TEXAS Labor Market Information

Better Metrics Using Wage Record Microdata

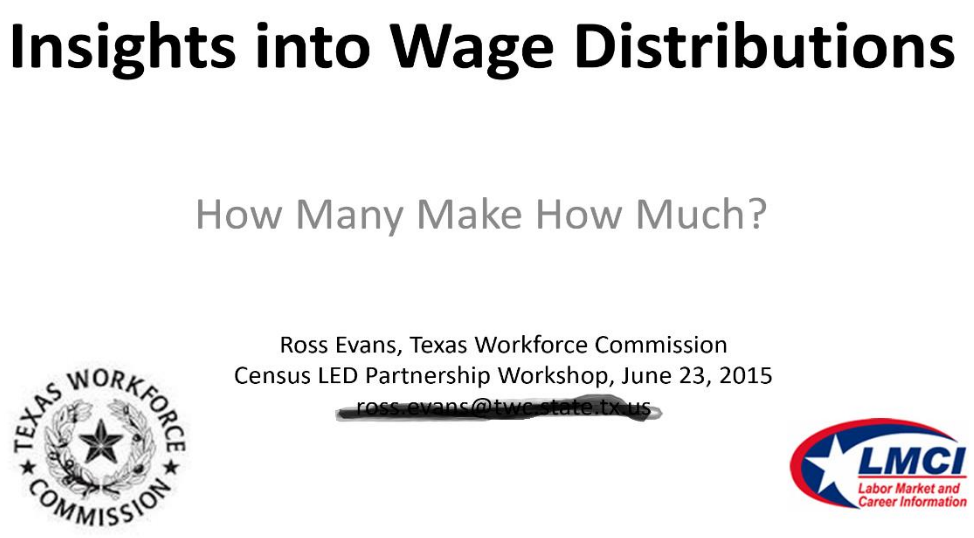
Analyzing the Distribution of Texas Wages

Census LED Partnership Workshop — May 15, 2024

Background

- This presentation follows up another from this workshop in 2015, which showed:
    - Wages naturally scale exponentially, not linearly.
    - Log-transforming the wage data produces a bell-shaped curve for the distribution.
    - The center mass of that curve and geometric mean fall significantly below the published QWI average wage statistic.
- Today we will try to show what a better metric could be.

