

Using Language Models to Understand Wage Premia

Sarah H. Bana

Stanford Digital Economy Lab

Institute for Human-Centered Artificial Intelligence

December 8, 2021

Massive Amounts of Data Created by Our Existence Online

One activity increasingly conducted online is job search

- Firms post jobs and interact with candidates
- Workers use job boards to identify and apply for vacancies

Large scale information about workers and firms, in close to real-time and preserving variation

Does the Text Content of a Job Posting Predict the Salary Offered?

And if so, How Do We Measure What Text Matters?

Growing evidence that job tasks and skills affect earnings

- Within occupation measures of tasks have predictive power for earnings ([Autor and Handel, 2013](#))
 - Very costly to capture through traditional (survey-based) methods

Does the Text Content of a Job Posting Predict the Salary Offered?

And if so, How Do We Measure What Text Matters?

Growing evidence that job tasks and skills affect earnings

- Within occupation measures of tasks have predictive power for earnings ([Autor and Handel, 2013](#))
 - Very costly to capture through traditional (survey-based) methods

Online data provides an opportunity to better understand the labor market

- Job title level heterogeneity is important for explaining applicant behavior ([Marinescu and Wolthoff, 2020](#))

Does the Text Content of a Job Posting Predict the Salary Offered?

And if so, How Do We Measure What Text Matters?

Growing evidence that job tasks and skills affect earnings

- Within occupation measures of tasks have predictive power for earnings ([Autor and Handel, 2013](#))
 - Very costly to capture through traditional (survey-based) methods

Online data provides an opportunity to better understand the labor market

- Job title level heterogeneity is important for explaining applicant behavior ([Marinescu and Wolthoff, 2020](#))
- Skills required relate to average wages of professionals across MSA-occupation cells ([Deming and Kahn, 2018](#))

Job Postings Give Us Insight Into Workers' Skills and Activities

"Sales Associate\n\nGap\n\n-\n\nSanta Clara, CA 95050\n\nPart-time, Temporary\n\nAs a Brand Associate, youre an integral part of our team and bring our brand to life for our customers. Youre responsible for engaging and connecting with our customers by providing excellent customer service resulting in brand loyalty. Youre an expert in product and use your knowledge and experience to educate, inform, inspire and wardrobe the customer. Through collaboration with your leadership team, youll deliver a best-in-class customer experience using an omni-channel approach.\n\nWHAT YOU'LL DO\n\n* Consistently treat all customers and employees with respect and contribute to a positive work environment.\n\n* Promote loyalty by educating customers about our loyalty programs.\n\n* Seek out and engage with customers to drive sales and service using suggestive selling.\n\n* Enhance customer experience using all omni-channel offerings.\n\n* Be accountable to personal goals which contribute to overall store goals and results.\n\n* Support sales floor, fitting room, cash wrap, back of house, as required.\n\n* Maintain a neat, clean and organized work center.\n\n* Handle all customer interactions and potential issues returns courteously and professionally.\n\n* Execute operational processes effectively and efficiently.\n\n\nJob Types: Part-time, Temporary\n\nsave this job a"

Sales Associate at Gap in Santa Clara, CA

Job Postings Give Us Insight Into Workers' Skills and Activities

"Sales Associate\n\nGap\n\n-\n\nSanta Clara, CA 95050\n\nPart-time, Temporary\n\nAs a Brand Associate, you're an integral part of our team and bring our brand to life for our customers. You're responsible for engaging and connecting with our customers by providing excellent customer service resulting in brand loyalty. You're an expert in product and use your knowledge and experience to educate, inform, inspire and wardrobe the customer. Through collaboration with your leadership team, you'll deliver a best-in-class customer experience using an omni-channel approach.\n\nWHAT YOU'LL DO\n\n* Consistently treat all customers and employees with respect and contribute to a positive work environment.\n\n* Promote loyalty by educating customers about our loyalty programs.\n\n* Seek out and engage with customers to drive sales and service using suggestive selling.\n\n* Enhance customer experience using all omni-channel offerings.\n\n* Be accountable to personal goals which contribute to overall store goals and results.\n\n* Support sales floor, fitting room, cash wrap, back of house, as required.\n\n* Maintain a neat, clean and organized work center.\n\n* Handle all customer interactions and potential issues/returns courteously and professionally.\n\n* Execute operational processes effectively and efficiently.\n\nJob Types: Part-time, Temporary\n\nsave this job a"

Sales Associate at Gap in Santa Clara, CA

You're responsible for engaging and connecting with our customers by providing excellent customer service resulting in brand loyalty.

