Policy Brief Notes

Johnny Dedyo

November 15, 2025

1 Dataset merge

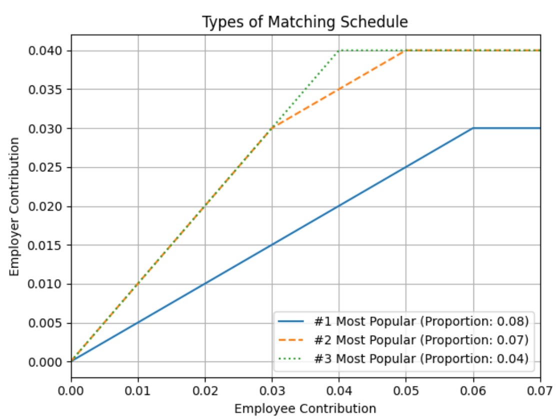
1.1 Questions

(a) Which dataset should I merge into?

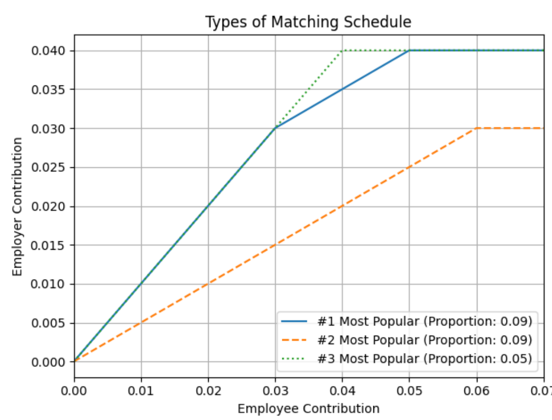
(b) Must `opr_pn` be numeric? (Getting some that have a value of, e.g., $2E2$.)

1.2 Figure 1

- The most common plan is actually no matching at all. So I filter those out when I make this figure.



(a) Old data



(b) New data

1.3 Table 6

- There is no filtering of the dataset for this table. Presume a match of 0 for plans that have no mention of any matching contributions.
- The fraction offering matching is calculated as the fraction of all plans that offer a nonzero match.
  - In practice, I calculate the company match if an employee fully exploits the match. If this is nonzero, the plan is marked as offering a nonzero match.
  - *More complicated* plans—those with special eligibility requirements, mid-year changes, or other complications—are also marked as offering a nonzero match since special rules for matching imply the existence of matching. These plans are *excluded* from all subsequent calculations for Table 6 since we cannot "code" them.
- The fraction of plans with each number of tiers is calculated as the fraction

$$\frac{\text{\# of plans with } n \text{ tiers}}{\text{\# of plans we coded (excluding \textit{More complicated})}}$$

- Right now the "Of which we can code" line is calculated as the fraction in of plans offering a nonzero match that are not *More complicated*. That is, the denominator is plans offering matching, not the whole universe.
  - The **Among Those Coded** line indicates that the denominator is now the number of plans we coded, i.e., the denominator of the fraction in the bullet above.

Table 6: Share of Plans Offering Matching, by Tier Structure (Old vs. New Data)

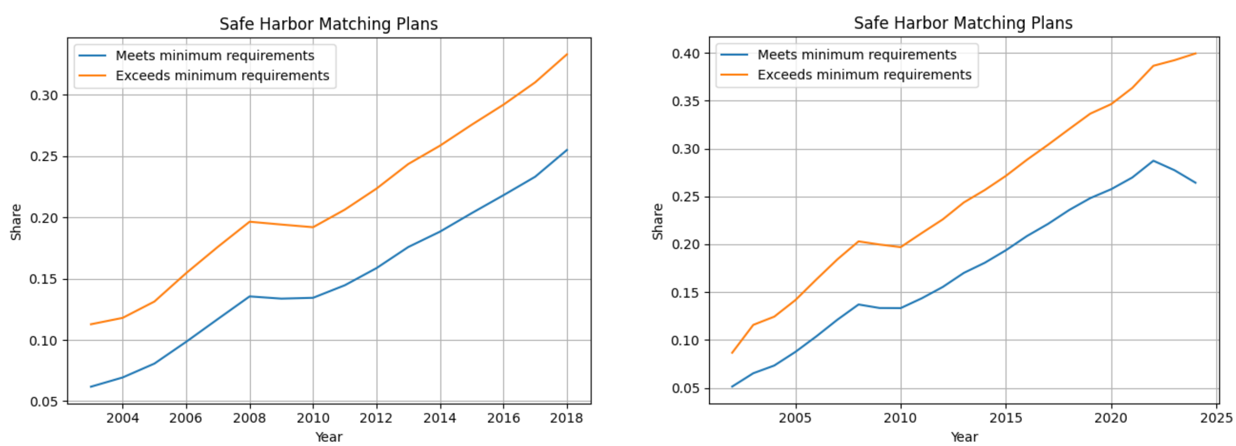
	Fraction of Plans (old data)	**Fraction of Plans (new data)**
Offering matching	0.761	0.804
Of which we can code	0.661	0.652
Among Those Coded:		
Single-tier	0.806	0.767
Two-tier	0.193	0.232
Three-tier	0.001	0.001

