Technological Change and Occupations over the Long Run

L. Kogan¹  D. Papanikolaou²  L. Schmidt³  B. Seegmiller³

¹MIT Sloan and NBER
²Kellogg School of Management and NBER
³MIT Sloan
Motivation

- A large set of questions require us to measure innovation outcomes.
  - Decline in measured productivity: Is innovative output low? Or is the relation with productivity weakened?
  - What is the relation between innovation and worker outcomes?

- To answer these questions we need:
  - a measure of innovation that is comparable across time and space
  - a way to identify exposure to technical change at the level of an individual worker

- This paper:
  - We create time-series indicators of technological change at the level of worker occupations
What we do

• We build on Kelly et al. (2018) to identify important patents by parsing the text of all 9 million patents issued by the USPTO since 1836.
  ▶ We posit that important patents are those that are distinct from prior patents but are closely related to future patents.
  ▶ Our technology indicators correlate with measured productivity at the aggregate and sectoral level

• For each patent, we identify a group of occupations that are likely to be significantly affected by the underlying invention.
  ▶ Occupations exposed to technological change experienced declines in employment and wages.

• Implementation requires us to measure distance between patent documents and occupation task descriptions.
  ▶ We do so using advances in text analysis
Measuring technological innovation

• Innovation is hard to measure directly.
  ▶ How do you measure ideas? R&D spending measures inputs not outputs.

• Our starting goal is patents. Why?
  ▶ By definition, patents relate to new inventions
    (though not all valuable inventions are patentable)
  ▶ They measure output not inputs
    (important if you think research productivity is slowing down)

• However, not all patents are equally valuable inventions.
  ▶ pro-patent shift in US policy (Hall and Zeidonis 2001)

• To create meaningful indices of innovation, we need to weigh important patents differently from ones that are trivial.
History of Biotech: How the "First" Biotech Patent Generated Millions

MARIE GODAR — 03/12/2015 - 4 MINS - MEDICAL

The Cohen-Boyer patents were now issued 35 years ago at Stanford University... so what were they and how did they shape the Modern Biotechnology Field?

Recombinant DNA (rDNA) products provided a new technology platform for a range of industries, resulting in over US$35 billion in sales for an estimated 2,442 new products.
...while others are not so useful
Clearly we need to **weigh patents by their importance.**

- Q: Can we identify important patents and relate them to worker occupations using text alone?
1. We identify significant patents as those that:
   - are distinct from previous patents but are related to subsequent patents (i.e., they are novel and impactful)
   - **Implementation:** We need to measure the similarity between a given patent and prior and subsequent patents (within a window).

2. We identify the exposure of occupation $j$ to technology as
   - The number of important patents that are related to the tasks occupation $j$ performs
   - **Implementation:** We need to measure the similarity between a given patent and occupation task descriptions (ONET/DOT)
Summary Report for:
13-2072.00 - Loan Officers

Evaluate, authorize, or recommend approval of commercial, real estate, or credit loans. Advise borrowers on financial status and payment methods. Includes mortgage loan officers and agents, collection analysts, loan servicing officers, and loan underwriters.

Sample of reported job titles: Business Banking Officer, Commercial Banker, Commercial Loan Officer, Corporate Banking Officer, Loan Officer, Mortgage Loan Officer, Mortgage Loan Originator, Portfolio Manager, Relationship Manager

View report: Summary | Details | Custom

Tasks | Technology Skills | Tools Used | Knowledge | Skills | Abilities | Work Activities | Detailed Work Activities | Work Context | Job Zone | Education | Credentials | Interests | Work Styles | Work Values | Related Occupations | Wages & Employment | Job Openings | Additional Information