What You'll Do:

- Consistently treat all customers and employees with respect and contribute to a positive work environment.
- Promote loyalty by educating customers about our loyalty programs.
- Seek out and engage with customers to drive sales and service using suggestive selling.
- Be accountable to personal goals which contribute to overall store goals and results.
- Support sales floor, fitting room, cash wrap, back of house, as required.

Job Postings Give Us Insight Into Workers' Skills and Activities

Seasonal Retail Commission Jewelry Sales, Full Time: Valley Fair\nMacy's\nSanta Clara, CA\nJob Overview:\n\nThe Seasonal Jewelry Sales associate is responsible for providing outstanding customer service in the Jewelry & Watch Complex. This includes meeting hourly sales and Star Rewards goals on a personal and departmental basis, demonstrating superior product knowledge to customers, building a clientele using the clientele system, offering the customer our extended service plan and creating a shopping experience that will make the customer feel welcome and comfortable.\n\nThe Seasonal Jewelry Sales associate will participate in ongoing training via the IPAD as well as attend vendor and trainer facilitated trainings. Additionally, all Jewelry Sales associates participate in all aspects of ensuring the department is operating efficiently including but not limited to processing goods, merchandising, taking price changes, setting and signing sales, doing RTVs, BOPS & Fulfillment.\n\nIn order to present our customers with the best holiday shopping experience, many of our Seasonal Jewelry Sales Associates arrive prior to store opening and remain after closing to ensure we are ready to make Macy's magic. Seasonal associates will have the opportunity to work a flexible schedule on a temporary basis, which may include early mornings, evenings, weekends and busy events such as the day after Thanksgiving and the day after Christmas. And don't forget - just in time for holiday shopping, you will receive an employee discount of up to 20% starting your first day!\n\nEssential Functions:\n\n* Be proficient in use of all POS and MPOS systems including Search and Send, My Client and More@Macys tablet app\n\n* Be proactive in assisting customers who are using devices to shop and compare, whether Macys devices or their own\n\n* Assist customers in all aspects of service, and qualify customer needs, figure out what's right for them by using steps of MAGIC & High Touch Selling\n\n* Offer to put purchase on customers Macys charge account\n\n* Suggest additional merchandise to compliment customer selection\n\n* Maintain a professional attitude with sincerity and enthusiasm that demonstrates Macys commitment to our customer\n\n* Participate in pre-selling and sales driving events including trunk shows to maximize sales\n\n* Stay informed on current promotional events and sales\n\n* Offer and promote benefits of extended service plan to all customers purchasing Fine Jewelry & Watches\n\n* Use clientele program to maintain customer profile and contact information to increase personal sales and build solid customer base\n\n* Perform all aspects of fulfillment process; including, but not limited to: picking, packing, labeling, shipping and BOPS\n\n* Offer to call other locations if merchandise is unavailable within store or offer Search & Send\n\n* Be knowledgeable of and perform sales support functions related to Jewelry (POS procedures, iPad, Ops, AP)\n\n* Ensure proper processing, presentation, organization, storing, and replenishment of stock. Maintain Recovery & Fill in Standards\n\n* Perform other duties as assigned, including but not limited to watch sizing, ear piercing\n\n* Adhere to asset protection programs and procedures to ensure audit compliance\n\n* Perform functions in an efficient manner, as directed by the supervisor\n\n* Regular, dependable attendance and punctuality\n\nQualifications:\n\nEducation/Experience:\n\n* High school diploma or equivalent preferred.\n\n* Previous selling experience is required, preferably in fine jewelry.\n\nCommunication Skills:\n\n* Effective written and verbal skills, ability to interpret instructional documents such as safety rules, operating and maintenance instructions, and procedure manuals.\n\n* Excellent written and verbal communication skills.\n\nMathematical Skills:\n\n* Basic math functions such as addition, subtraction, multiplication, and division.\n\n* Able to use a calculator.\n\nReasoning Ability:\n\n* Self-starter, able to work independently and as part of a team and must have good time management skills.\n\nPhysical Demands:\n\n* This position involves constant moving, talking, hearing, reaching, and standing.\n\n* Involves standing for at least two consecutive hours.\n\n* Involves lifting at least 30 lbs.\n\n* May occasionally involve stooping, kneeling, crouching, and climbing ladders.\n\n* Vision abilities include close vision, color vision, depth perception, and focus adjustment.\n\nOther Skills:\n\n* Superior organizational and time management skills.\n\n* Must be able to multi-task in a fast-paced environment.\n\n* Must be able to build relationships and influence others.\n\n* Must possess a thirst for knowledge.\n\n* Ability to collaborate and function as a member of a team.\n\n* Must possess a strong sense of urgency.\n\n* Should be comfortable with the use of computers and frequent use of RF equipment.\n\nWork Hours:\n\n* Flexible with scheduling and available to work retail hours, which may include day, evening, and weekend hours.\n\n* Must be able to work overtime hours as needed.\n\n* This job requires the ability to work in a fast-paced environment and to be able to work in a fast-paced environment and to be able to work in a fast-paced environment.