1.4 Figure 2

- Filter out all *More complicated* plans—again, these are those with special eligibility requirements, mid-year changes, or other complications that make it impossible for us to encode them accurately in simple tabular form.
  - Question: does it make sense to filter out the *More complicated* plans, or should I leave them in the denominator?
- Presume a match of 0 for plans that have no mention of any matching contributions.
- Define basic safe harbor as a 100% match on the first 3% of contributions and 50% on the next 2%.
- Define enhanced safe harbor as 100% on the first 4%.
- Define QACA safe harbor as 100% on the first 1% and 50% on the next 5% OR 100% on the first 3.5%.
- Define "Exceeds minimum requirements" according to this specification:

  > Plan sponsors can make additional matching contributions on top of these guidelines and maintain their safe harbor status as long as the additional match satisfies:
  >
  > (a) contributions above 6% of salary are not matched, and
  >
  > (b) the additional matching contribution on top of the safe harbor match do not exceed 4% of compensation.

- The blue line plots the proportion of plans that perfectly fit one of the three safe harbor specifications above (i.e., those that are more generous are not counted).
- The orange line counts plans that are more generous than the corresponding safe harbor plan, in addition to those that meet the criteria exactly.

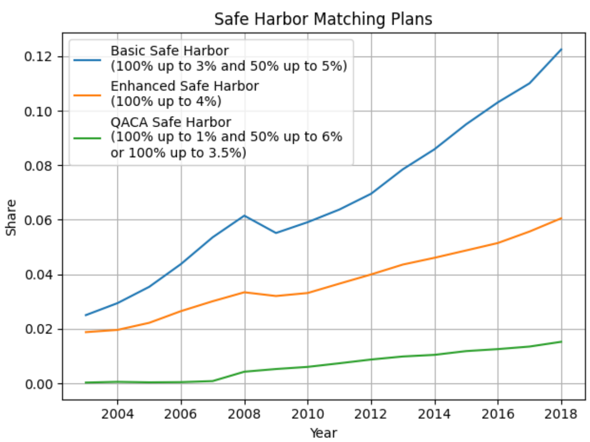


(a) Old data

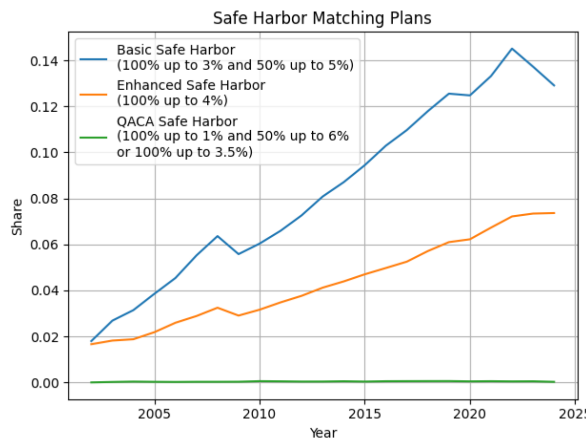
(b) New data

1.5 Figure 3

- There is no filtering of the dataset for this figure. The denominator is our whole universe of plans.
- Presume a match of 0 for plans that have no mention of any matching contributions.
- Safe harbor plans are defined as for the previous figure.
- Each line is the proportion of one of the three flavors of safe harbor plan in the full sample of plans, by year.
  - It does *not* count plans that exceed the safe harbor requirements as the corresponding safe harbor plan. These lines plot the proportion of plans that meet the exact safe harbor specifications.