2015 presentation

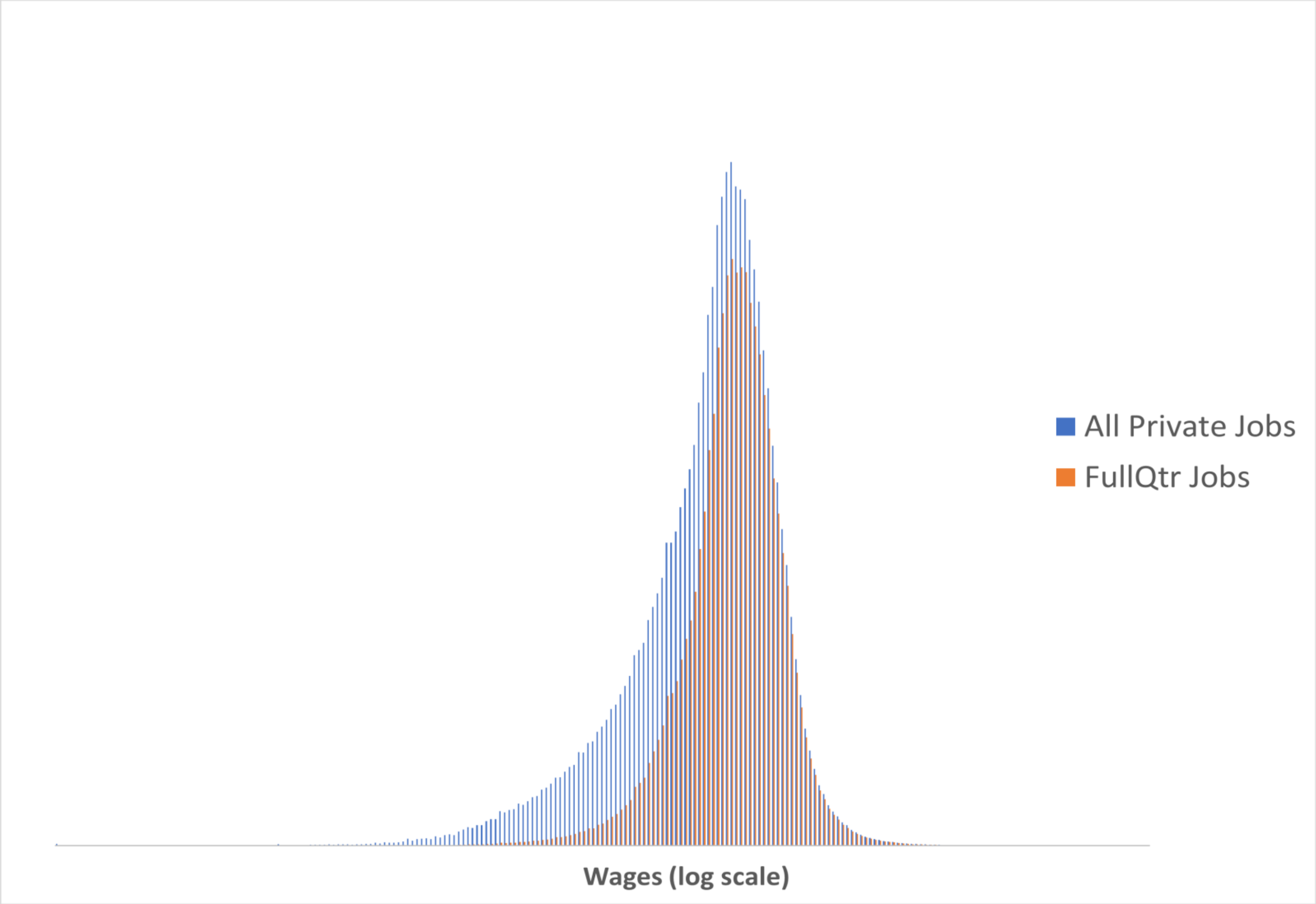


Insights into Wage Distributions

How Many Make How Much?

Ross Evans, Texas Workforce Commission
Census LED Partnership Workshop, June 23, 2015
[redacted]

A general primer on wage distributions …



- Wages scale exponentially. Log transformation yields a bell-shaped curve.
- Wage values in this generic chart grow by 10 percent per interval. Values range from 1 cent to 10's or 100's of millions of dollars per job.
- The universe skews to the left, mostly due to partial-quarter employment — an artifact of clunky data collection.
- Culling those jobs leaves a more symmetrical Full Quarter subset.

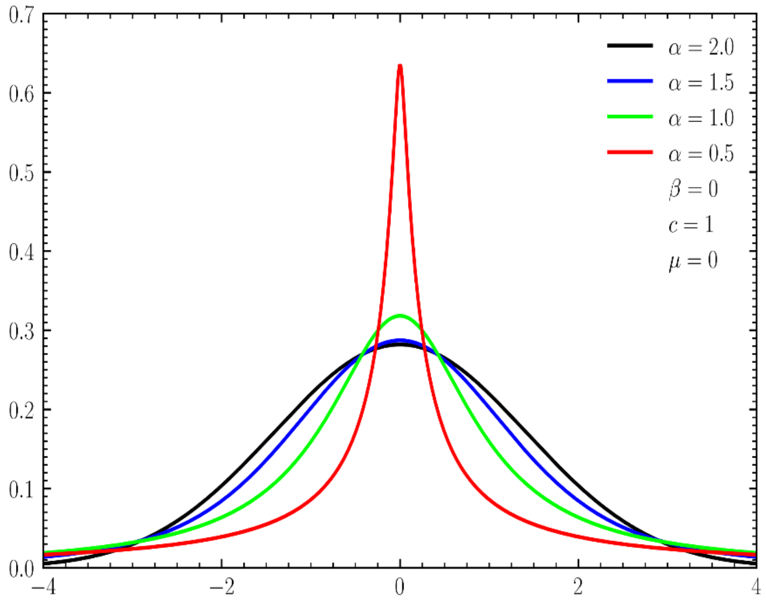
New math, new meaning for 'Stable'

- This is **not** the same use of the word "Stable" as Census QWI uses – meaning the subset of Full Quarter jobs.
- Here we reserve the use of that word for what mathematicians call a general branch of statistics: Stable Distributions.