Sales Associate at Macy's in Santa Clara, CA

Job Overview: The Seasonal Jewelry Sales associate is responsible for providing outstanding customer service in the Jewelry & Watch Complex. This includes meeting hourly sales and Star Rewards goals on a personal and departmental basis, demonstrating superior product knowledge to customers, building a clientele using the clientele system, offering the customer our extended service plan and creating a shopping experience that will make the customer feel welcome and comfortable.

Essential Functions:

- Be proficient in use of all POS and MPOS systems including My Client and More@Macys tablet app

Education/Experience:

- High school diploma or equivalent preferred.
- Previous selling experience required, preferably in fine jewelry.

Do These Roles Pay Different Salaries? How Different?

Sales Associate at Gap in Santa Clara, CA

You're responsible for engaging and connecting with our customers by providing excellent customer service resulting in brand loyalty.

What You'll Do:

- Consistently treat all customers and employees with respect and contribute to a positive work environment.
- Promote loyalty by educating customers about our loyalty programs.
- Seek out and engage with customers to drive sales and service using suggestive selling
- Be accountable to personal goals which contribute to overall store goals and results.
- Support sales floor, fitting room, cash wrap, back of house, as required.

Sales Associate at Macy's in Santa Clara, CA

Job Overview: The Seasonal Jewelry Sales associate is responsible for providing outstanding customer service in the Jewelry & Watch Complex. This includes meeting hourly sales and Star Rewards goals on a personal and departmental basis, demonstrating superior product knowledge to customers, building a clientele using the clientele system, offering the customer our extended service plan and creating a shopping experience that will make the customer feel welcome and comfortable.

Essential Functions:

- Be proficient in use of all POS and MPOS systems including My Client and More@Macys tablet app

Education/Experience:

- High school diploma or equivalent preferred.
- Previous selling experience required, preferably in fine jewelry.

Do These Roles Pay Different Salaries? How Different?

Sales Associate at Gap in Santa Clara

- **Salary from the metadata:**

\$32,500

- **Salary from the prediction model:**

\$33,479

Sales Associate at Macy's in Santa Clara

- **Salary from the metadata:**

\$41,500

- **Salary from the prediction model:**

\$40,881

Growing Set of Tools Available to Researchers

A number of papers have started to use job postings as a source of information on firms' demands

Growing Set of Tools Available to Researchers

A number of papers have started to use job postings as a source of information on firms' demands

I add a few additional tools

- New data source (Greenwich.HR) on salaries from the metadata of postings
 - Salaries for over 60 percent of job postings, and increasing over time
- Natural language processing (NLP) methods to distill the full text of job postings

NLP Turns Words into Context-Dependent Vectors

- Modern NLP models (e.g. BERT) help turn words into context-dependent vectors
- Having salaries as an outcome turns this into a classical supervised learning problem
 - Instead of counting words ourselves, and determining what may or may not be important, we can let the data decide

Elements Create a 21st Century Version of Hedonic Regression

In the Style of Rosen (1974)

Both sides of the market are heterogeneous → equilibrium prices

- Prices shed light on costs

Distribution of valuations for any job attribute (skills, amenities, etc.) in postings

- Major complement to the labor force statistics currently collected

Limitations

- Do not distinguish between signaling and human capital
- Prior to the negotiation process

This Paper: Use NLP to Predict Salaries

*To the Best of My Knowledge, First Paper to Use **Full Text** of Job Postings*

Model

- Supervised learning model, currently trained on 857,477 postings
 - Key layer: BERT embeddings (context dependent dimensions)
- Predicts salaries from job postings using the text with out-of-sample $R^2 = 0.84$
 - 21 percent (15 p.p.) increase over a model with occupation x MSA fixed effects

