A Long-Run Anatomy of Task Exposures to Technology

(Discussion)

Advanced analytics:
new methods and applications for macroeconomic policy

Johannes Zahner

Philipps-Universität Marburg

July 2022

Research question

Research question(s):

(i) the development of a novel NLP-based methodology for assessing the relevance of patents to tasks over time, and (ii) the dynamic effects of technology on labor productivity, working hours, and labor share.

Research Design

General idea: Occupation → tasks → task exposure to technology

- Occupation → task (topics): Topic modelling on task description of > 1000 occupation
- Task → task exposure: (i) similarity between patent titles and task topics relying on word embeddings and (ii) PCA to extract three factors: (a) replacing workers, (b) augmenting workers, (c) idiosyncratic innovations.

Results

1. Majority of the variance in patent relevance over time is explained by factor 1 (manual-biased technologies).
2. Factor 2 (cognitive-biased technology) becomes only relevant around 1990.
3. Shock to manual-biased technologies $\rightarrow$ replace workers (employment $\downarrow$, labor share $\downarrow$)
4. Shock to cognitive-biased technologies $\rightarrow$ augment workers (employment $\uparrow$, labor productivity $\uparrow$)

Results

1. Majority of the variance in patent relevance over time is explained by factor 1 (manual-biased technologies).
2. Factor 2 (cognitive-biased technology) becomes only relevant around 1990.
3. Shock to manual-biased technologies $\rightarrow$ replace workers (employment $\downarrow$, labor share $\downarrow$)
4. Shock to cognitive-biased technologies $\rightarrow$ augment workers (employment $\uparrow$, labor productivity $\uparrow$)

Results

1. Majority of the variance in patent relevance over time is explained by factor 1 (manual-biased technologies).
2. Factor 2 (cognitive-biased technology) becomes only relevant around 1990.
3. Shock to manual-biased technologies $\rightarrow$ replace workers (employment $\downarrow$, labor share $\downarrow$)
4. Shock to cognitive-biased technologies $\rightarrow$ augment workers (employment $\uparrow$, labor productivity $\uparrow$)

Results

1. Majority of the variance in patent relevance over time is explained by factor 1 (manual-biased technologies).
2. Factor 2 (cognitive-biased technology) becomes only relevant around 1990.
3. Shock to manual-biased technologies $\rightarrow$ replace workers (employment $\downarrow$, labor share $\downarrow$)
4. Shock to cognitive-biased technologies $\rightarrow$ augment workers (employment $\uparrow$, labor productivity $\uparrow$)

Questions and comments I

Tasks → task exposure to technology

Linguistic difficulties in identification of task exposure:

- Difficulty to measure complementary tasks, say public speaking and hotel service being replaced by online meetings: Correlation between linguistic similarity in task and technology and the likelihood to be replaced by technology.
- Vague language: Explicit use of vague language (e.g. patent trolls).
- Are patent titles proper representation of research efforts? Specifically:
  - There could be incentives to make the titles sound different to avoid/gain the attention of unions or the press?
  - The argument in favour of context-specific embeddings (i.e. BERT) does not seem very relevant when the text is so short.
  - Recent research (Baumgärtner, Zahner, 2021) suggests that in areas where language is highly technical, embeddings trained on small but representative samples of text are superior to large-scale, state-of-the-art general language models from Google, etc.
  - Are titles informative in the first place (see next slide)?.

Questions and comments I

Tasks → task exposure to technology

Linguistic difficulties in identification of task exposure:

- Difficulty to measure complementary tasks, say public speaking and hotel service being replaced by online meetings: Correlation between linguistic similarity in task and technology and the likelihood to be replaced by technology.
- Vague language: Explicit use of vague language (e.g. patent trolls).
- Are patent titles proper representation of research efforts? Specifically:
  - There could be incentives to make the titles sound different to avoid/gain the attention of unions or the press?
  - The argument in favour of context-specific embeddings (i.e. BERT) does not seem very relevant when the text is so short.
  - Recent research (Baumgärtner, Zahner, 2021) suggests that in areas where language is highly technical, embeddings trained on small but representative samples of text are superior to large-scale, state-of-the-art general language models from Google, etc.
  - Are titles informative in the first place (see next slide)?.