- Analyze applicants’ financial status, credit, and property evaluations to determine feasibility of granting loans.
- Obtain and compile copies of loan applicants' credit histories, corporate financial statements, and other financial information.
- Meet with applicants to obtain information for loan applications and to answer questions about the process.
- Explain to customers the different types of loans and credit options that are available, as well as the terms of those services.
- Review loan agreements to ensure that they are complete and accurate according to policy.
- Approve loans within specified limits, and refer loan applications outside those limits to management for approval.
- Handle customer complaints and take appropriate action to resolve them.
- Stay abreast of new types of loans and other financial services and products to better meet customers’ needs.
- Review and update credit and loan files.
- Submit applications to credit analysts for verification and recommendation.
- Compute payment schedules.
- Analyze potential loan markets and develop referral networks to locate prospects for loans.
- Set credit policies, credit lines, procedures and standards in conjunction with senior managers.
- Confer with underwriters to aid in resolving mortgage application problems.
- Market bank products to individuals and firms, promoting bank services that may meet customers' needs.
- Work with clients to identify their financial goals and to find ways of reaching those goals.
- Negotiate payment arrangements with customers who have delinquent loans.
- Prepare reports to send to customers whose accounts are delinquent, and forward irreconcilable accounts for collector action.
Occupation Task Description: Example

Summary Report for:
19-3011.00 - Economists

Conduct research, prepare reports, or formulate plans to address economic problems related to the production and distribution of goods and services or monetary and fiscal policy. May collect and process economic and statistical data using sampling techniques and econometric methods.


Also see: Environmental Economists

Tasks

Study economic and statistical data in area of specialization, such as finance, labor, or agriculture.

Conduct research on economic issues and disseminate research findings through technical reports or scientific articles in journals.

Compile, analyze, and report data to explain economic phenomena and forecast market trends, applying mathematical models and statistical techniques.

Supervise research projects and students' study projects.

Teach theories, principles, and methods of economics.

Study the socioeconomic impacts of new public policies, such as proposed legislation, taxes, services, and regulations.

Formulate recommendations, policies, or plans to solve economic problems or to interpret markets.

Explain economic impact of policies to the public.

Provide advice and consultation on economic relationships to businesses, public and private agencies, and other employers.

Forecast production and consumption of renewable resources and supply, consumption, and depletion of non-renewable resources.

Develop economic guidelines and standards and prepare points of view used in forecasting trends and formulating economic policy.

Knowledge

Mathematics — Knowledge of arithmetic, algebra, geometry, calculus, statistics, and their applications.
Method and compositions are provided for replication and expression of exogenous genes in microorganisms. Plasmids or virus DNA are cleaved to provide linear DNA having ligatable termini to which is inserted a gene having complementary termini, to provide a biologically functional replicon with a desired phenotypic property. The replicon is inserted into a microorganism cell by transformation. Isolation of the transformants provides cells for replication and expression of the DNA molecules present in the modified plasmid. The method provides a convenient and efficient way to introduce genetic and physiological changes in microorganisms.
PROCESS FOR PRODUCING BIOLOGICALLY FUNCTIONAL MOLECULAR CHIMERAS

The invention was supported by generous grants of NIH, NSF and the American Cancer Society.

CROSS-REFERENCE TO RELATED APPLICATIONS

This application is a continuation-in-part of application Ser. No. 959,288, filed Nov. 9, 1978, which is a continuation of application Ser. No. 687,430 filed May 17, 1976, now abandoned, which was a continuation-in-part of application Ser. No. 520,691, filed Nov. 4, 1974, now abandoned.

BACKGROUND OF THE INVENTION

1. Field of the Invention

Although transfer of plasmids among strains of E. coli and other Enterobacteriaceae has long been accomplished by conjugation and/or transduction, it has not been previously possible to selectively introduce particular species of plasmid DNA into these bacterial hosts or other microorganisms. Since microorganisms that have been transformed with plasmid DNA contain autonomously replicating extrachromosomal DNA species having the genetic and molecular characteristics of the parent plasmid, transformation has enabled the selective cloning and amplification of particular plasmid genes.

The ability of genes derived from totally different biological classes to replicate and be expressed in a particular microorganism permits the attainment of functional biological activities by combining genes from different species.