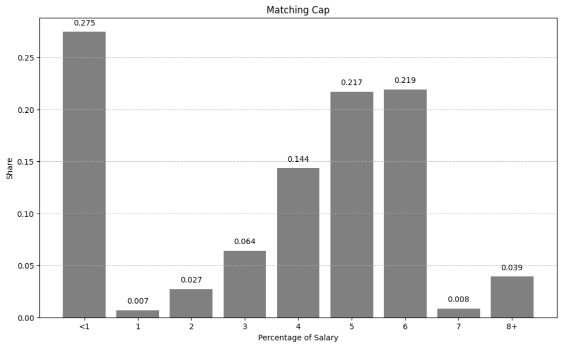
(a) Old data



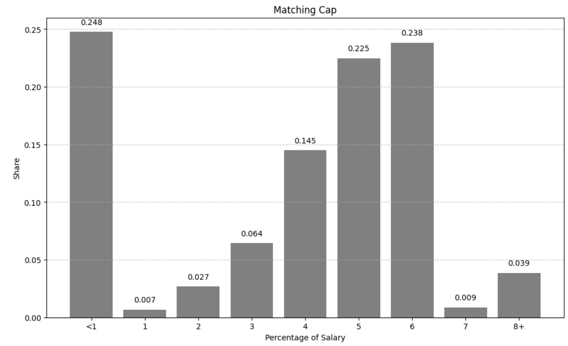
(b) New data

1.6 Figure 4

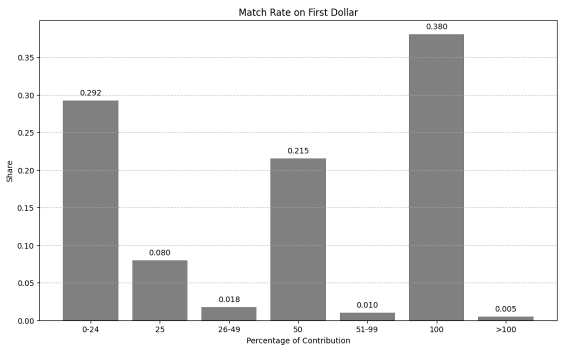
- For Figures 4a–c, select plans from the year 2017 that are not *More complicated*. So the denominator here excludes plans with special eligibility requirements, mid-year changes, or other complications.
- Presume a match of 0 for plans that have no mention of any matching contributions.
- Each of these barplots tallies the proportion of plans in each bin, so the bar heights in each subfigure add up to 1.
- For Figure 4a, put a plan in bin i if $i - 0.5 \leq \texttt{match\_cap} < i + 0.5$. For the < 1 bin, the lower limit is $-\infty$; for the $8+$ bin, the upper limit is ∞.
  - `match_cap` is defined as the maximum employee contribution that will receive some (nonzero) level of employer matching.
- For Figure 4b, put a plan in the bin (lb, ub) if $lb - 0.5 \leq \texttt{match\_rate\_1} < ub + 0.5$. (for example, the 0-24 bin would be $-0.5 \leq \texttt{match\_rate\_1} < 24.5$.
  - `match_rate_1` is defined as the percentage employer match on the first dollar of employee contributions.
- For Figure 4c, calculate the employer matching contribution assuming that the employee fully exploits his match. Again, put a plan in bin i if $i - 0.5 \leq \texttt{max\_match} < i + 0.5$. For the 0% bin, the lower limit is $-\infty$; for the $> 8\%$ bin, the upper limit is ∞.



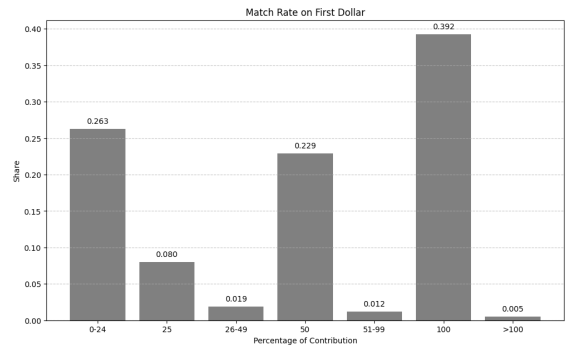
(a) Matching cap (old)



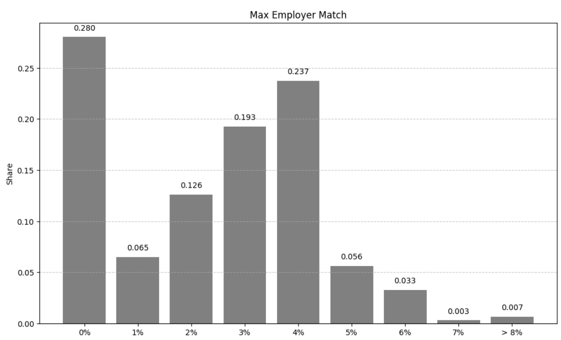
(b) Matching cap (new)