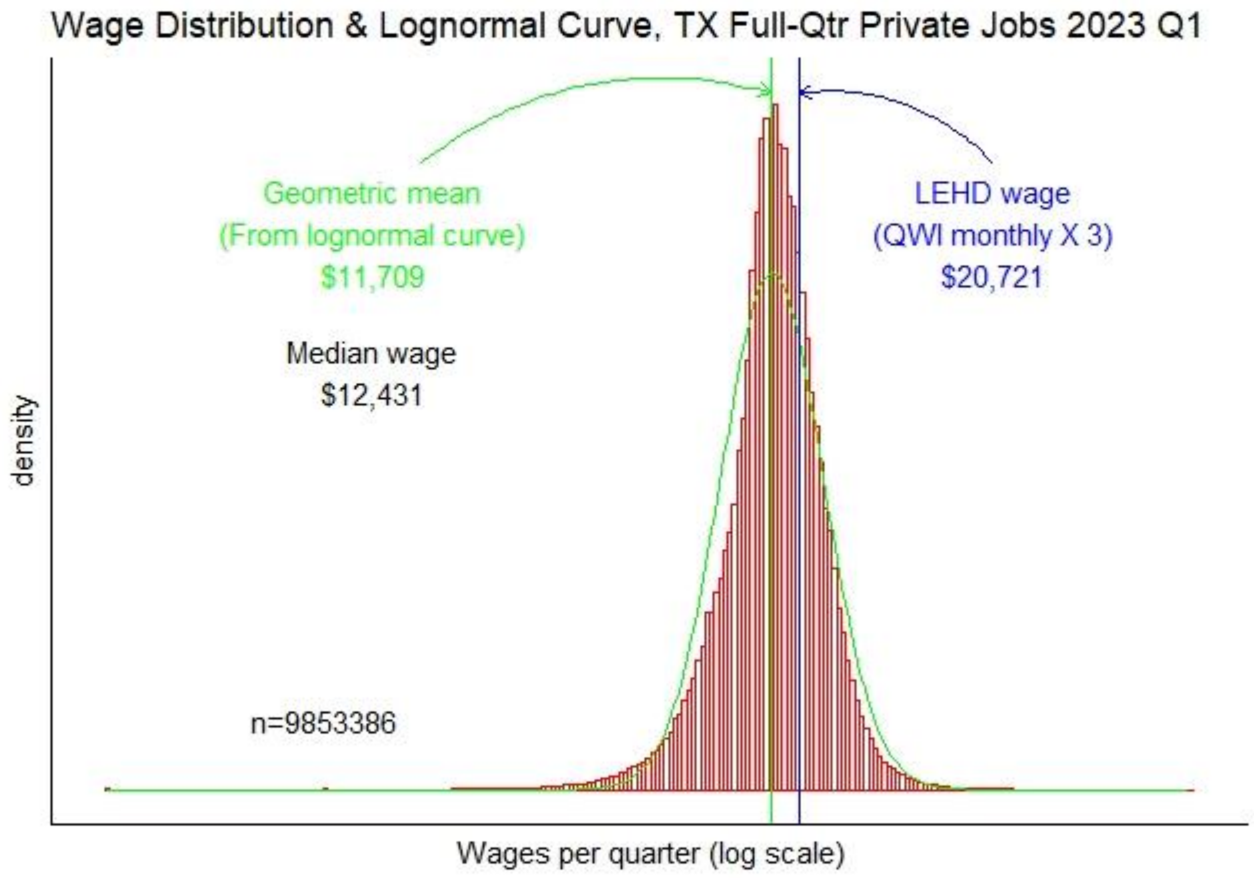
The Stable equation is a better fit than the commonly used Normal curve. It accounts for the variable shape and skew in the data. (Normal is just a special case within the Stable family.)

Because we log-transform the empirical data first, we will examine the fit of Lognormal and Logstable models.

These are density plots, which means that the area under the curves remains constant. They always contain exactly 100 percent of all jobs in the sample.