The process of this invention employs novel plasmids, which are formed by inserting DNA having one or more intact genes into a plasmid in such a location as to permit retention of an intact replicator locus and system (replicon) to provide a recombinant plasmid molecule. The recombinant plasmid molecule will be referred to as a “hybrid” plasmid or plasmid “chimera.” The plasmid chimera contains genes that are capable of expressing at least one phenotypical property. The plasmid chimera is used to transform a susceptible and competent microorganism under conditions where transformation occurs. The microorganism is then grown under conditions which allow for separation and harvesting of transformants that contain the plasmid chimera.

The process of this invention will be divided into the following stages:

I. preparation of the recombinant plasmid or plasmid chimera;

II. transformation or preparation of transformants;

and

III. replication and transcription of the recombinant plasmid in transformed bacteria.

Preparation of Plasmid Chimera

In order to prepare the plasmid chimera, it is necessary to have a DNA vector, such as a plasmid or phage, which can be cleaved to provide an intact replicator locus and system (replicon), where the linear segment has ligatable termini in a nucleotide sequence modified to provide base pairs for covalent ligation to the DNA segment containing the desired gene.
Approach 1: Represent document as sparse word vectors

- For two documents $i$ and $j$, we construct $V_i$ and $V_j$ as a (sparse) word vector of length $W$ (i.e. the size of the set union for terms in $(i,j)$)
  - Example: $D1 = \{\text{dog, eat, food}\}$ and $D2 = \{\text{cat, eat, food}\}$ leads to $V_1 = [1, 0, 1, 1]$ and $V_2 = [0, 1, 1, 1]$

- This ‘bag of words’ approach works well when the two documents are written in the same ‘language’, for instance when they contain well defined technical terms.

- We can measure similarity across documents based on a distance measure (cosine similarity) between $V_1$ and $V_2$.
  - We will use this approach when measuring the distance between two patent documents.
  - Not all words are equally informative, so we need appropriate weights.
Assigning weights to individual words

• Not all words are equally informative. Similar documents should share **uncommon** words

  ▶ The challenge is isolate these important terms. For example: word ‘electricity’ first appears in a patent in 1880; it should be weighted differently in 1880 than if it appears in 1980.

• Weigh word $w$ in patent document $d$ by

$$
\frac{f_{w,d}}{\sum_{w'} f_{w',d}} \times \log \left( \frac{\# \text{ documents before } t}{1 + \# \text{ documents before } t \text{ that include term } w} \right)
$$

Term Frequency (TF)  Backward Inverse Document Frequency (BIDF)

▶ TF: how important is word $w$ to document $d$

▶ BIDF: how much information provided by word $w$

▶ $TFBIDF_{w,d} = TF_{w,d} \times BIDF_{w,d}$

• We then compute cosine similarities using $V_{i,t} = TFBIDF_{i,t}$. 
Patent Similarity Example: Sewing Machine

Connection indicates similarity in excess of 50%
Patent Similarity Example: Moving Pictures

Apparatus for Exhibiting Photographs of Moving Objects (493,426)

- Camera lantern (546,093)
- Vitascope (578,185)
- Machine for exhibiting and taking pictures (553,369)
- Kinetographic camera (593,376)
- Roll holder camera and picture exhibitor (542,334)
- Phantoscope (586,953)
- Projecting kinetoscope (707,934)
- Kinetographic camera (560,800)
- Vitascope (673,992)
- Phantoscope (629,063)

- Picture exhibitor (380,977)
- Method of producing instantaneous photographs (452,966)
- Camera (528,140)
Patent Similarity Example: Telephone

Improvement of Transmitters and Receivers for Electric Telegraphs (161,739)

- Improvement in telephonic telegraph-receivers (178,399)
- Telephone relay (255,333)
- Telegraph repeater (231,477)
- Quadruplex Telegraph (254,297)
- Electric Burglar-Alarm (225,271)
- Telegraphic repeater (250,774)
- Electric signal-bell (228,851)
- Improvement in electro magnetic alarms (197,416)
- Improvement in telegraphy (174,465)
- Telephonic system (284,594)
- Testing and breaking circuits (260,043)