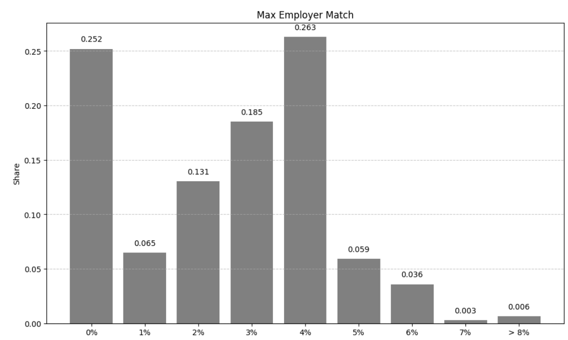
(c) Match rate on first dollar (old)



(d) Match rate on first dollar (new)



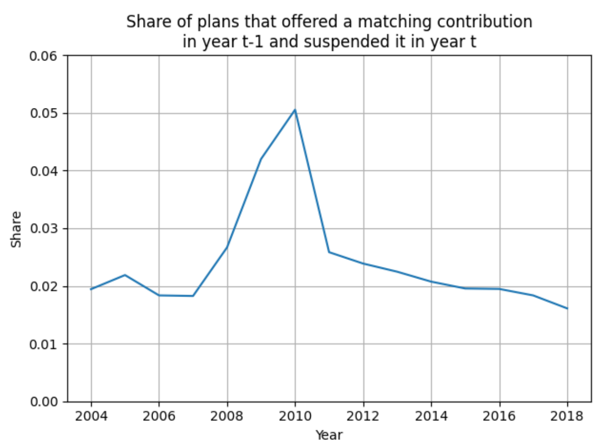
(e) Max employer match(old)



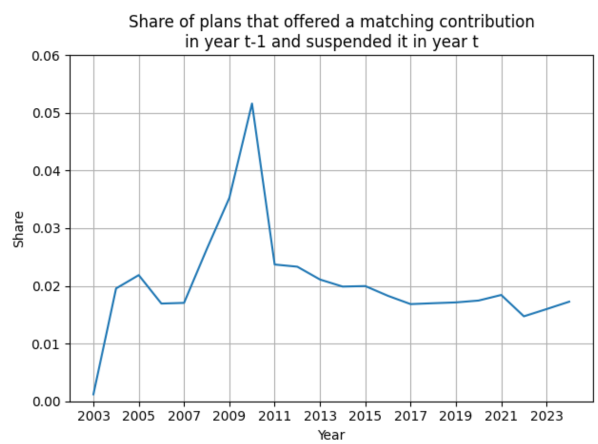
(f) Max employer match (new)

1.7 Figure 5

- There is no filtering of the dataset. The denominator for these figures is the whole universe of plans.
- Presume a match of 0 for plans that have no mention of any matching contributions.
- Define a "plan" as a pair (`opr_ein`, `opr_pn`). Index the dataset on this pair.
- For each index—(`opr_ein`, `opr_pn`)—and year, flag whether there is a match offered (i.e., the first match rate is positive OR the plan is *More complicated*).
  - *More complicated* plans are presumed to have a nonzero match since a plan must have special features to be classified as *More complicated*, but a unilateral 0 match could not have features like special eligibility requirements, a mid-year change, etc.
- For each year, t, plot the proportion of plans for which this flag was `True` in year $t-1$ AND is `False` in year t



(a) Figure counting *More complicated* as a matching contribution (old)



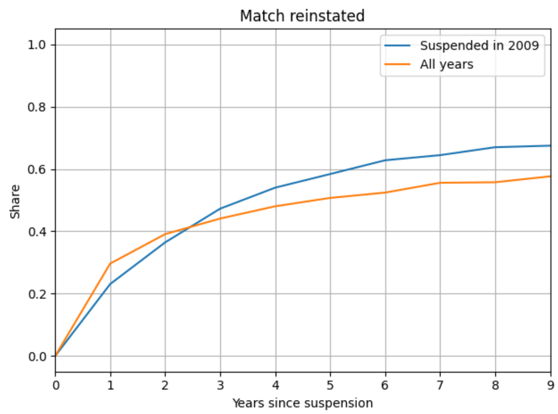
(b) Figure counting *More complicated* as a matching contribution (new)