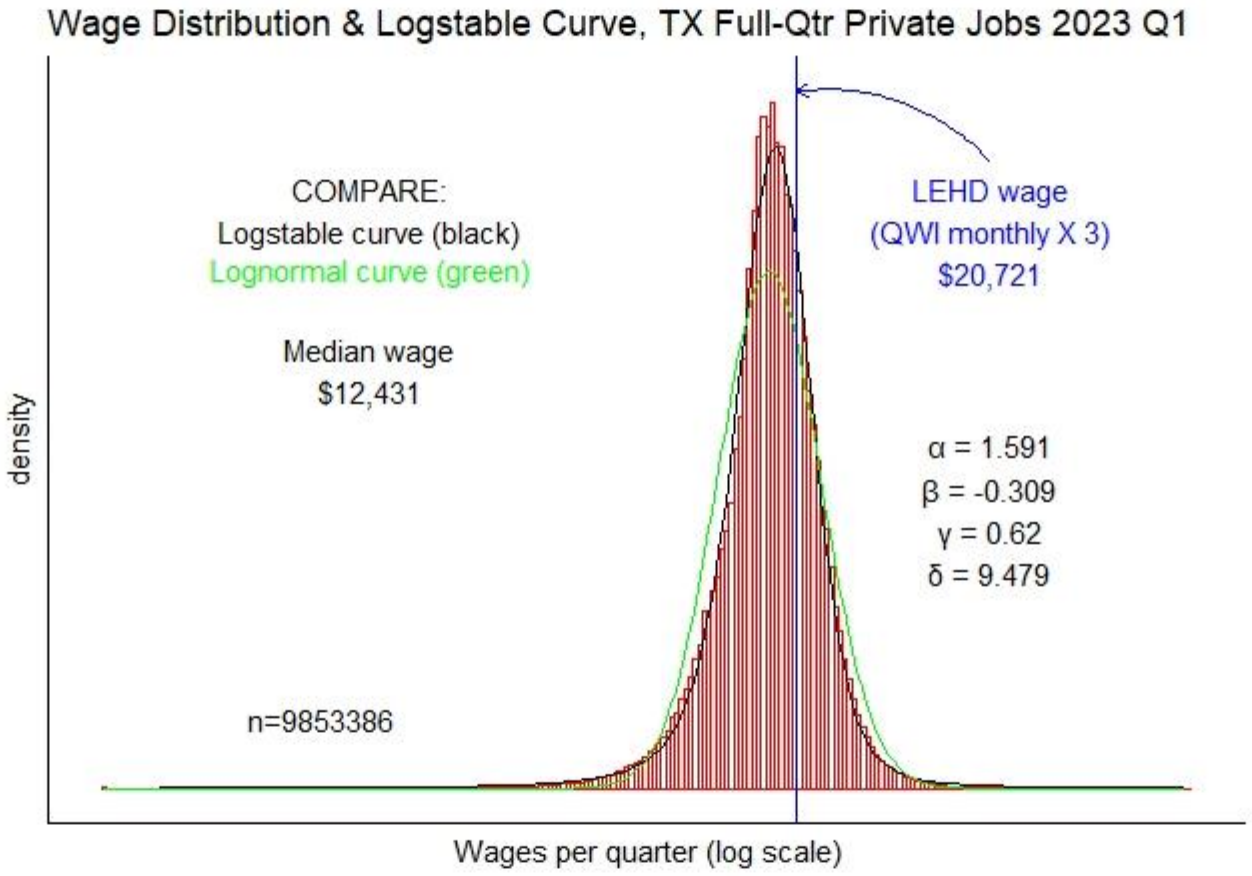


Distribution vs. QWI average wage



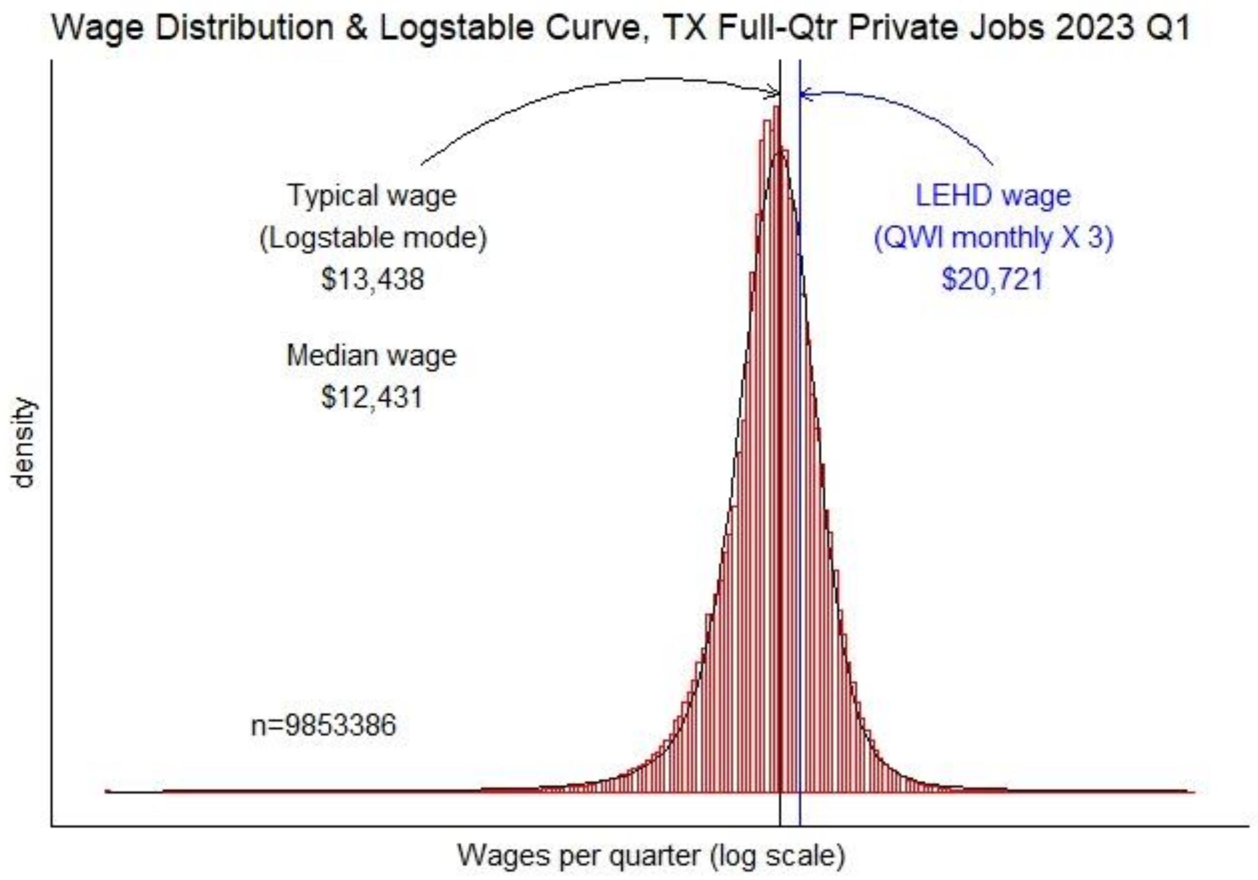
- For Full Quarter private jobs, compare the log-transformed wage distribution to the QWI average.
- Empirical data binned in histogram at 10% wage growth per interval.
- As a first-order approximation, the geometric mean can be calculated. The Lognormal curve is plotted around it using the derived mean and standard deviation of all log-transformed data points.

