# Estimating Local Wage Growth from Glassdoor Salary Data

Andrew Chamberlain, Ph.D.
Chief Economist, Glassdoor

# The Problem

- What is the median wage for **data scientists** in **Seattle** in **June 2017**?

- This is **really hard** to know:

  - Local wage statistics from BLS are only released **once a year.**

  - BLS occupations **aren't detailed enough** (no "data scientist" SOC category).

# Our Idea

- Glassdoor is a job site. We ask users to leave **anonymous salaries** while browsing jobs.
- We've collected **several million** of these salaries at the **job title, company** and **metro** level (for 700,000+ employers) since 2008.
  - **Huge scale**: We have **34,000+** Walmart salaries.
- Lets use these data to estimate local wage trends.

# The Result: Local Pay Reports

Glassdoor Economic Research

**Metro Area**

- United States
- Atlanta
- Boston
- Chicago
- Houston
- Los Angeles
- New York City
- Philadelphia
- San Francisco
- Seattle
- Washington D.C.

## United States | April 2017

**Methodology:** Based on millions of anonymous salary reports shared on Glassdoor. The Local Pay Reports apply a proprietary machine-learning algorithm to estimate median annual base pay for full-time workers. April 2017 figures are current as of April 25, 2017. Metro areas are based on U.S. Census Bureau "core based statistical areas" (CBSAs). All year-over-year estimates for median salary growth are based on 3-month moving averages. Median base pay is presented in nominal dollars, and is not adjusted for inflation or seasonally adjusted. (*) denotes Bureau of Labor Statistics data.

A list of frequently asked questions is available here. Complete methodology is available here.

### Job Title

Design

| | Median Base Pay | YoY |
|---|---|---|
| Graphic Designer | $ 43,550 | 2.1% |

Education

| | Median Base Pay | YoY |
|---|---|---|
| Professor | $ 90,007 | 5.8% |
| Teacher | $ 46,017 | 2.5% |

Engineering

| | Median Base Pay | YoY |
|---|---|---|
| Mechanical Engineer | $ 73,349 | 2.4% |
| Design Engineer | $ 70,616 | -0.3% |
| Electrical Engineer | $ 77,098 | 2.5% |

### U.S. Quick Facts

| | April 2017 | YoY |
|---|---|---|
| U.S. Median Pay | $ 51,350 | 2.7% |

### Industry

| | Median Base Pay |
|---|---|
| Accounting & Legal | $ 52,251 |
| Aerospace & Defense | $ 54,684 |
| Architecture & Civil Engineering | $ 52,794 |
| Arts & Entertainment | $ 46,495 |
| Automotive | $ 46,911 |
| Banking & Financial Services | $ 51,076 |
| Beauty & Fitness | $ 45,562 |

# How Does it Work?

# Main Challenge We Faced

- Glassdoor salaries exhibit **composition bias**.

- The mix of salaries reported on Glassdoor over time differs from overall labor market.

    - **Geography**: If early adopters are from high wage areas, **average wages fall** as content gets more representative.

    - **Firms**: A surge in salaries from high-wage employers can cause **average wages to rise** due to composition effects.

- We need **condition on observables** to correct for this, and hold constant the composition of our sample over time.

