Power Transformations for Consumer Expenditure Data

Taylor J. Wilson
Branch of Information and Analysis
2018 Microdata Users Workshop
18-20 July 2018

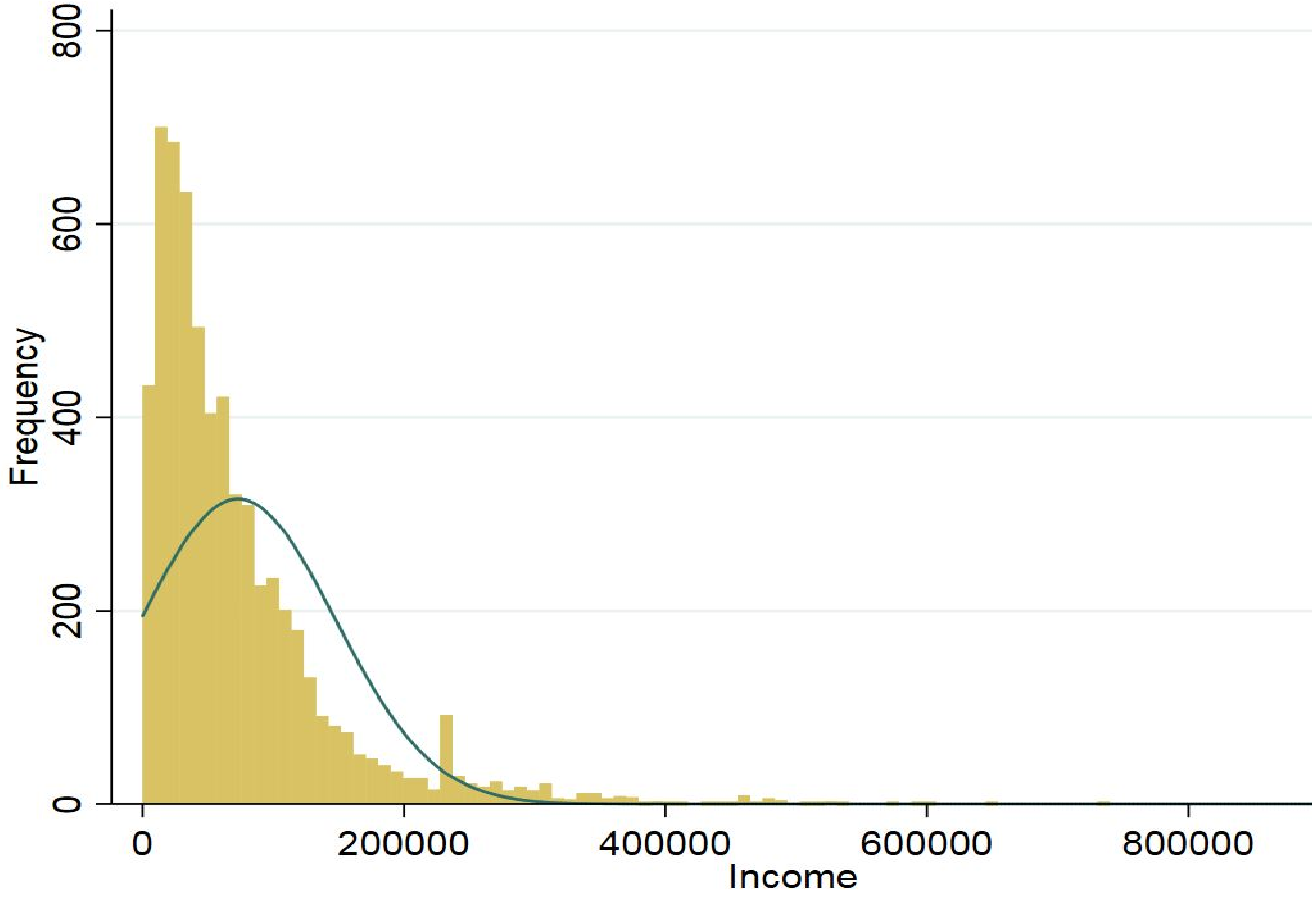
Roadmap

- CE data distributions
- Defining the power transformation
- Picking the optimal lambda
- Income-Expenditure relationship

Data Distributions

- Most expenditure and income data exhibit a right skew.
- This results from a fixed lower bound and extreme values in the right tail.
- May need to correct the distribution into a normal shape.