Illustrate the power of words for a representative sample of postings

This Paper: Use NLP to Predict Salaries

To the Best of My Knowledge, First Paper to Use **Full Text** of Job Postings

Application

- Create a counterfactual posting by adding a marginal characteristic
- Predict the salary of the counterfactual posting
- Compare the counterfactual posting's salary to the original posting's salary
→ Difference is an estimate of the wage premium for that job
- Applied to online certifications
 - First independent estimates

Approach scalable to any marginal job characteristic

Outline

- Data
- Model
- Empirical Approach
- Application to Certifications

Data

Two Sources of Job Posting Data

Burning Glass Technologies

- Job postings collected from over 40,000 online job boards and company websites
- 200+ million postings over the past decade
- Used by researchers for a variety of applications
 - Recessions and technological change ([Hershbein and Kahn, 2016](#))
 - Skill requirements ([Deming and Kahn, 2018](#))
 - AI skills demand ([Alekseeva et al., 2021](#))
 - Monopsony ([Azar et al., 2020](#))
 - General Purpose Technology (GPT) detection ([Goldfarb et al., 2021](#))
- Key attribute: Posting text
 - To the best of my knowledge, other researchers have only used keywords

Two Sources of Job Posting Data

Greenwich.HR

- Job postings from millions of sources
- 62 million postings between April 2019 - Sept 2020
 - 37 million (\approx 60 percent) have salaries
- Salary data that comes from the metadata of the posting

Some Job Board User Interfaces Ask Recruiters To Input Salaries

What's the pay?

Tell job seekers the pay and receive up to two times more applications.



What is the pay for this job?

Exact rate

\$ 75.00 per hour

Some Job Board User Interfaces Ask Recruiters To Input Salaries

What's the pay?
Tell job seekers the pay and receive up to two times more applications.



What is the pay for this job?

Exact rate ▾

\$ 75.00 per hour ▾

What's the pay?
Tell job seekers the pay and receive up to two times more applications.



Range

- Starting at
- Up to
- ✓ Exact rate

\$ 75.00 per hour ▾

Some Job Board User Interfaces Ask Recruiters To Input Salaries

What's the pay?

Tell job seekers the pay and receive up to two times more applications.



What is the pay for this job?

Exact rate ▾

\$ 75.00 per hour ▾

What's the pay?

Tell job seekers the pay and receive up to two times more applications.



Range
Starting at
Up to
✓ Exact rate

\$ 75.00 per hour ▾

Company description

Tell potential applicants what your company does and what it's like to work there.

Compensation

Show estimate from LinkedIn members for Mark, Manager at Flexis in Greater Atlanta Area ⓘ

I'll provide my own

Base salary

| | | | | |
|-----|----------|---|----------|----------|
| USD | \$88,000 | - | \$90,000 | Per year |
|-----|----------|---|----------|----------|

Additional compensation

| | | | | |
|-----|----------|---|----------|----------|
| USD | \$20,000 | - | \$27,000 | Per year |
|-----|----------|---|----------|----------|

Base salary and additional compensation will be added together on your job.

There is Evidence of This Data on the Jobseeker's Side



Jobs

Upload/Build Resume

Salaries & Advice ▾

Recommended Jobs

Data Scientist in Palo Alto, CA



Job Type ▾

Date Posted ▾

Pay ▾

Distance ▾

Easy Apply Only

621 Data Scientist Jobs in Palo Alto, CA

Create Job Alert.

Get similar jobs sent to your email

Save

Sort by: Relevancy | Date



TODAY

Data Scientist

Analysts | CA - Redwood City | Contractor

Easy Apply



Data Scientist

Analysts | Redwood City, CA | Contractor

Any

\$20,000+

\$40,000+

\$60,000+

\$80,000+

\$100,000+

\$120,000+



How Selected are the Posted Salaries?

Discussion and Comparison to the Current Population Survey (CPS)

Salaries collected from metadata may differ in their selection from explicitly posted salaries

How are salaries collected from the metadata selected?

- Partially a function of the posting website's listing protocols
 - More likely a decision at the firm level as opposed to the job level

How Selected are the Posted Salaries?