1.8 Figure 6

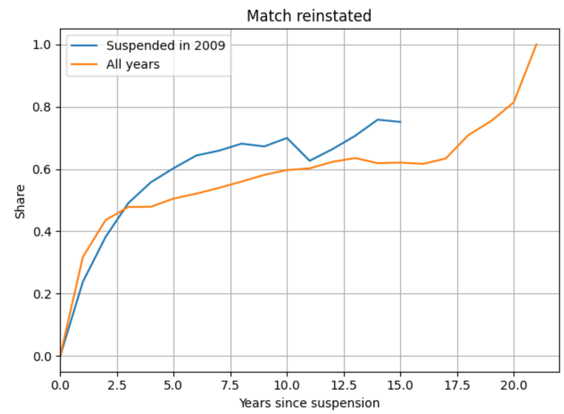
- No filtering on the data. The denominator is our whole universe of plans.
- Presume a match of 0 for plans that have no mention of any matching contributions.
- Define a "plan" as a pair $(\texttt{opr\_ein}, \texttt{opr\_pn})$. Index the dataset on this pair.
- Define a plan as offering a match if `match_rate_1` > 0 (i.e., there is a nonzero match on the first dollar of employee contributions).
  - Note that this defines a *More complicated* plan as NOT offering a match. (This choice was made so that the max employer contribution is well-defined, since *More complicated* means that we can't easily describe the plan in a table).
- Define a plan as suspending a match in year t if the plan offered a match in year $t - 1$ but does not offer one in year t. If it suspended in multiple years, take the earliest one.
- For Figure 6a, calculate the maximum employer matching contribution for each plan (and year) assuming the employee fully exploits the match. Then plot the average as a function of the number of years since the match was suspended.
- For Figure 6b, define a match as being reinstated in year t if the plan did not offer a match in year $t - 1$ but does offer one in year t. Plot the proportion of plans that have reinstated as a function of the number of years since the suspension.
- Important question: are we happy with how I handled the *More complicated* plans here? I am unsure whether I should filter them out altogether, treat them as a match (but then how do I calculate the max employer contribution?), or something else.

(a) Max employer contribution (old)

(b) Max employer contribution (new)



(c) Match reinstated (old)



(d) Match reinstated (new)

1.9 Figure 7

- Filter out *More complicated* plans.
- Presume a match of 0 for plans that have no mention of any matching contributions.
- Define a "plan" as a pair (`opr_ein`, `opr_pn`). Index the dataset on this pair.
- For Figure 7a, the "Mean" line plots the maximum employee contribution that will receive employer matching. It is a simple average across all observations for the year.
  - The "Within firm" line plots the results of a fixed effects regression. Specifically, it plots the year t coefficient, β_t in the regression specification

    $$\text{AvgMatchCap}_t = \sum_i \alpha_i \text{Firm}_i + \sum_j \beta_j \text{Year}_j + \epsilon_t$$

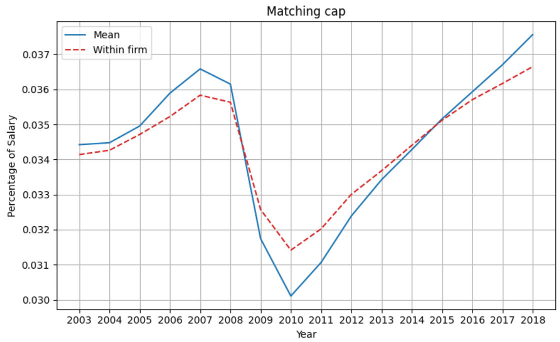
    where Firm_i is a dummy variable for the presence of the i^{th} (`opr_ein`, `opr_pn`) pair in year t, α_i is the fixed effect of Firm_i, and Year_j is a dummy for year j (so for year t, only Year_t is nonzero).
- For Figure 7b, the "Mean" line plots the match rate on the first dollar of participant contributions. It is a simple average across all observations for the year.
  - The "Within firm" line plots the results of a fixed effects regression. Again, it plots the year t coefficient, β_t in the regression specification

    $$\text{AvgMatchRate}_t = \sum_i \alpha_i \text{Firm}_i + \sum_j \beta_j \text{Year}_j + \epsilon_t$$

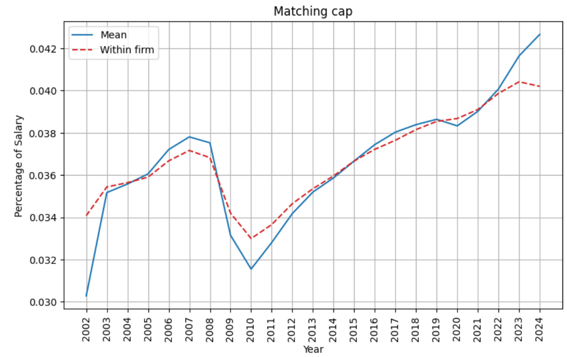
    where the variables are defined as above.
- For Figure 7c, the "Mean" line plots the maximum employer contribution, assuming that the participant fully exploits his match. It is a simple average across all observations for the year.
  - The "Within firm" line plots the results of a fixed effects regression. Again, it plots the year t coefficient, β_t in the regression specification

    $$\text{AvgMaxMatch}_t = \sum_i \alpha_i \text{Firm}_i + \sum_j \beta_j \text{Year}_j + \epsilon_t$$