Income Distribution – First Quarter 2017



Source: Public Use Microdata 2016 – First quarter 2017

Defining the Power Transformation

- Power Transformations can be expressed as a simple power in a linear context.
- We care about finding a λ that works for our purposes.
- Log transformation is a special case of power transforms.

$$y_{trans} = \begin{cases} \dfrac{y^{\lambda} - 1}{\lambda} & (\lambda \neq 0) \\ \ln(y) & (\lambda = 0) \end{cases}$$

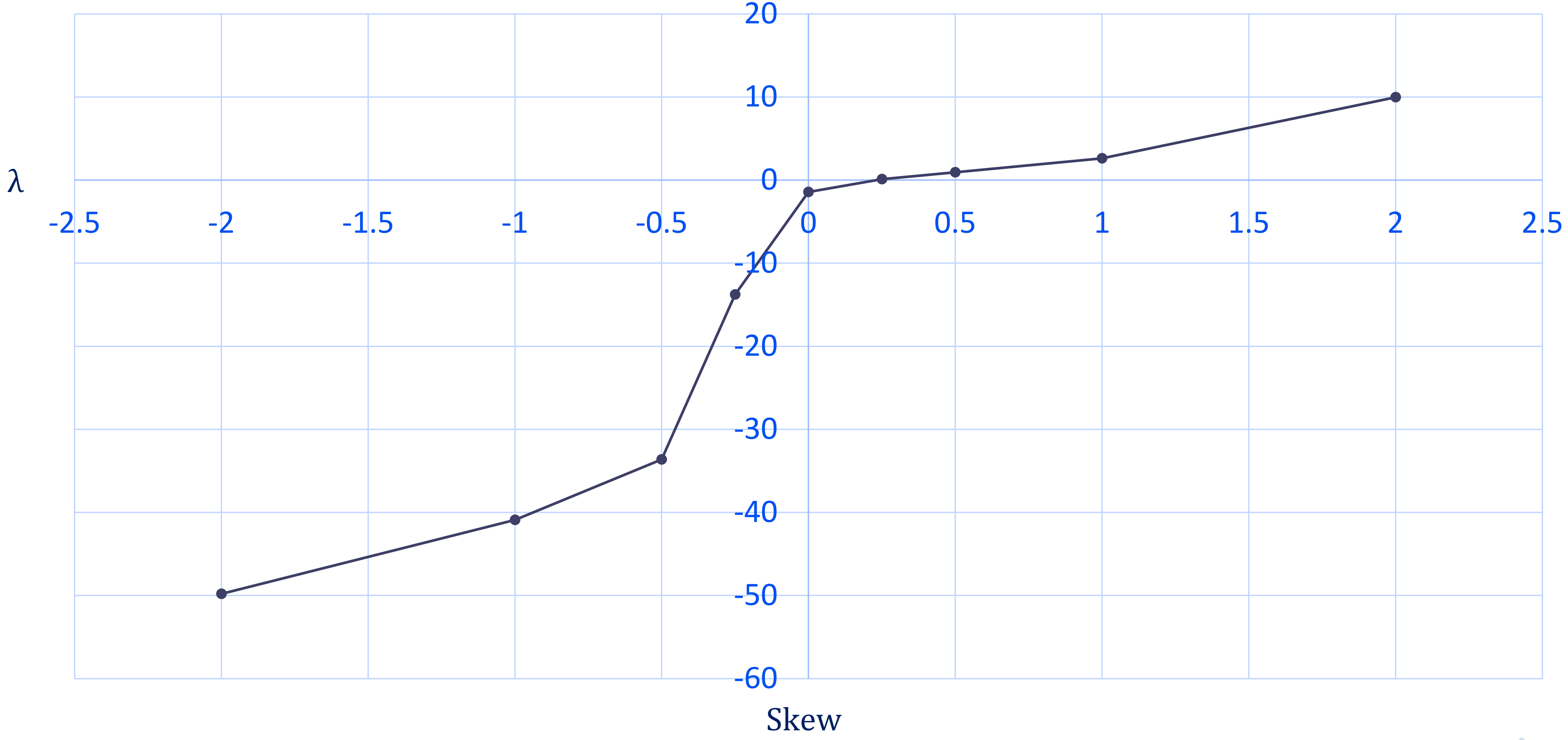
Picking the Optimal Lambda

- Choose lambda such that skew is minimized.