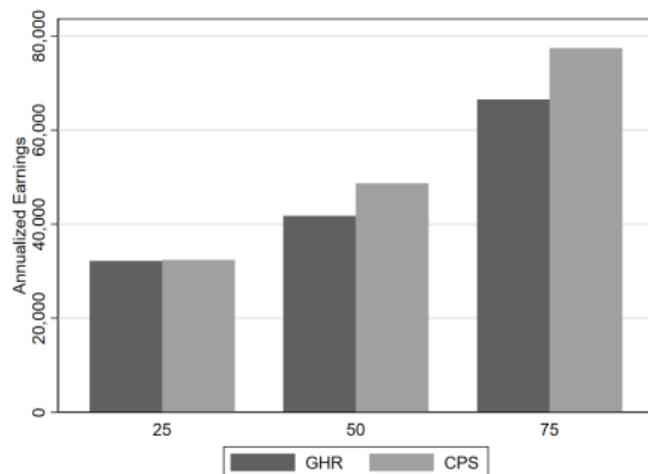
Discussion and Comparison to the Current Population Survey (CPS)

Salaries collected from metadata may differ in their selection from explicitly posted salaries

How are salaries collected from the metadata selected?

- Partially a function of the posting website's listing protocols
 - More likely a decision at the firm level as opposed to the job level

CPS measures usual weekly earnings, before taxes, and including any overtime pay, commission, or tips usually received



Model

Two Innovations

1. Supervised machine learning
2. Natural language processing that turns words into vectors

Simple Regression for Salary Prediction

Starting point: Linear regression

- We might run a regression of important characteristics of job postings (skills, amenities, education levels, experience levels) on salaries
- But why not include interactions?
 - Even just pairwise interactions will yield too many covariates

Need a process allowing for interactions that have high weights in predicting salary

Supervised Machine Learning (ML)

Brief Introduction

Supervised learning excels at identifying patterns from data when both inputs (X 's) and outputs (Y 's) are given

Objective: Minimize loss function (same as OLS)

- With data-driven interactions, there is a potential for “overfitting”
- To ensure that the model is learning true patterns from the data, evaluate “out-of-sample” (postings that model has not seen before)

Equipped with Tools to Characterize Interactions

But How to Take Stock of Words?

- Count relevant words?

Equipped with Tools to Characterize Interactions

But How to Take Stock of Words?

- Count relevant words?
- Count relevant phrases?

Equipped with Tools to Characterize Interactions

But How to Take Stock of Words?

- Count relevant words?
- Count relevant phrases?

Issues:

- Likely to be a lot of sparse matrices
- From our postings earlier, “preferred” and “desired” imply the same concept

Computer Scientists Have Approaches to Turn Words Into Vectors

Word embeddings – vectors for words, such that words closer to one another are similar in meaning

In this paper, I use **BERT** by [Devlin et al. \(2018\)](#)

- Currently the basis for Google Search
- **Trained on** English Wikipedia (2,500M words) and BooksCorpus (800M words)
- **Trained using** Masked Language Modeling (Self-Supervised Learning)

This Approach Facilitates Turning a Posting into a Matrix

Each Token (Word) is Represented by a Vector

1. Each posting is tokenized (split into words or subwords)
2. Allocate the posting to the training (in-sample) or testing (out-of-sample) data
3. Train the model on the outcome, $\ln(\text{salary})$
4. Evaluate on the out-of-sample postings

► Model Structure

Model Performance

Table: Out-of-Sample Coefficient of Variation

| | Occupation FEs (limited sample) | Occupation FEs | Occupation FEs and MSA FEs | Occupation x MSA FEs | NLP Model |
|---------|------------------------------------|----------------|-------------------------------|-------------------------|-----------|
| R^2 | 0.612 | 0.590 | 0.621 | 0.695 | 0.843 |
| No. Obs | 203,007 | 214,281 | 214,281 | 214,281 | 214,281 |
| No. FEs | 784 | 785 | 1,614 | 41,455 | – |

Model Performance

Table: Out-of-Sample Root Mean Squared Error

| | Occupation FEs (limited sample) | Occupation FEs | Occupation FEs and MSA FEs | Occupation x MSA FEs | NLP Model |
|---------|------------------------------------|----------------|-------------------------------|-------------------------|-----------|
| RMSE | 0.323 | 0.330 | 0.318 | 0.317 | 0.202 |
| No. Obs | 203,007 | 214,281 | 214,281 | 214,281 | 214,281 |
| No. FEs | 784 | 785 | 1,614 | 41,455 | – |

Empirical Approach

Text Injection Experiments

Methods from Bana, Brynjolfsson, Rock and Steffen (2021)