- Improvement in telegraph-receivers (161,739)
- Telephone relay (255,333)
- Telegraph repeater (231,477)
- Quadruplex Telegraph (254,297)
- Electric Burglar-Alarm (225,271)
- Telegraphic repeater (250,774)
- Electric signal-bell (228,851)
- Improvement in electro magnetic alarms (197,416)
- Improvement in telegraphy (174,465)
- Telephonic system (284,594)
- Testing and breaking circuits (260,043)
However, the previous approach does not deal with synonyms.

- For example, if $D_3 = \{\text{canine, eat, food}\}$ then

\[
\text{distance}(D_1, D_2) = \text{distance}(D_1, D_3)
\]

- This creates a bias towards low similarity if the two documents use different vocabulary
  - e.g. patent documents vs occupation task descriptions

- Our (current) approach: use word embeddings (e.g. word2vec).
  - Each word $x_k$ is represented as a 300-dimensional vector (arbitrary basis).
  - The (cosine) distance between two vectors is related to the probability they are synonyms (i.e., they are used in the same context within a set of documents).
  - We use word vectors provided by Pennington et al. (2014) that were trained on 42 billion word tokens of web data from Common Crawl.
Text Analysis Basics: Representing Text as Data

**Approach 2:** Represent documents as weighted averages of word vectors:

- Each document is a weighted average of word vectors

\[ V_i = \sum_{x_k \in A_i} w_{i,k} x_k \]

- Now, \( V_i \) is no longer sparse but has lower dimensionality than before.

- Here \( w_{i,k} \) is the term-frequency-inverse-document-frequency (TFIDF) defined as

\[ w_{i,k} \equiv TF_{i,k} \times IDF_k \]

- As before, a word will receive a higher weight if it appears multiple times in a document and if it is relatively infrequent.

- IDF is computed separately for patents and job descriptions.
## Patents and Occupations: Similarity Examples

<table>
<thead>
<tr>
<th>Occupation</th>
<th>US Patent #</th>
<th>Patent Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>Loan Interviewers and Clerks (434131)</td>
<td>4,736,294</td>
<td>Data processing methods and apparatus for managing vehicle financing</td>
</tr>
<tr>
<td></td>
<td>5,611,052</td>
<td>Lender direct credit evaluation and loan processing system</td>
</tr>
<tr>
<td></td>
<td>5,673,402</td>
<td>Computer system for producing an illustration of an investment repaying a mortgage</td>
</tr>
<tr>
<td></td>
<td>5,870,721</td>
<td>System and method for real time loan approval</td>
</tr>
<tr>
<td></td>
<td>5,940,811</td>
<td>Closed loop financial transaction method and apparatus</td>
</tr>
<tr>
<td>Cashiers (412011)</td>
<td>4,541,057</td>
<td>System for performing combined financial transactions with single dispensing of cash</td>
</tr>
<tr>
<td></td>
<td>4,814,985</td>
<td>Sales limit indicator for an electronic cash register</td>
</tr>
<tr>
<td></td>
<td>5,055,657</td>
<td>Vending type machine dispensing a redeemable credit voucher upon payment interrupt</td>
</tr>
<tr>
<td></td>
<td>5,085,435</td>
<td>Method of using a random number supplier for the purpose of reducing currency handling</td>
</tr>
<tr>
<td></td>
<td>5,224,162</td>
<td>Electronic cash system</td>
</tr>
</tbody>
</table>
## Patents and Occupations: Similarity Examples (cont)