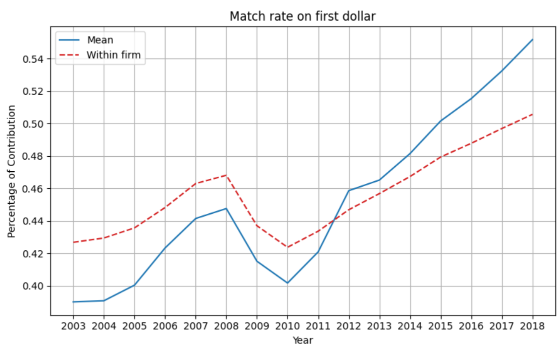
    where the variables are defined as above.



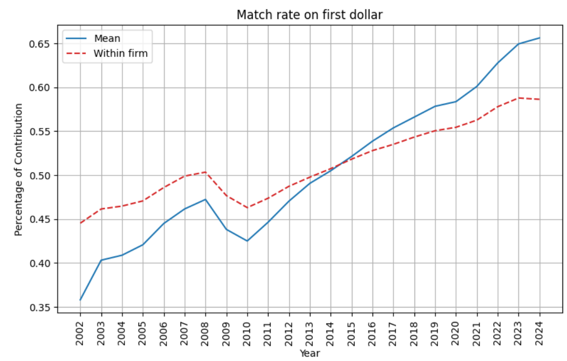
(a) Matching cap (old)



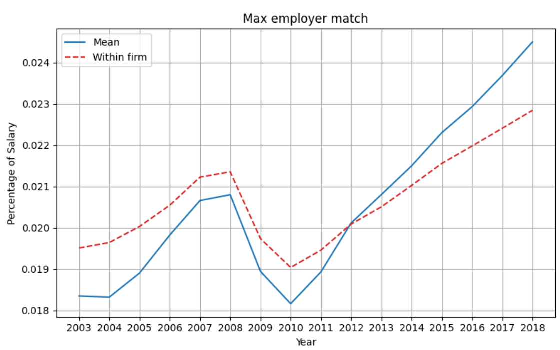
(b) Matching cap (new)



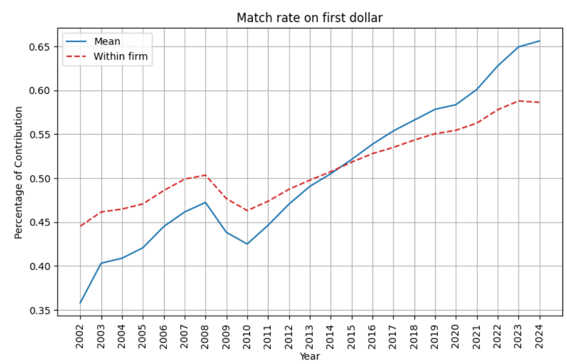
(c) Match rate on first dollar (old)



(d) Match rate on first dollar (new)



(e) Max employer match (old)



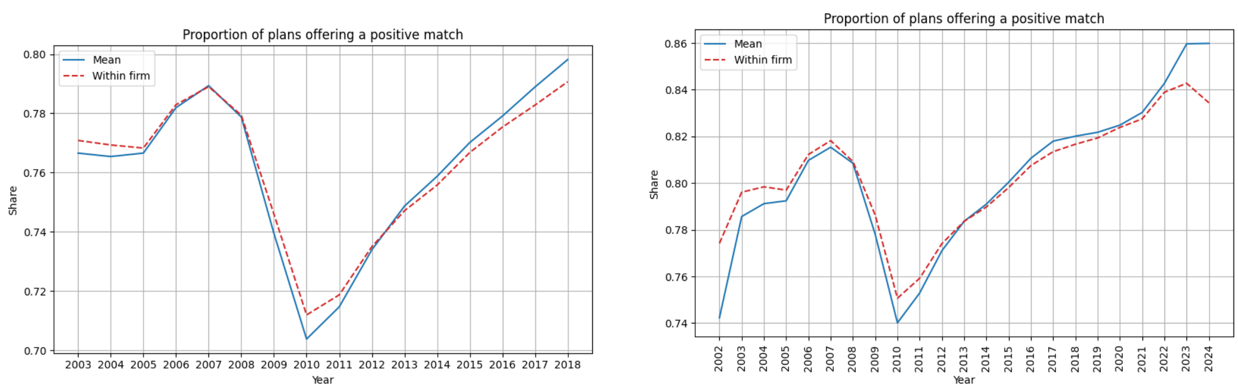
(f) Max employer match (new)