1. Construct a set of counterfactual postings by adding a phrase to the end of a posting
2. Use the salary model to predict the salary of the counterfactual posting
3. Take the difference between the predicted salary of the counterfactual posting (with the additional characteristic) and the predicted salary of the original posting

Provides a premium associated with the characteristic

► Formal Treatment

Construct a Variety of Counterfactuals

We might not see many truck driver job postings with computer networking skills requested

- Limit the sample to occupations that have requested that characteristic in a prior time period

This approach also lends itself to heterogeneity

Application to Certifications

As of 2020, over 500k Online Certificates and Digital Badges

Exceeding the number of degrees and certificates from postsecondary institutions



coursera

Professional Certificates

Get job-ready for an in-demand career

Break into a new field like information technology or data science. No prior experience necessary to get started.



[Explore Certificates →](#)

As of 2020, over 500k Online Certificates and Digital Badges

Exceeding the number of degrees and certificates from postsecondary institutions



coursera

Professional Certificates

Get job-ready for an in-demand career

Break into a new field like information technology or data science. No prior experience necessary to get started.



[Explore Certificates →](#)

More accessible to larger swathes of the population

As of 2020, over 500k Online Certificates and Digital Badges

Exceeding the number of degrees and certificates from postsecondary institutions



coursera

Professional Certificates

Get job-ready for an in-demand career

Break into a new field like information technology or data science. No prior experience necessary to get started.



[Explore Certificates](#) →

More accessible to larger swathes of the population

- Shorter
- More narrowly focused
- Offered in a flexible time frame

May increase productivity

As of 2020, over 500k Online Certificates and Digital Badges

Exceeding the number of degrees and certificates from postsecondary institutions



coursera

Professional Certificates

Get job-ready for an in-demand career

Break into a new field like information technology or data science. No prior experience necessary to get started.



[Explore Certificates](#) →

More accessible to larger swathes of the population

- Shorter
- More narrowly focused
- Offered in a flexible time frame

May increase productivity **but** currently, no clear way to understand potential earnings consequences

As of 2020, over 500k Online Certificates and Digital Badges

Exceeding the number of degrees and certificates from postsecondary institutions



coursera

Professional Certificates

Get job-ready for an in-demand career

Break into a new field like information technology or data science. No prior experience necessary to get started.



[Explore Certificates →](#)

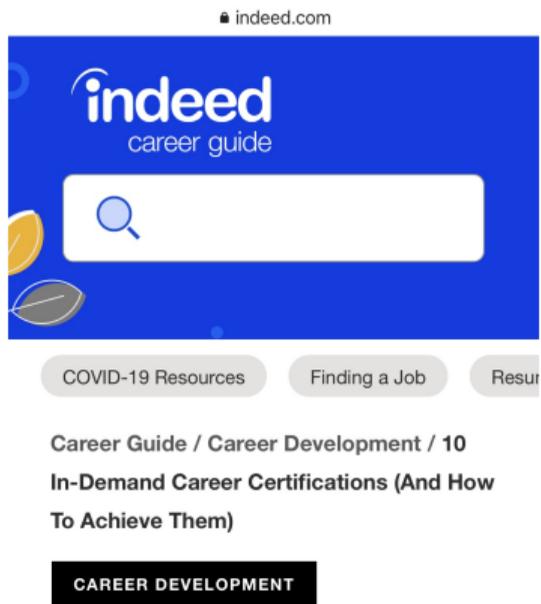
More accessible to larger swathes of the population

- Shorter
- More narrowly focused
- Offered in a flexible time frame

May increase productivity **but** currently, no clear way to understand potential earnings consequences

Growing at an unprecedented rate

Applied to In-Demand Certifications



- First independent estimates of the prices
 - Important due to the principal-agent problems inherent with certifications
- Measurement can be shared close to real-time and with heterogeneity

