204 and 205, correspond to very different regulatory tasks, that’s why they are in different parts. If there is something after the dot, it means some changes where made within the regulation.

What I am hoping to show is that there has been an increasing number of tasks assigned to the Federal Reserve over time. We will see what these tasks are, but I am expecting it will be consumer protection stuff and also the system soundness stuff that has been added. We will see that Congress gives new authorities with some regularity, but seldom walks these authorities back, and within the domain in which they give authority there is tremendous volume of activity. I hope to produce a figure like the following

X-axis is time

Y-axis is a bunch of rows corresponding to task

Big circle when the task is assigned turning to a right arrow until the task stopped.

Dots, squiggles, or some other marker of policy activity within each task

Somehow I will layer on top of that activity by Congress touching upon the same policy matters. It will show typically much less activity.

This is a lot harder than what has been done before to illustrate increasing delegation to the bureaucracy, but it is quite a bit better for reasons I could get into.

**Step 1(b)** – Since 2000, these things are published as separate PDFs easily downloaded.  
  
<https://www.govinfo.gov/app/collection/lsa/LSAMONTHLY/2000/12>

Typically, there is an annual addition for Title 12, seems to be usually published in December. The PDFs are pretty clean and easy to read. Getting us to 2023 will give us nearly 60 years of data. Going back earlier in time is possible, but I think a bridge to cross later if we really need it.