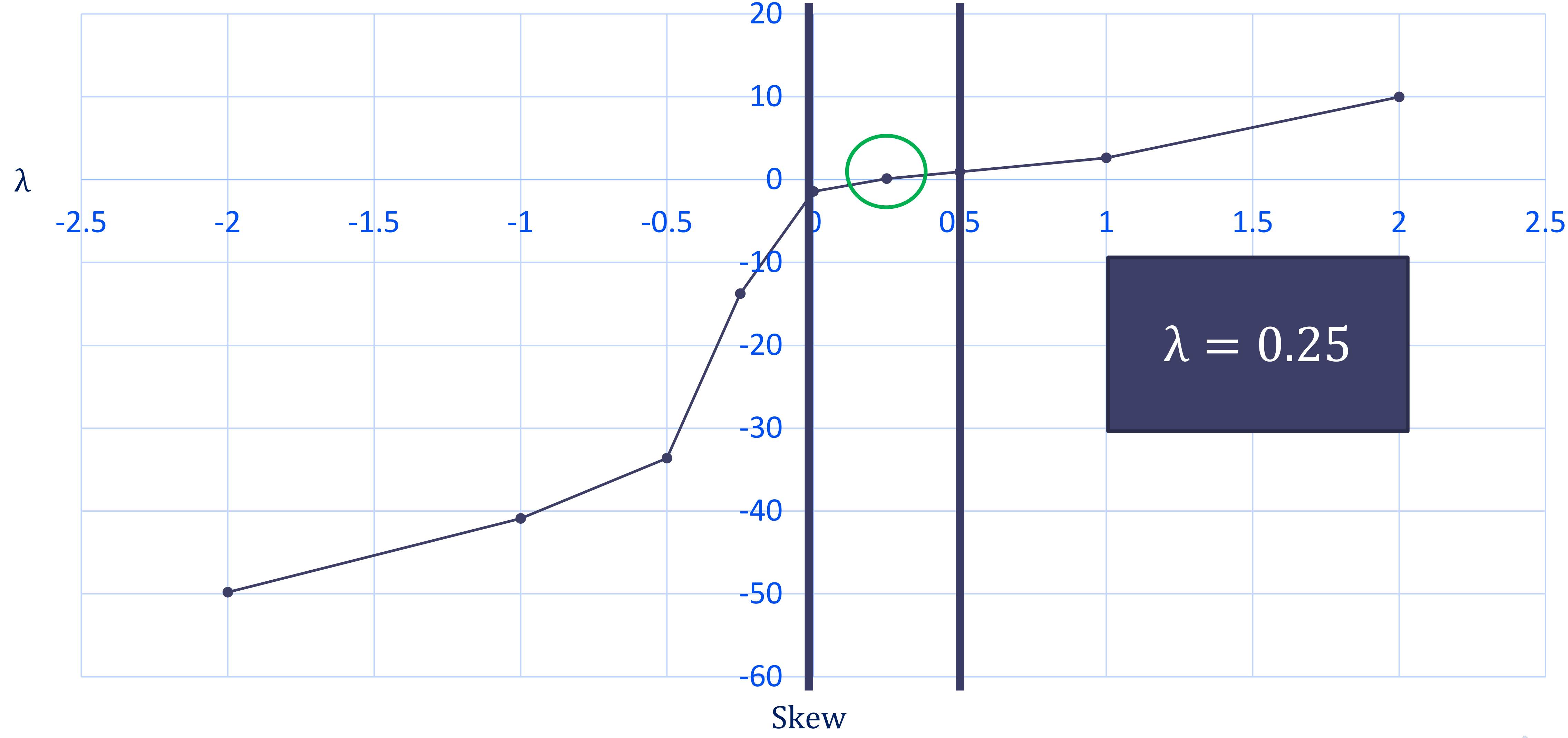
- Statistical programs estimate this with maximum likelihood estimation.