1.10 Figure 8

- Define a "plan" as a pair $(\texttt{opr\_ein}, \texttt{opr\_pn})$. Index the dataset on this pair.
- Define a *More complicated* plan as having a nonzero match (since if the plan had a match of 0, it would not be classified as *More complicated*, as argued under Figure 5).
- Presume a match of 0 for plans that have no mention of any matching contributions.
- For all other plans, calculate the maximum employer match if the employee fully exploits his plan. If this is nonzero, flag the plan as having a nonzero match.
- The "Mean" line plots the proportion of plans that offer a nonzero match as defined above. It is a simple average across all observations for the year.
  - The "Within firm" line plots the results of a fixed effects regression. It plots the year t coefficient, β_t in the regression specification

$$\text{PropPositiveMatch}_t = \sum_i \alpha_i \text{Firm}_i + \sum_j \beta_j \text{Year}_j + \epsilon_t$$

where Firm_i is a dummy variable for the presence of the i^{th} $(\texttt{opr\_ein}, \texttt{opr\_pn})$ pair in year t, α_i is the fixed effect of Firm_i, and Year_j is a dummy for year j (so for year t, only Year_t is nonzero).



(a) Old data

(b) New data

1.11 Figure 9

- Define a "plan" as a pair $(\texttt{opr\_ein}, \texttt{opr\_pn})$. Index the dataset on this pair.
- Filter out *More complicated* plans (since we cannot easily define their match rates and caps).
- Presume that plans with no mention of matching contributions do not match at all.
- Filter out all plans with a non-positive match, since we want statistics conditional on a positive match being offered.
- For Figure 9a, the "Mean" line plots the maximum employee contribution that will receive employer matching, conditional on the plan offering a nonzero match. It is a simple average across all observations for the year. (As in Figure 7a, but conditional on a positive match being offered).
  - The "Within firm" line plots the results of a fixed effects regression. Specifically, it plots the year t coefficient, β_t in the regression specification

    $$\text{AvgMatchCap}_t = \sum_i \alpha_i \text{Firm}_i + \sum_j \beta_j \text{Year}_j + \epsilon_t$$

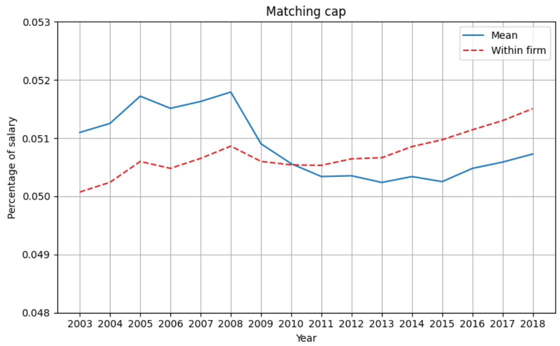
    where Firm_i is a dummy variable for the presence of the i^{th} $(\texttt{opr\_ein}, \texttt{opr\_pn})$ pair in year t, α_i is the fixed effect of Firm_i, and Year_j is a dummy for year j (so for year t, only Year_t is nonzero).
- For Figure 9b, the "Mean" line plots the match rate on the first dollar of participant contributions, conditional on the plan offering a nonzero match. It is a simple average across all observations for the year. (Similar to Figure 7b, but conditional on a positive match being offered).
  - The "Within firm" line plots the results of a fixed effects regression. Again, it plots the year t coefficient, β_t in the regression specification

    $$\text{AvgMatchRate}_t = \sum_i \alpha_i \text{Firm}_i + \sum_j \beta_j \text{Year}_j + \epsilon_t$$