10 In-Demand Career Certifications

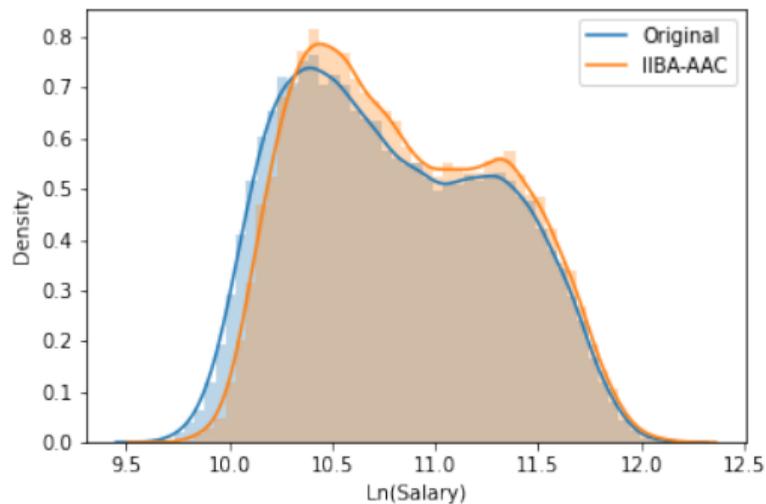
“In-Demand” Career Certifications

According to Indeed in 2021

| Category | Abbreviation | Certification Title | Cost (Lower) | Cost (Upper) |
|--------------------|--------------|---|-----------------|-----------------|
| Project Management | PMP | Project Management Professional | \$405 | \$555 |
| Project Management | CAPM | Certified Associate in Project Management | \$225 | \$300 |
| Business Analyst | CBAP | Certified Business Analysis Professional | \$475 | \$575 |
| Business Analyst | IIBA-AAC | IIBA Agile Analysis Certification | \$450 | \$575 |
| Supply Chain | CPIM | Certified in Production and Inventory Management | \$495 | \$690 |
| Supply Chain | CSCP | Certified Supply Chain Professional | \$695 | \$969 |
| Supply Chain | CLTD | Certified in Logistics, Transportation and Distribution | \$475 | \$625 |
| Computer Network | CCIE | Cisco Certified Internetwork Expert | \$2050 | \$2050 |
| Computer Network | CCNP | Cisco Certified Network Professional | \$300 | \$300 |

Example: International Institute of Business Analysis – Agile Analysis Certification (IIBA-AAC)

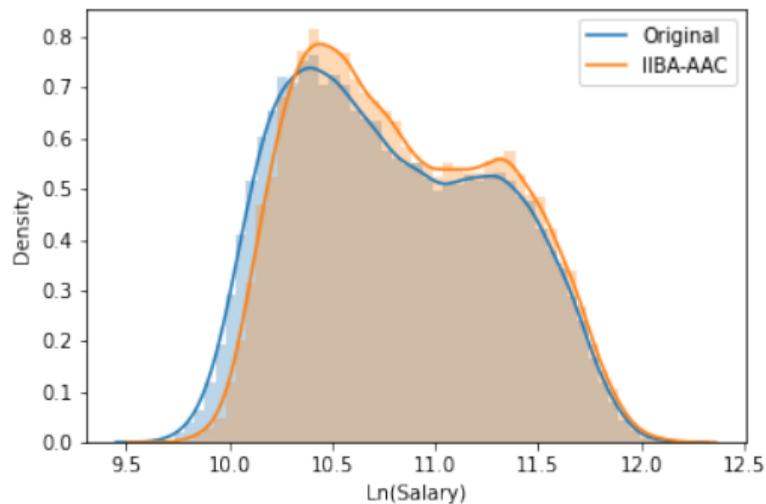
All Postings



0.060 log point increase \approx \$3013

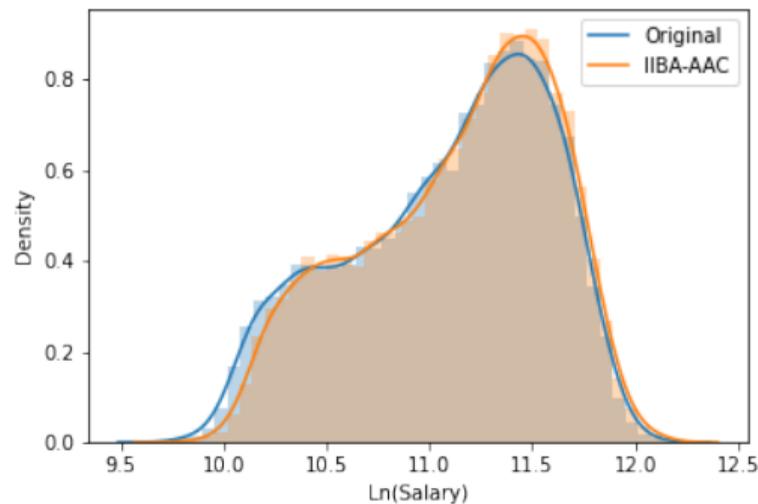
Example: International Institute of Business Analysis – Agile Analysis Certification (IIBA-AAC)