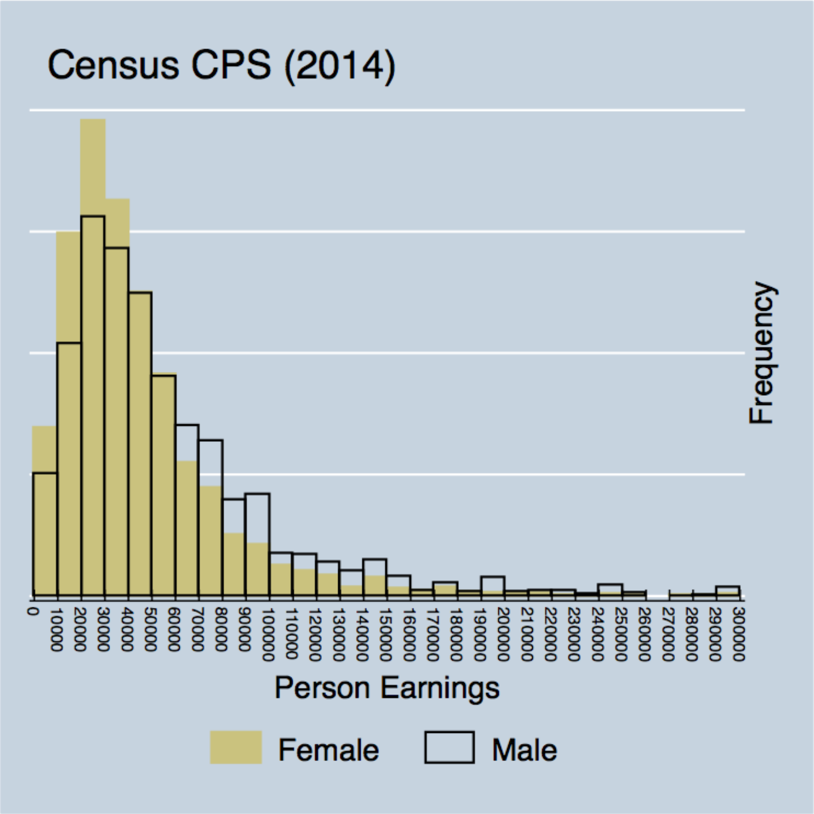
# Our Solution

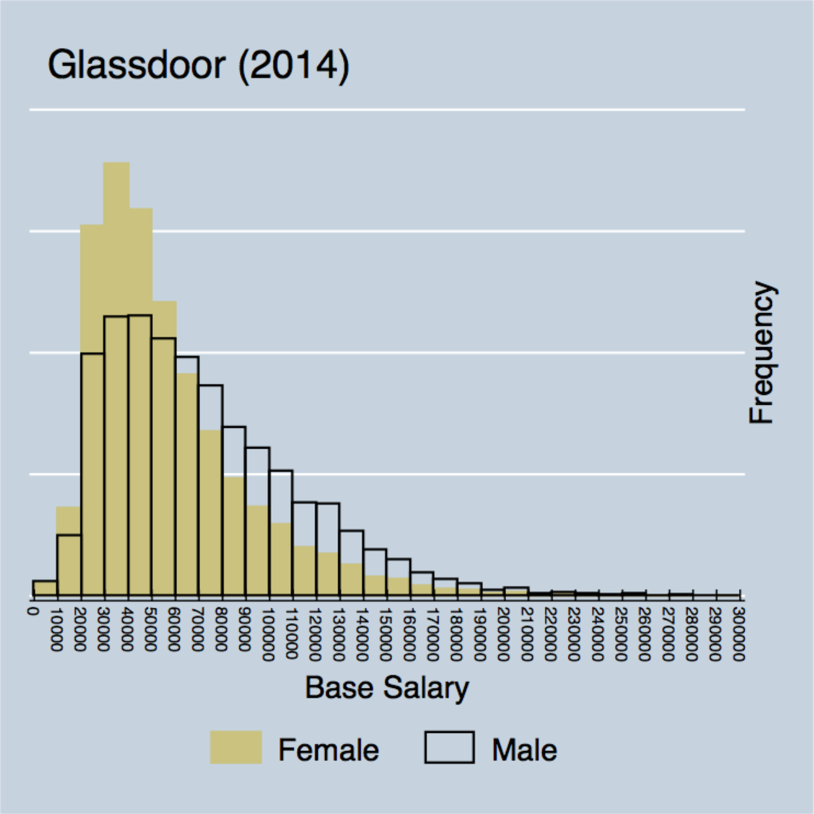
- We estimate the separate impact of **many salary observables** using a simple ML model (analogous to fixed effects).

- **We control for features** like job title, seniority, company, industry, employer size, metro, time period, and **many** more.

- We'll then reassemble these "building blocks" of salaries into **conditional local wage estimates** at the job title and metro level, at different points in time.