Questions and comments I

Tasks → task exposure to technology

Linguistic difficulties in identification of task exposure:

- Difficulty to measure complementary tasks, say public speaking and hotel service being replaced by online meetings: Correlation between linguistic similarity in task and technology and the likelihood to be replaced by technology.
- Vague language: Explicit use of vague language (e.g. patent trolls).
- Are patent titles proper representation of research efforts? Specifically:
  - There could be incentives to make the titles sound different to avoid/gain the attention of unions or the press?
  - The argument in favour of context-specific embeddings (i.e. BERT) does not seem very relevant when the text is so short.
  - Recent research (Baumgärtner, Zahner, 2021) suggests that in areas where language is highly technical, embeddings trained on small but representative samples of text are superior to large-scale, state-of-the-art general language models from Google, etc.
  - Are titles informative in the first place (see next slide)?.

Questions and comments I

Tasks → task exposure to technology

Linguistic difficulties in identification of task exposure:

- Difficulty to measure complementary tasks, say public speaking and hotel service being replaced by online meetings: Correlation between linguistic similarity in task and technology and the likelihood to be replaced by technology.
- Vague language: Explicit use of vague language (e.g. patent trolls).
- Are patent titles proper representation of research efforts? Specifically:
  - There could be incentives to make the titles sound different to avoid/gain the attention of unions or the press?
  - The argument in favour of context-specific embeddings (i.e. BERT) does not seem very relevant when the text is so short.
  - Recent research (Baumgärtner, Zahner, 2021) suggests that in areas where language is highly technical, embeddings trained on small but representative samples of text are superior to large-scale, state-of-the-art general language models from Google, etc.
  - Are titles informative in the first place (see next slide)?.

Questions and comments I

Tasks → task exposure to technology

Linguistic difficulties in identification of task exposure:

- Difficulty to measure complementary tasks, say public speaking and hotel service being replaced by online meetings: Correlation between linguistic similarity in task and technology and the likelihood to be replaced by technology.
- Vague language: Explicit use of vague language (e.g. patent trolls).
- Are patent titles proper representation of research efforts? Specifically:
  - There could be incentives to make the titles sound different to avoid/gain the attention of unions or the press?
  - The argument in favour of context-specific embeddings (i.e. BERT) does not seem very relevant when the text is so short.
  - Recent research (Baumgärtner, Zahner, 2021) suggests that in areas where language is highly technical, embeddings trained on small but representative samples of text are superior to large-scale, state-of-the-art general language models from Google, etc.
  - Are titles informative in the first place (see next slide)?.

Questions and comments I

Tasks → task exposure to technology

Linguistic difficulties in identification of task exposure:

- Difficulty to measure complementary tasks, say public speaking and hotel service being replaced by online meetings: Correlation between linguistic similarity in task and technology and the likelihood to be replaced by technology.
- Vague language: Explicit use of vague language (e.g. patent trolls).
- Are patent titles proper representation of research efforts? Specifically:
  - There could be incentives to make the titles sound different to avoid/gain the attention of unions or the press?
  - The argument in favour of context-specific embeddings (i.e. BERT) does not seem very relevant when the text is so short.
  - Recent research (Baumgärtner, Zahner, 2021) suggests that in areas where language is highly technical, embeddings trained on small but representative samples of text are superior to large-scale, state-of-the-art general language models from Google, etc.
  - Are titles informative in the first place (see next slide)?.

Questions and comments I

Tasks → task exposure to technology

Linguistic difficulties in identification of task exposure:

- Difficulty to measure complementary tasks, say public speaking and hotel service being replaced by online meetings: Correlation between linguistic similarity in task and technology and the likelihood to be replaced by technology.
- Vague language: Explicit use of vague language (e.g. patent trolls).
- Are patent titles proper representation of research efforts? Specifically:
  - There could be incentives to make the titles sound different to avoid/gain the attention of unions or the press?
  - The argument in favour of context-specific embeddings (i.e. BERT) does not seem very relevant when the text is so short.
  - Recent research (Baumgärtner, Zahner, 2021) suggests that in areas where language is highly technical, embeddings trained on small but representative samples of text are superior to large-scale, state-of-the-art general language models from Google, etc.
  - Are titles informative in the first place (see next slide)?.

Patent Titles

Patent 42(0 000)

Title: Gear train

Abstract: *A gear train enables rotational drive in either sense of a driven member le preventing transmission of rotation in either sense from the driven to the driving member. Each member of the train comprises a pair of discs attached on their faces with each disc having asymmetric teeth on its periphery and each disc of the pair being a reflection about a diametral line of the other. The asymmetrics in each member of the train are different.*

Claims: *What we claim is: 1. A gear train in which rotation [..] 2. A gear train as [...] 5. A gear train as [...]*

Patent Titles

Patent 42(0 000)