Today we have a curve that fits better



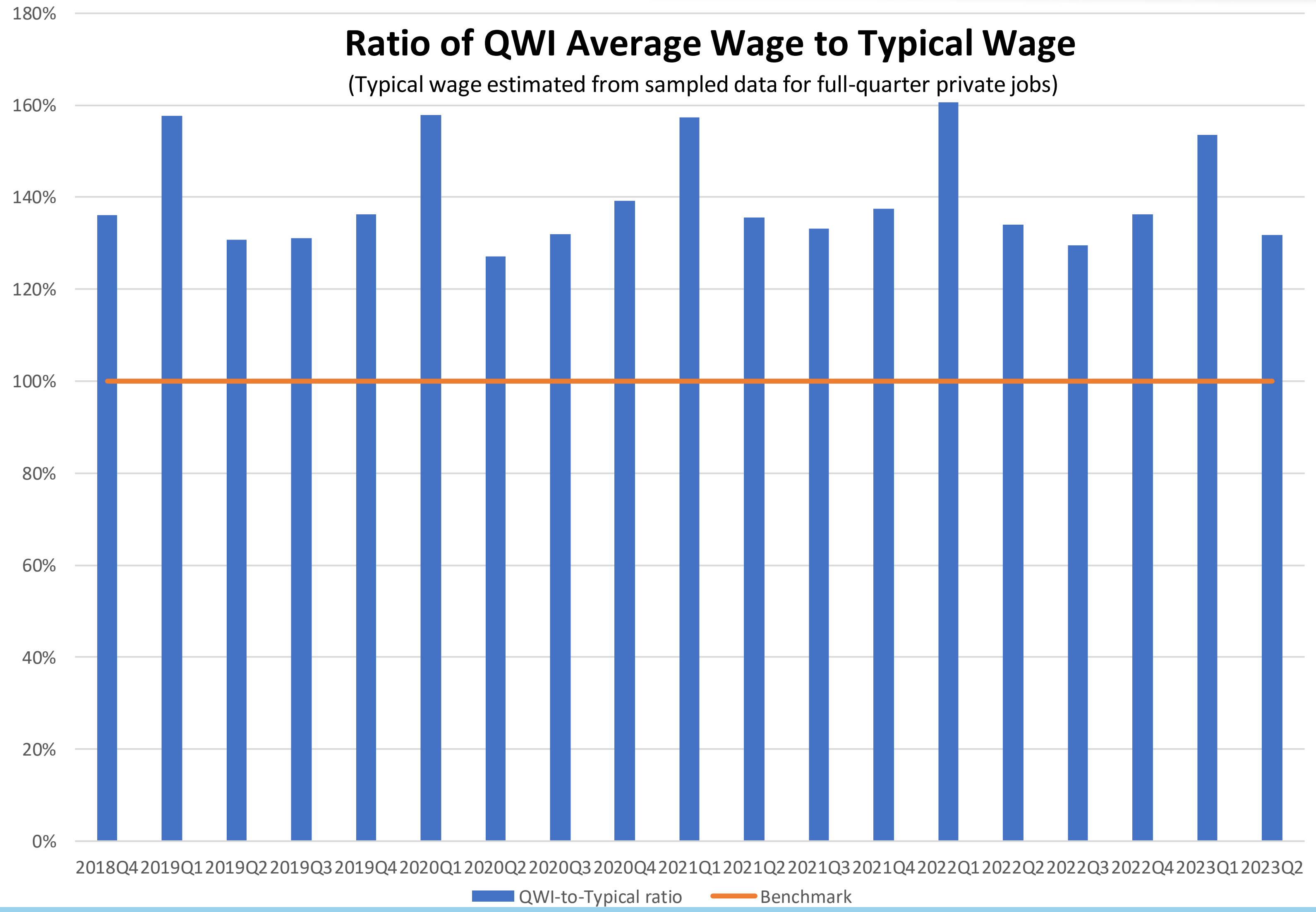
- Using Stable software, estimate the 4 parameters that define this curve from log-transformed data:
  - Alpha – shape
  - Beta – skew
  - Gamma – similar to standard deviation
  - Delta – similar to mean
- That estimation step is compute-intensive for large datasets!
- Using those parameters, plot the Logstable curve.
- This curve is a much better fit than Lognormal:
  - Closer to the higher peak
  - Closer to the necklines and shoulders
  - Heavier tails. Model fits when the cumulative distribution quantile >.01 and <.99, per ecdfHT plot (not shown).

Mode of that curve is the ‘Typical Wage’