# First, Some Perspective




Bias is **less severe** than most economists assume.



glassdoor®

# How It Works

- We use a penalized machine learning model known as “elastic net.”

- It’s a generalized linear model (GLM) that’s a compromise between LASSO and ridge regression.

- Computationally, it solves:

$$\hat{\beta} = \underset{\beta_0,\beta}{\operatorname{argmin}} \frac{1}{2N} \sum_{i=1}^{N} (\underbrace{y_i - \beta_0 - x_i^T \beta}_{\text{OLS}})^2 + \lambda \left[ \underbrace{\frac{(1-\alpha)}{2} \|\beta\|_2^2}_{\text{“ridge” penalty}} + \underbrace{\alpha \|\beta\|_1}_{\text{LASSO penalty}} \right]$$



glassdoor®

# Why Penalized Regression

- We have **lots** of predictors, with many highly correlated.
- The **penalty term**: (1) omits less important and redundant predictors (like LASSO), and (2) shrinks correlated predictors toward each other (like ridge).
- It is **stable and computationally efficient** (runs pretty fast at scale).
- **The proof of the pudding is in the eating**. It delivers reliable, low-variance estimates month after month.

# Doing This in Practice

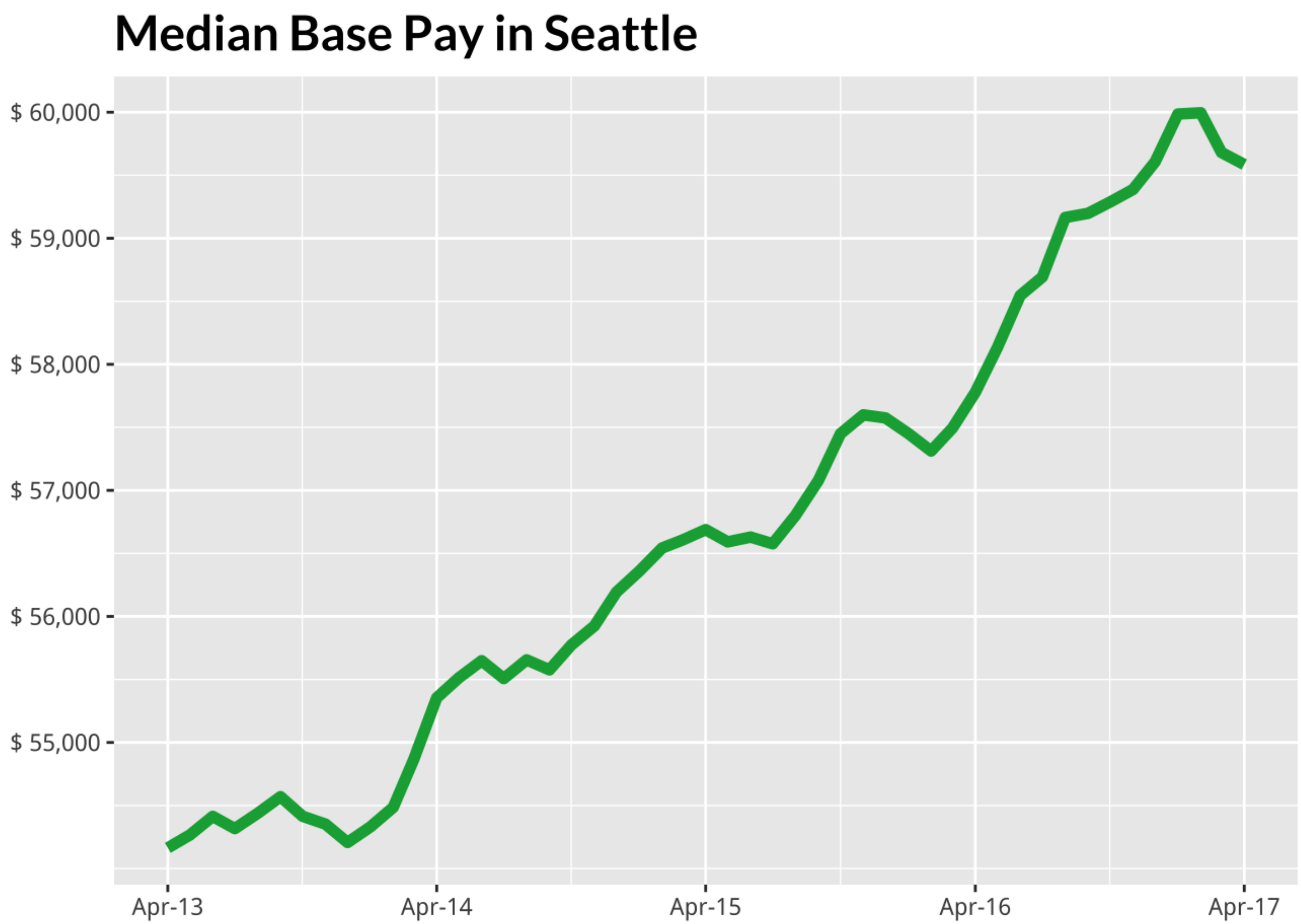
- We pull a large SQL query of **several million** U.S. salaries for all time.
  - **Base pay**, **full-time** workers (future versions will include bonuses, commissions and total comp.).
- We estimate **elastic net regression** of **log salary** on a large number of features in *R.*
- We run a Python script to re-assemble estimated beta coefficients into pay estimates at job title /metro level.