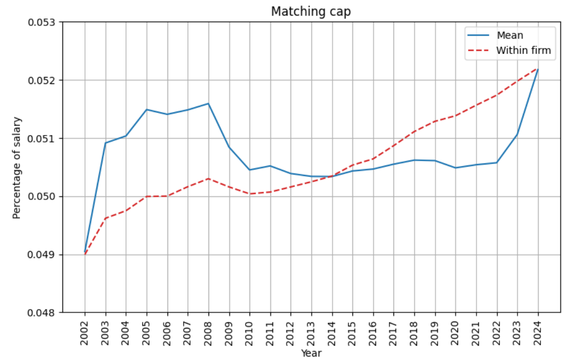
    where the variables are defined as above.
- For Figure 9c, the "Mean" line plots the maximum employer contribution, assuming that the participant fully exploits his match and conditional on the plan offering a nonzero match. It is a simple average across all observations for the year. (Similar to Figure 7c, but conditional on a positive match being offered).
  - The "Within firm" line plots the results of a fixed effects regression. Again, it plots the year t coefficient, β_t in the regression specification

    $$\text{AvgMaxMatch}_t = \sum_i \alpha_i \text{Firm}_i + \sum_j \beta_j \text{Year}_j + \epsilon_t$$

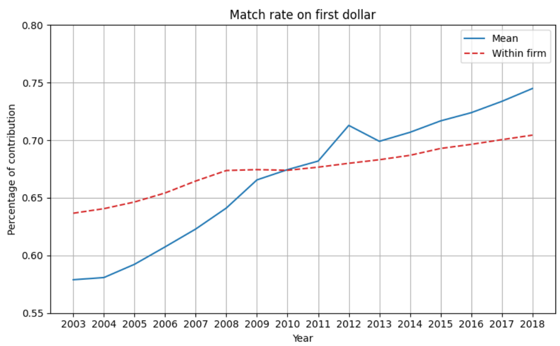
    where the variables are defined as above.



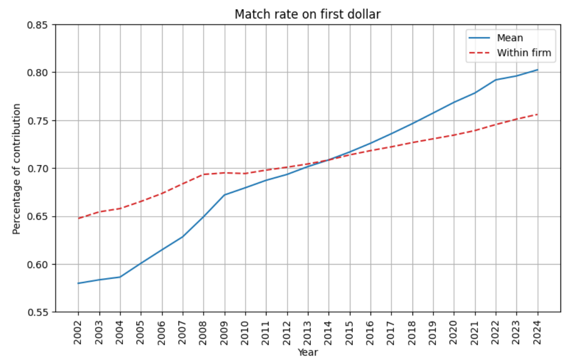
(a) Matching cap, conditional on nonzero match (old)



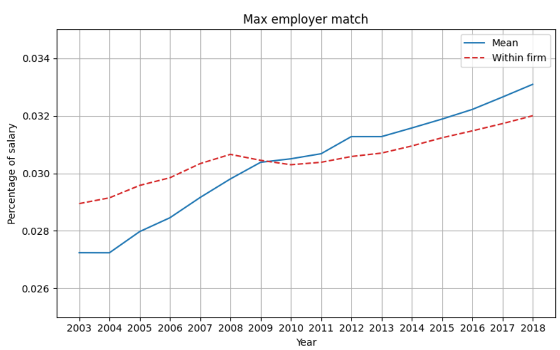
(b) Matching cap, conditional on nonzero match (new)



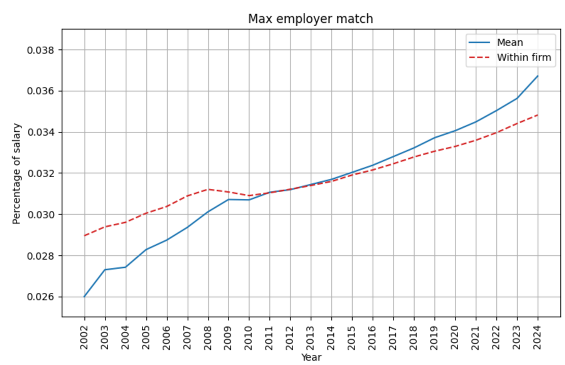
(c) Match rate on first dollar, conditional on nonzero match (old)



(d) Match rate on first dollar, conditional on nonzero match (new)



(e) Maximum employer matching contribution, conditional on nonzero match (old)



(f) Maximum employer matching contribution, conditional on nonzero match (new)

1.12 Updates

(a) Figure 4: split 0 from 1-24

(b) Figure 5 (and others): Recreate with LLM training data.

(c) Proportion of plans with matching, average match_rate, match_cap, etc. over time (t on x-axis, avg on y-axis).

(d) Add fixed effects for Figure 6 and cut after 15 years

(e) Add standard errors to graphs (see if sample is smaller for 2024 where stuff gets weird).

(f) Add boring tables on sample sizes, etc, to get a sense of "what's in" the dataset.

(g) What are the differences in average employer contribution and avg participant contribution by plans we can/cannot code.