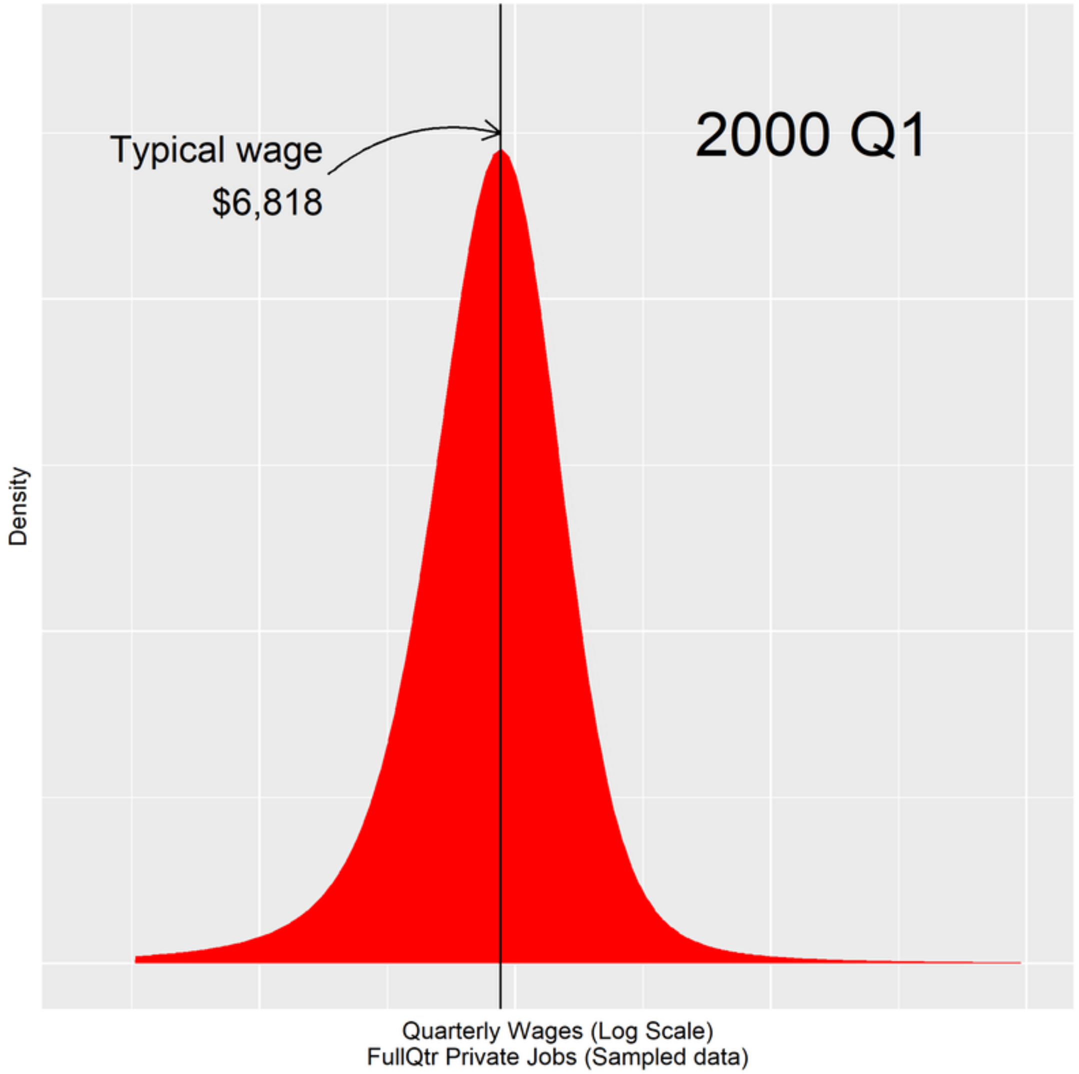
- The mode is a good representation of rank-and-file jobs’ wages. And the zenith of this continuous curve is the optimal estimate of the mode. This is **the Goldilocks metric**.
- The distribution shows a real-world phenomenon of the labor market, which organizes itself organically around that peak.
- Remember the Peter Principle!
- By any of the measures we looked at, the QWI “average” wage is far above what typical jobs pay. Here it falls near the **72nd percentile**.
- In dollar terms **QWI is 54 percent higher**.

Recent QWI vs. ‘Typical’ benchmark



- Using the Typical wage metric as a benchmark, QWI has exceeded it for most recent 19 quarters.
- The ratio above the benchmark has been at least 27 percent and as much as 60 percent.
- There is a seasonal pattern, with the highest ratios occurring in Q1 of each year – driven by annual bonuses.
- Not shown here: The ratio varies even more in subset samples of different industries.