# An Example

- **Software Engineer** in **Seattle** in **June 2016**

| Feature | Description | Value | Coefficient |
|---|---|---|---|
| G | Job Title | Software Engineer | 0.476 |
| L | Metro | Seattle | 0.12 |
| GL | Job Title x Metro | Software Engineer x Seattle | 0.08 |
| P | Pay Period | Annual | 0.118 |
| GP | Job Title x Pay Period | Software Engineer x Annual | 0.002 |
| Y | Time Period | June 2016 | 0.07 |
| GY | Job Title x Time Period | Software Engineer x June 2016 | 0.036 |
| LY | Metro x Time Period | Seattle x June 2016 | 0.021 |
| Int | Intercept | N.A. | 10.65 |
| Sum | | | 11.573 |
| Base Pay Estimate (June 2016) | | | $106,192 |

# Overall Metro Trends






glassdoor®

# How Does it Perform?

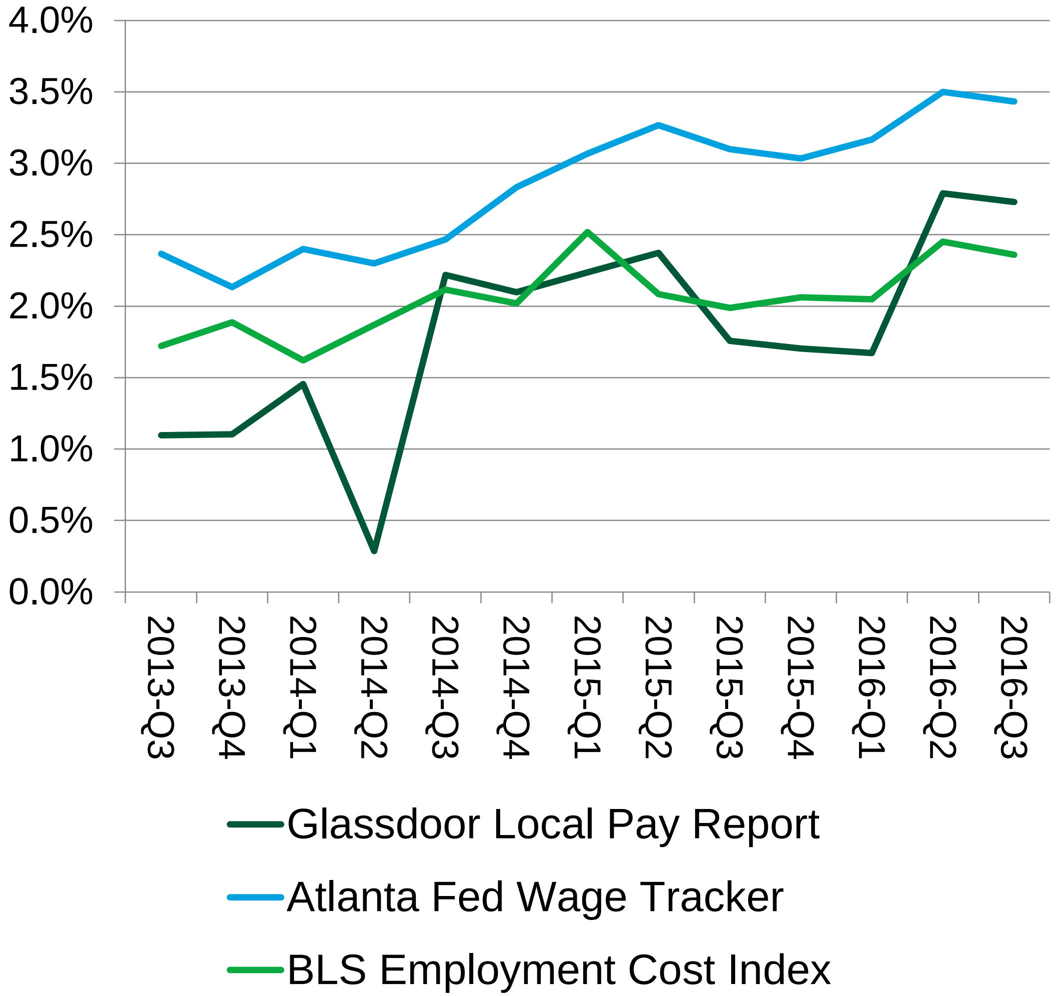
# Pretty Close on Overall Pay Levels

Table 2. Comparison of Median U.S. Pay from BLS Estimates and Glassdoor's "Local Pay Reports" Model

| Year | Glassdoor Local Pay Report Estimate (U.S. Median Base Pay, Full Time Workers) | BLS Estimate (U.S. Median Wages and Salaries, Full Time Workers) | Percentage Error |
|---|---|---|---|
| 2015 | $44,282 | $42,068 | 5.3% |
| 2014 | $43,808 | $41,132 | 6.5% |

*Source: U.S. Bureau of Labor Statistics, "Labor Force Statistics from the Current Population Survey: Table 39. Median weekly earnings of full-time wage and salary workers by detailed occupation and sex." Available at http://www.bls.gov/cps/cpsaat39.htm.*