All Postings



0.060 log point increase \approx \$3013

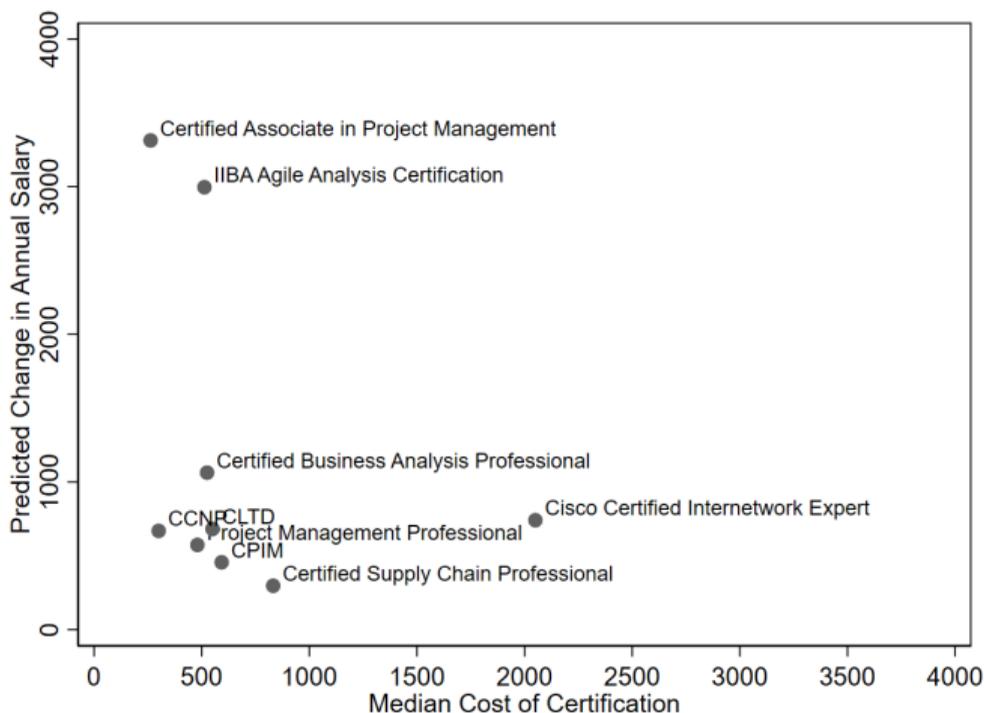
Only Postings in Certain Occupations



0.047 log point increase \approx \$3140

Applied to In-Demand Certifications

- All of these certifications are associated with a positive effect on earnings
- Not systematically related to type of certification or cost
- Substantial range of premia



Conclusion

The Text of the Posting Matters!

And We Can Use It!

- Supervised learning model using metadata of salaries explains significantly more variation than a model with occupation and location FEs
- Countless possibilities to evaluate characteristics of jobs
 - Close to real-time
 - Differences by place and occupation
- As firms and workers make strategic decisions about their human capital, this information is a crucial input

Wide Variety of Applications

In Reskilling and Other Domains

- Currently, advice around upskilling is quite broad
 - Skills may be valued differently in different places
 - Marginal skill could be different in different roles
 - Advice should change over time, based on labor market conditions

A new paradigm, underpinned by better labor market information, could substantially improve America's approach to workforce education and training ([Bonvillian and Sarma, 2021](#))

Thank you!
Any feedback is appreciated

sarah.bana@gmail.com
@SarahHBana

Formal Treatment of Text Injection Method

Trained model can be described as

$$Y = f(X|\beta)$$

- Y is the outcome, in this case the salary
- X is the posting text
- β are the learned parameter vector of weights derived from the BERT layer and training from the process

Recall, β is high dimensional and contains many interaction terms, differentiating it from counting words

For a given posting i , we can add text t

$$y_i = f(x_i|\beta)$$

Posting without added text

$$y_{i,0} = f(x_i, t_i = 0|\beta)$$

Posting with added text

$$y_{i,t} = f(x_i, t_i = t|\beta)$$

Central Limit Theorem (CLT) for Consistency and Inference

The outcome of interest is the average value of t on salary. This amounts to an expectation:

$$\mathbb{E}[f(x_i, t_i = t|\beta) - f(x_i, t_i = 0|\beta)]$$

- Randomly sample from postings
- Treat postings as independent and identically distributed (i.i.d)
- Draw on the Central Limit Theorem (CLT) for consistency and inference

▶ [Back to Text Injection Intuition](#)