<table>
<thead>
<tr>
<th>Occupation</th>
<th>US Patent #</th>
<th>Patent Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>Packers and Packagers, Hand (537064)</td>
<td>3,876,858</td>
<td>Shrink-film hole-burning device</td>
</tr>
<tr>
<td></td>
<td>3,931,701</td>
<td>Automatic produce-bagging machine that uses factory-roll polyethylene net tubing</td>
</tr>
<tr>
<td></td>
<td>4,098,398</td>
<td>Container for recycle of motor oil</td>
</tr>
<tr>
<td></td>
<td>4,266,698</td>
<td>Opening arrangement for packing containers of thin plastic film together with a packing container provided with the opening arrangement</td>
</tr>
<tr>
<td></td>
<td>4,912,913</td>
<td>Bag sealing machine</td>
</tr>
<tr>
<td>Shipping, Receiving, and Traffic Clerks (435071)</td>
<td>5,233,532</td>
<td>System for mailing and collecting items</td>
</tr>
<tr>
<td></td>
<td>5,481,464</td>
<td>System for collecting and shipping items</td>
</tr>
<tr>
<td></td>
<td>5,656,799</td>
<td>Automated package shipping machine</td>
</tr>
<tr>
<td></td>
<td>5,666,493</td>
<td>System for managing customer orders and method of implementation</td>
</tr>
<tr>
<td></td>
<td>6,148,291</td>
<td>Container and inventory monitoring methods and systems</td>
</tr>
</tbody>
</table>
So far, we have created distance measures between patents and between patents to occupations.

Next steps:

1. Create indices of technological change
2. Identify occupation exposures
Measuring patent importance

- Important patents are both novel (fewer past connections) and impactful (have more future connections)
- Our importance score measures both impact and novelty

\[ \xi_{j}^{0,\tau} = \frac{FS_{j}^{0,\tau}}{BS_{j}^{0,\tau}} \]

- **Future Impact (forward similarity)**

\[ FS_{j}^{0,\tau} = \sum_{i \in F} \rho_{j,i}, \]

- **Novelty (backward similarity):**

\[ BS_{j}^{0,\tau} = \sum_{i \in B} \rho_{j,i}, \]

\( B_{j,\tau} \) and \( F_{j,\tau} \) is set of patents granted in the \( \tau \) calendar years prior to, and following, \( j \)’s application year, respectively.
Significant Patents, Examples

Airplane patent is at the 99th percentile in terms of our importance measure (it has 19 cites over its lifetime)
The present invention is an apparatus, system, and method for providing reservations for restroom use. In one embodiment, a passenger on an airplane may submit a reservation request to the system for restroom use. The reservation system determines when the request can be accommodated and notifies the passenger when a restroom becomes available. The system improves airline safety by minimizing the time passengers spend standing while an airplane is in flight.
Creating time-series indices

Next, we construct indices of technological progress.

- One issue: part of the time-variation in our importance indicator may capture shifts in language (or differences in OCR quality)
  - Solution: remove year FEs, denote adjusted quality measure by $\tilde{q}$.
  - Assumption: shifts in language should affect all patents symmetrically.

- Our approach: count the # of patents at the right tail of the distribution (breakthroughs)

$$\eta_t = \frac{1}{\kappa_t} \sum_{i \in \Gamma_t} 1(\tilde{q}_{i,t} \geq \tilde{q}_{90})$$

scale by US population $\kappa_t$
Breakthrough patents—based on breakthrough counts

# of breakthrough patents per 1000 people

year

1840 1860 1880 1900 1920 1940 1960 1980 2000

0 0.02 0.04 0.06 0.08 0.1
Breakthrough patents and aggregate productivity

A. Early period (1889–1957)
Labor Productivity (Kendrick)

B. Post-war period (1948–2007)
Total Factor Productivity (BLS)

- Figures plot increase in **average** productivity to a one-standard deviation increase in our index
Breakthrough patents and industry productivity

A. Early period (1899–1954)
Labor Productivity (Kendrick)

B. Post-war period (1987–2016)
Total Factor Productivity (NAICS 4-digit)

- Figures plot increase in **average** productivity to a one-standard deviation increase in our index
Occupation-specific indices of technical change

- Follow similar approach as before, with some adjustments:
  - Denote by $\rho_{i,j}$ each element of the patent (i) X occupation (j) matrix.
  - To account for shifts in language, remove time FEs from all elements.
  - Impose sparsity: set the bottom 80% of patent-occupation pairs to zero.
  - Re-scale the remainder 20% of pairs so they range between (0,1).
  - Denote the adjusted similarity measure by $\tilde{\rho}_{i,j}$.