# Pretty Close on Wage Growth




| | **Correlation** |
|---|---|
| Atlanta Fed Wage Growth Tracker | **0.79** |
| BLS Employment Cost Index | **0.75** |



glassdoor®

# Some Interesting Pay Trends

- Median pay rising 3.8% YOY in **Los Angeles**, but only 1.0% in **Houston** (low energy prices).
- Median pay for **data scientists** is leveling off (1.1% YOY) as technical skill profile for that role flattens out.
- **Registered Nurse** pay is rising 4% YOY, well above U.S. average.
- **Retail cashier** pay up 5% YOY, reflecting minimum wage hikes and 500K+ job openings.

THE WALL STREET JOURNAL.

REAL TIME ECONOMICS | EMPLOYMENT

# How Big Data Is Creating a Detailed Picture of U.S. Earnings

A new monthly report from Glassdoor uses millions of employees' reviews of their employers to estimate U.S. wage growth

By JOSH ZUMBRUN

Nov 4, 2016 6:16 am ET

For a typical worker negotiating salary, it's helpful to know how earnings of people in your occupation, and in your city, are changing. Turn to the Labor Department's monthly employment report, however, and you'll mostly find

# For More Information

- Glassdoor's **Local Pay Reports** are available **free** at:

glassdoor.com/research

# Thank You

Andrew Chamberlain, Ph.D.

**Web:** Glassdoor.com/research

**Email:**

andrew.chamberlain@glassdoor.com