- Visualize skew by plotting it against the parameter. For convenience select parameters with function names.

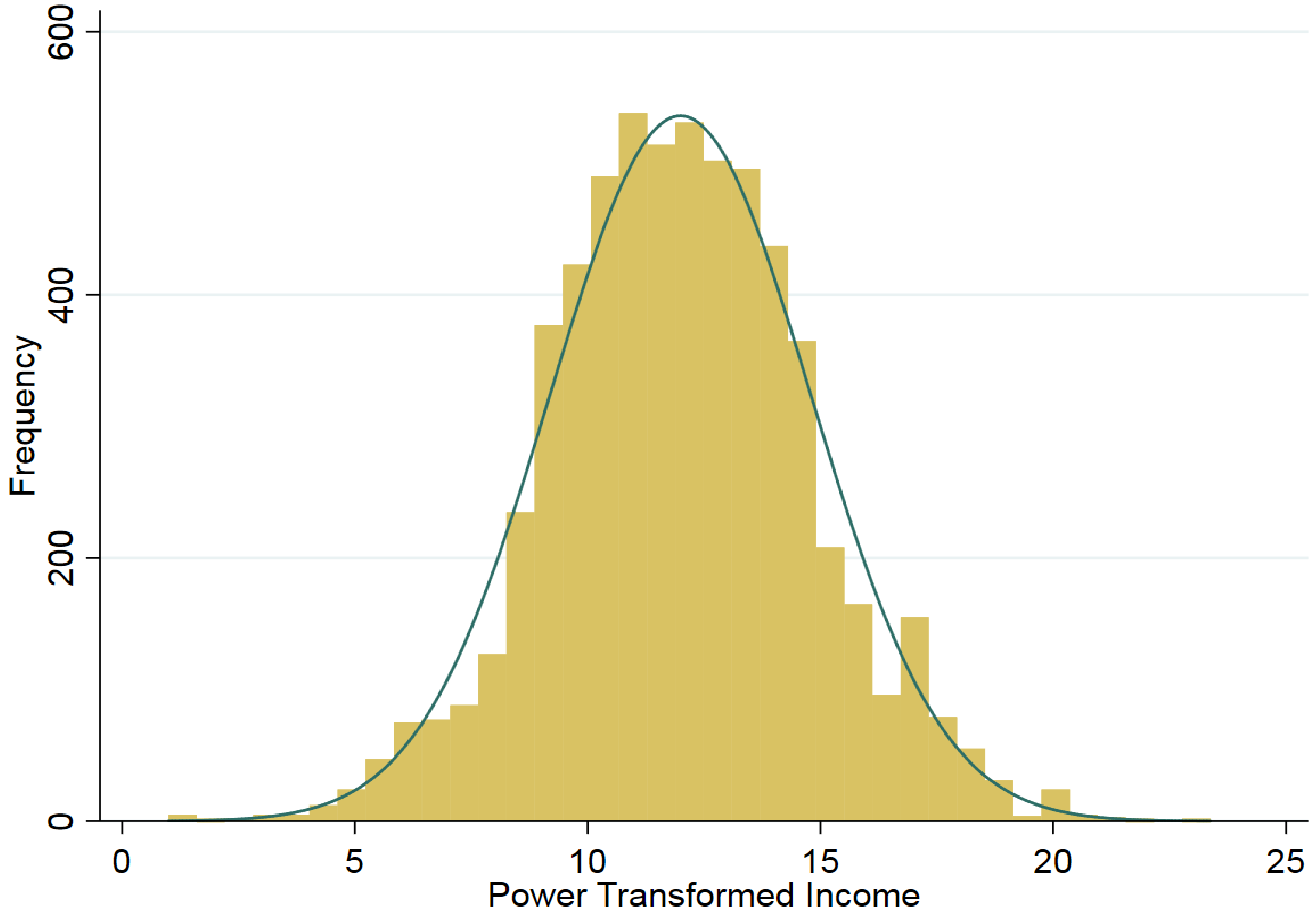
Skew by Selected Lambda Parameters