Shape and Typical Wage change over time



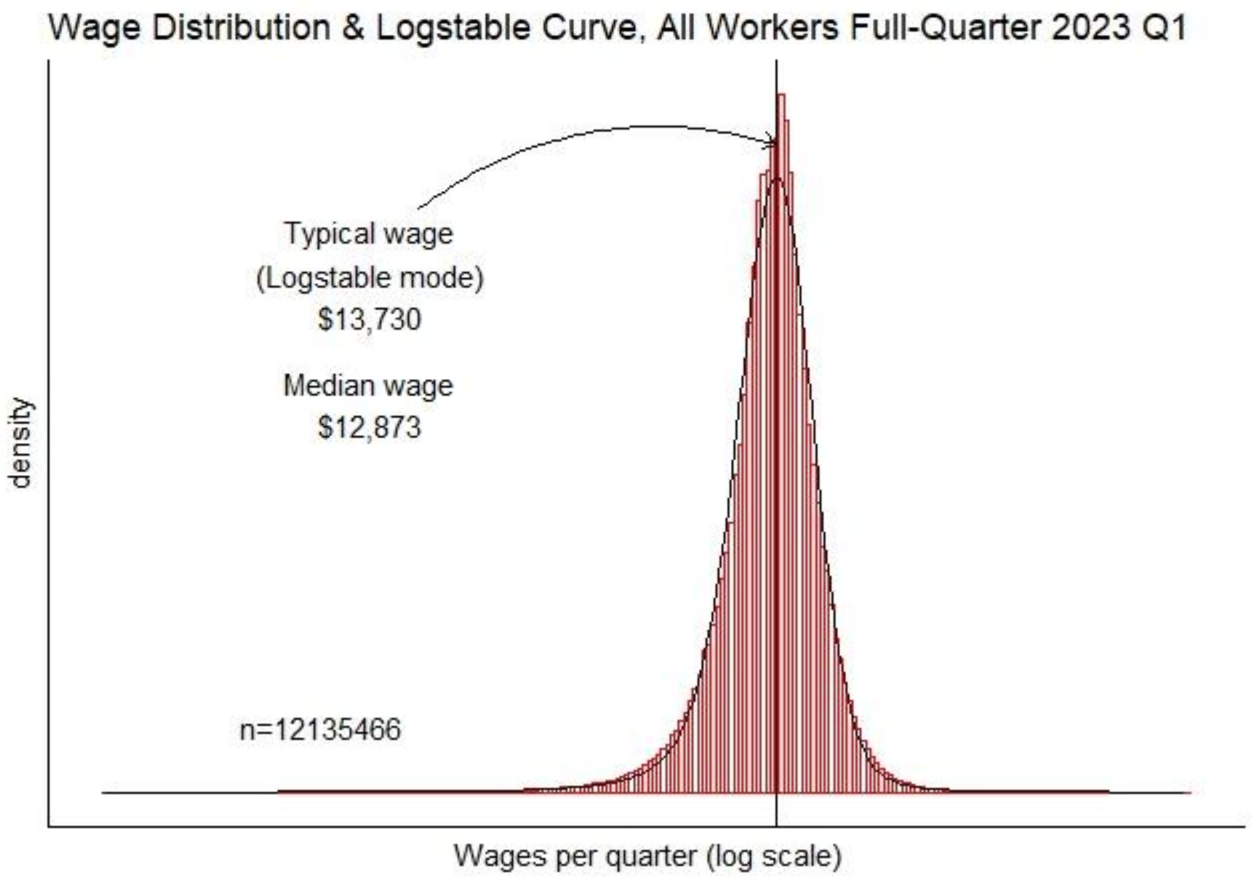
- The labor market is dynamic. So the shape and position of the wage distribution are constantly changing.
- This reflects seasonal, cyclical and secular changes in the economy.
- While the shape is fluid, the overall distribution migrates from left to right with wage inflation.
- Typical wage grew from:
  - 2000 Q1 $6,818
  - 2023 Q1 $13,500

  - 2000 Q2 $7,125
  - 2023 Q2 $13,841

(Due to sampling error, 2023 Q1 wage differs from earlier slide)

A different view per worker, not per job

- As an additional view of the microdata, it can be re-aggregated per SSN across all jobs — here including private, state/local government.
- This totals all wages each worker makes, including moonlighting, multiple part-time jobs and job changes within each quarter.
- The only records excluded from this Full Quarter subset are those who are not employed in Texas at all in the prior and succeeding quarters.
- This view would be even more useful if it included data from other states.
- The per-worker view is more comparable to some other datasets, such as income in the survey-based CPS.



Humble suggestions and caveats

By this analysis, the QWI average wage is more representative of high-earning employees.

The Typical wage from the Logstable model would be a better fit for most rank-and-file jobs.

- QWI aligns with the QCEW "Average Weekly Wage" from BLS. But that is because **QCEW and QWI share the same flaw:** They both estimate the arithmetic mean of a variable (wages) that does not scale arithmetically!
- The Logstable model is compute-intensive, and software options are limited. There are tradeoffs between resources and accuracy. So it **may be hard to implement in a production workflow**.
- The **median wage is more easily computable**, although the **value of the median is even lower than the Typical wage**. The median of the raw universe of wage records is pulled artificially low by partial-quarter employment, but is reasonable for the Full Quarter Jobs subset.

Citations

Univariate Stable Distributions
Models for Heavy Tailed Data
John P. Nolan in *Springer Series in Operations Research and Financial Engineering* *(2020)*

Royuela-del-Val, Javier & Simmross-Wattenberg, Federico & Alberola-López, Carlos, 2017. "**libstable: Fast, Parallel, and High-Precision Computation of α-Stable Distributions in R, C/C++, and MATLAB**," Journal of Statistical Software, Foundation for Open Access Statistics, vol. 78(i01).

Received: 29 November 2019 Revised: 18 October 2020 Accepted: 20 October 2020. Wiley. DOI: 10.1002/asmb.2590 SPECIAL ISSUE PAPER A graphical diagnostic for heavy tailed data John P. Nolan

Ross Evans

ross.evans@twc.texas.gov