Title: Gear train

Abstract: *A gear train enables rotational drive in either sense of a driven member le preventing transmission of rotation in either sense from the driven to the driving member. Each member of the train comprises a pair of discs attached on their faces with each disc having asymmetric teeth on its periphery and each disc of the pair being a reflection about a diametral line of the other. The asymmetrics in each member of the train are different.*

Claims: *What we claim is: 1. A gear train in which rotation [..] 2. A gear train as [...] 5. A gear train as [...]*

Patent Titles

Patent 42(0 000)

Title: Gear train

Abstract: *A gear train enables rotational drive in either sense of a driven member le preventing transmission of rotation in either sense from the driven to the driving member. Each member of the train comprises a pair of discs attached on their faces with each disc having asymmetric teeth on its periphery and each disc of the pair being a reflection about a diametral line of the other. The asymmetrics in each member of the train are different.*

Claims: *What we claim is: 1. A gear train in which rotation [..] 2. A gear train as [...] 5. A gear train as [...]*

Questions and comments II

Tasks → task exposure to technology

- Average patent embedding per year: (i) Should all patents be weighted equally and (ii) are the patents independent of each other?
  - Breakthrough patents (which are very relevant) could crowd out other smaller patents, which would reduce the research effort for your measurement.
  - Averaging their embeddings could bias your index towards larger industries that naturally file more patents and towards tasks with many patent trolls.
    → A possible alternative would be to measure similarity on a patent basis and use the threshold from the de-trending exercise or focus on outliers from the beginning.
- What about long-term effects/lags in implementation, i.e. innovation in trains replacing horse-related tasks?

De-trending exercise

What about terms that do not exist at all in your reference corpus?

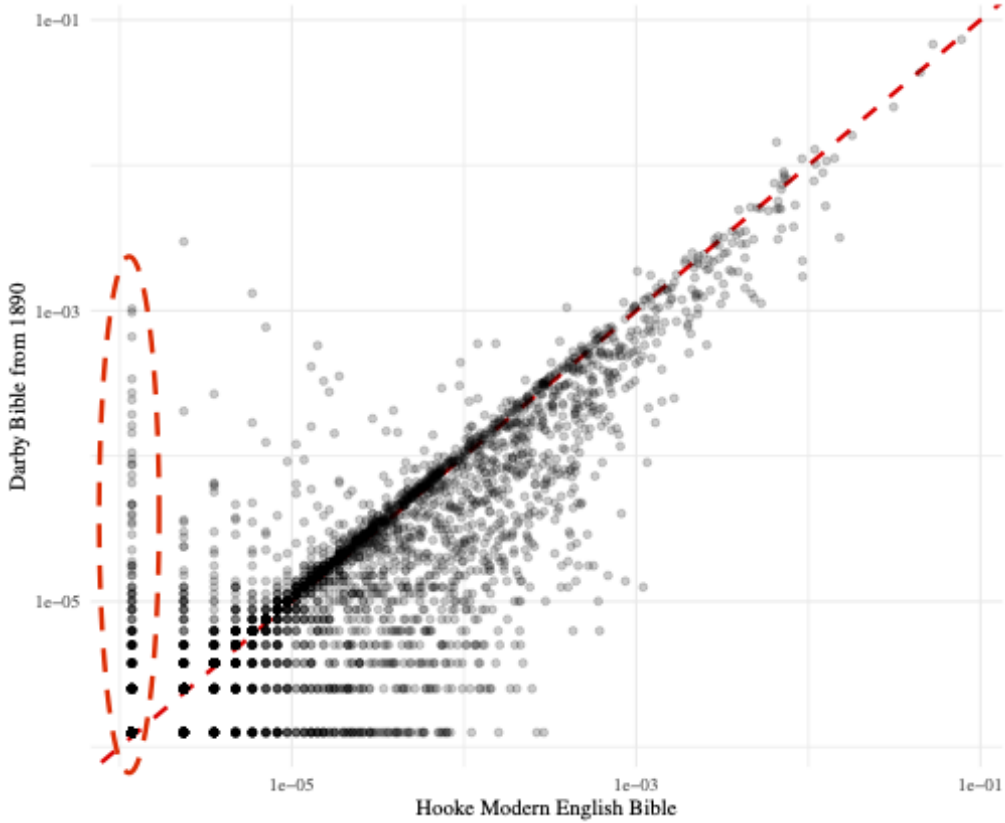

Figure: Relative frequency of terms in two Bible editions (1871 vs. 1965)

Further Comments

- How many dimensions in BERT?
- AR(1) instead of de-trend?
- Structural breaks in patent relevance?
- manual-biased = low-skilled work and cognitive-biased = high-skilled?
- How many task topics? 70 or 45?
- How can we interpret your results in term of magnitude?