- Our index then sums up occupation exposures across breakthrough patents:

$$\eta_{j,t} = \frac{1}{\kappa_t} \sum_{i \in \Gamma_t^c} \tilde{\rho}_{i,j} \times 1(\tilde{q}_{i,t} \geq \tilde{q}_{p90})$$
Technological Change and Occupations: Examples
Breakthrough patents and occupation outcomes

- Figures plot change in annualized employment/wage growth over time, in response to a one-standard deviation increase in our index.
Technological Change and Occupations: Composition

![Graph showing changes in occupational composition over time.](image)

Legend:
- Service
- Production, Transportation, & Material Moving
- Management, Business, & Financial
- Education, Legal, Community Service, Arts, & Media
- Sales & Office
- Natural Resources, Construction, & Maintenance
- Healthcare Practitioners
- Computer, Engineering & Science

Year:
- 1850
- 1860
- 1870
- 1880
- 1890
- 1900
- 1910
- 1920
- 1930
- 1940
- 1950
- 1960
- 1970
- 1980
- 1990
- 2000
Exposures of tasks to technological change

• One way to summarize trends is to examine how technical change is related to occupations performing different task types.

• Use task category scores \((T_{j,w})\) in Acemoglu and Autor (2011)
  
  ▶ Tasks \(w\) fall into: non-routine cognitive (analytical), non-routine cognitive (interpersonal), non-routine manual (interpersonal) non-routine manual (physical), routine cognitive, or routine manual.
  
  ▶ \(T\) normalized to mean zero and unit standard deviation.

• The task innovation exposure score for task \(w\) in year \(t\) is then given by

\[
\lambda_{w,t} = \sum_j \eta_{j,t} \times T_{j,w} \times \omega_j
\]

Here \(\omega_j\) is SOC labor-supply weights
Exposures of tasks to technological change

![Graph showing the exposures of tasks to technological change over time for different types of tasks such as Non-routine Cognitive (Analytical), Non-routine Cognitive (Interpersonal), Non-routine Manual (Interpersonal), Non-routine Manual (Physical), Routine Cognitive, and Routine Manual. The x-axis represents the years from 1850 to 2000, and the y-axis represents the exposures. The graph indicates the trends and changes in exposures over the years.]
Breakthrough Patents: Breakdown by Technology Classes

Decade


- Agriculture and Food (A0, A2)
- Electricity and Electronics (H0)
- Health and Entertainment (A6)
- Lighting, Heating, Nuclear (F2, G2)
- Transportation (B6)
- Chemicals and Metallurgy (C)
- Engineering, Construction, and Mining (E0, E2, F0, F1)
- Instruments, Information (G, Y1)
- Manufacturing Process (B0, B2, B3, B4, B8, D0, D1, D2)
- Weapons (F4)
Exposures of tasks to technological change

Manufacturing Process (B0–8,D0–2)
Exposures of tasks to technological change

Transportation (B7)

- Non-routine Cognitive (Analytical)
- Non-routine Cognitive (Interpersonal)
- Non-routine Manual (Interpersonal)
- Non-routine Manual (Physical)
- Routine Cognitive
- Routine Manual
Exposures of tasks to technological change

Electricity and Electronics (H0)
Exposures of tasks to technological change

Instruments and Information (G,Y1)
Conclusion

- We create text-based indicators of the exposure of occupations to technological change.
- Our indices are negatively correlated with future employment and wage growth.
- Recent technological wave appears qualitatively different than previous waves: it is a lot more related to occupations emphasizing cognitive tasks than before.
- Open questions and next step:
  - Average outcomes obscure heterogeneity
  - Relate innovation to individual worker outcomes using Census data.