Skew by Selected Lambda Parameters

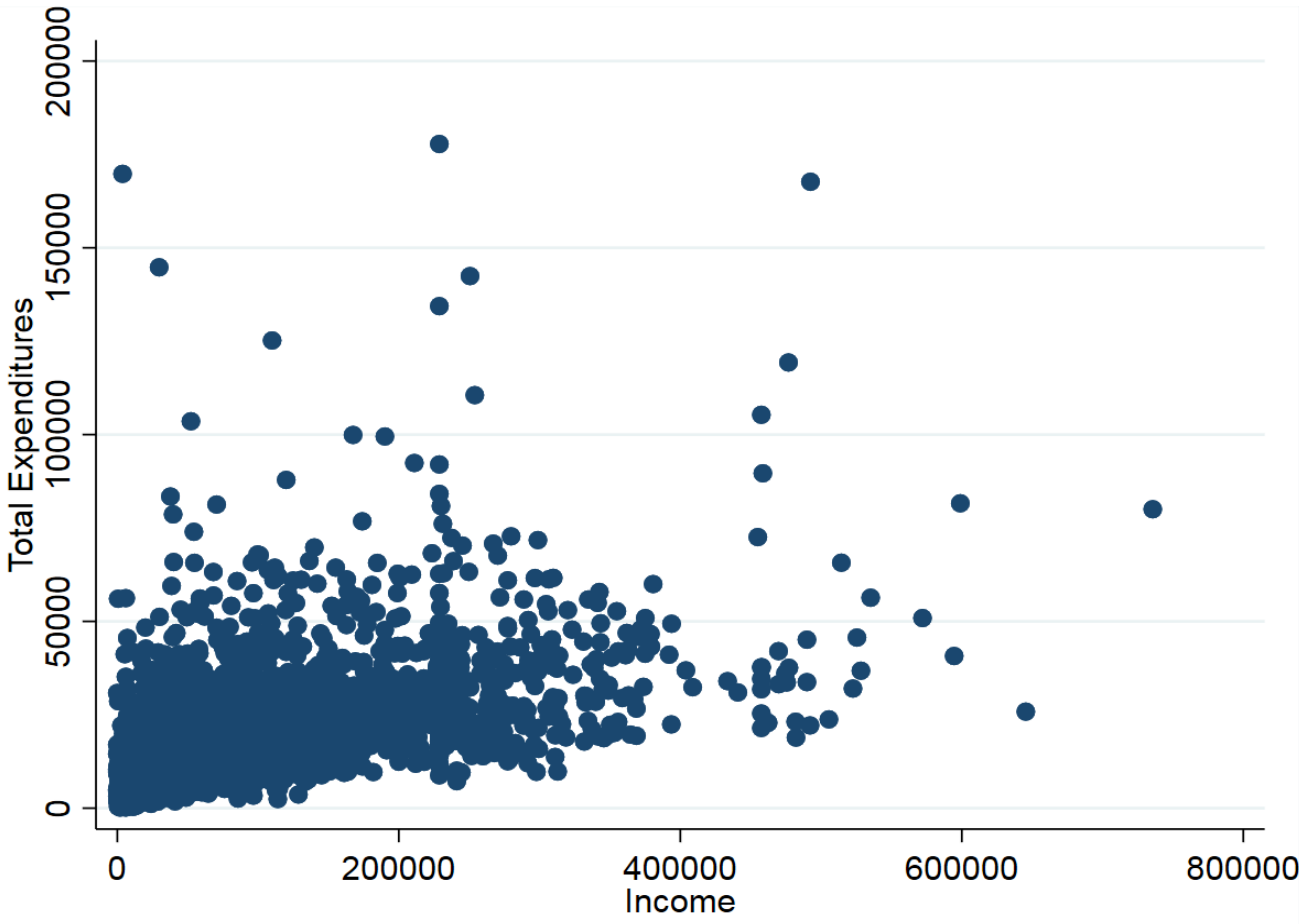


Transformed Income Distribution – First Quarter 2017



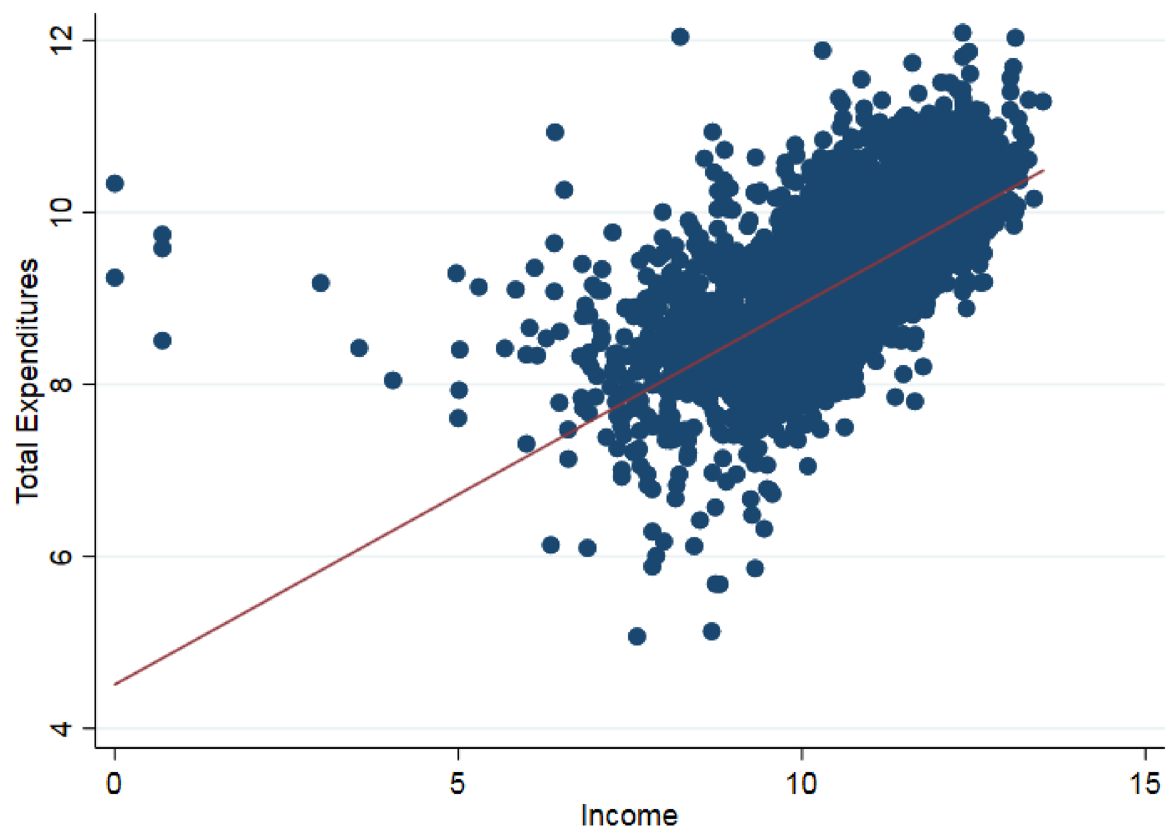
Income-Expenditure Relationship

- Bivariate income-expenditure relationship is heteroskedastic.
- Explore two strategies
  - Fix heteroskedasticity with a log transformation
  - Fix heteroskedasticity with a power transformation.



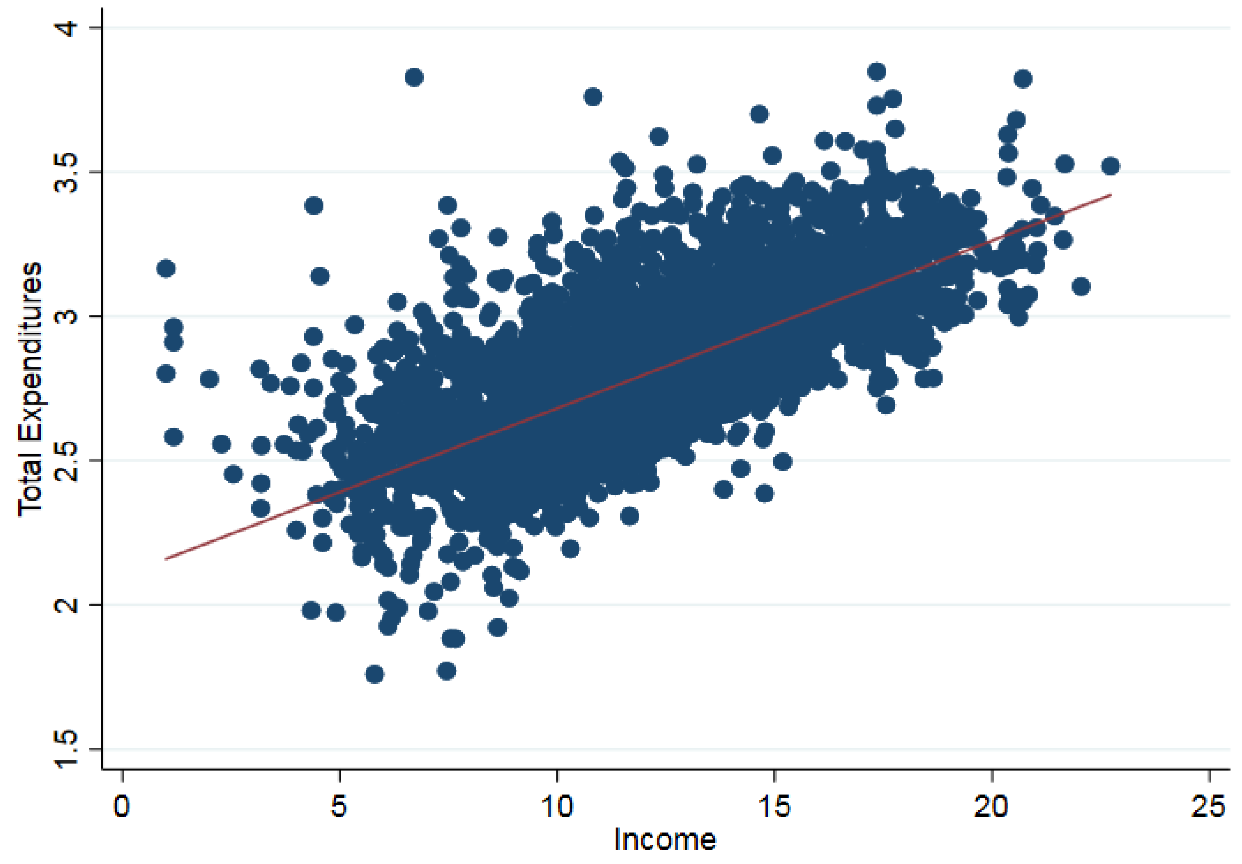
Income-Expenditure Relationship

- Log transformation improves the relationship between the variables.
- Outliers still seem to exist in the left tail of the distribution.
- Regression line is still biased.



Income-Expenditure Relationship

- Optimal lambda creates a much better shape to the data.
- Outliers are treated such that the bias is lessened or eliminated.
- Regression line describes the data more accurately.



Down side of power transformation

- Log transformations allow for direct interpretability of the coefficient as a constant elasticity.

$$\varepsilon_{y,x} = \frac{\mathrm{dy}}{\mathrm{dx}}\frac{x}{y} = \beta_1$$

- Power transformations require some additional calculus to 'correct' the beta and turn it into an elasticity at a point (x, y).

$$\varepsilon_{y,x} = \frac{\mathrm{dy}}{\mathrm{dx}}\frac{x}{y} = \beta_1\left(\frac{x^{m-1}}{y^{n-1}}\right)\left(\frac{m}{n}\right)\left(\frac{x}{y}\right)$$

Questions?

The paper accompanying this presentation can be found at the following address:

https://www.bls.gov/cex/ce_methodology.htm#reports

Contact Information

Taylor J. Wilson
(202) 691-6550
Wilson.Taylor@bls.gov