Labor Market Segmentation & the Income Distribution: New Evidence from Internal Census Bureau Data

Markus P. A. Schneider ¹  Ellis Scharfenaker ²

¹Department of Economics, University of Denver
²Department of Economics, University of Utah

October 24, 2018
The research in this paper was conducted while all authors were Special Sworn Status researchers of the U.S. Census Bureau at the Rocky Mountain Research Data Center at the University of Colorado.

Any opinions and conclusions expressed herein are those of the authors and do not necessarily represent the views of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed.
Based on Schneider (2013), which used public-use data (1996-2007)

**Idea:** Does the observed labor income distribution offer evidence of distinct segments?
Based on Schneider (2013), which used public-use data (1996-2007)

- **Idea:** Does the observed labor income distribution offer evidence of distinct segments?
  - Does the best-fitting functional form provide insights into how labor markets work?
Based on Schneider (2013), which used public-use data (1996-2007)

- **Idea:** Does the observed labor income distribution offer evidence of distinct segments?
  - Does the best-fitting functional form provide insights into how labor markets work?

- **Limits the RDC allows us to overcome:**
  - Public-use data is top-coded; RDC data is not.*

---

*Actually, it is, but the top-code limits are higher.*
Based on Schneider (2013), which used public-use data (1996-2007)

**Idea:** Does the observed labor income distribution offer evidence of distinct segments?
  - Does the best-fitting functional form provide insights into how labor markets work?

**Limits the RDC allows us to overcome:**
  - Public-use data is top-coded; RDC data is not.
  - Expansion in coverage (1975 - 2017)
Based on Schneider (2013), which used public-use data (1996-2007)

- **Idea:** Does the observed labor income distribution offer evidence of distinct segments?
  - Does the best-fitting functional form provide insights into how labor markets work?

- **Limits the RDC allows us to overcome:**
  - Public-use data is top-coded; RDC data is not.
  - Expansion in coverage (1975 - 2017)
  - More comprehensive consideration of models
Recent work downplays explanatory parsimony (vs. goodness-of-fit)
- Fully non-parametric models or highly flexible parametric distribution (e.g. GB2) are favored
Motivation

- Recent work downplays explanatory parsimony (vs. goodness-of-fit)
  - Fully non-parametric models or highly flexible parametric distribution (e.g. GB2) are favored
- We seek to revive two historical traditions:
  1. Statistical equilibrium models $\Rightarrow$ underlying market processes
  2. Theory of labor market segmentation

We solve at least one problem in the labor market segmentation literature: sorting workers into segments exogenously!
The model

- Likelihood for a $K$-component mixture model:

$$L[\Theta | \{x_i\}_n] = \prod_{i=1}^{N} \sum_{k=1}^{K} \lambda_k \ p_k [x_i | \theta_k]$$

(1)

$\lambda_k$ and $p_k$ are the component weight and pdf respectively.
The model

- Likelihood for a the $K$-component mixture model:

$$L[\Theta|\{x_i\}_n] = \prod_{i=1}^{N} \sum_{k=1}^{K} \lambda_k \, p_k \left[ x_i | \theta_k \right]$$

(1)

- Specifically: two-component exponential / log-normal mixture model:

$$p[x|\alpha, \mu, \sigma, \lambda] = \lambda \, \text{Exp}[x|\beta] + (1 - \lambda) \, \text{lgN}[x|\mu, \sigma]$$

(2)

$\lambda_k$ and $p_k$ are the component weight and pdf respectively.
Model components

- Exponential: characterized by memoryless property
  - Part-time (less than 35 hrs. on average), low-wage (?) jobs
  - **Speculation:** frequently changing hours & approx. constant real wage

- Log normal: described by evolution of incomes à la Gibrat’s law
  - Full-time (35+ hrs. on average)
  - **Speculation:** relatively constant hours & wage variation
Estimation

**Data:** Wage & salary data (WSAL_VAL) using weights (MARSUPWT)

- Both MLE & MCMC (Bayesian) estimation
  - Adjusted for censoring, since even the internal data is top-coded.
- Broad consideration of alternative models
ML estimates

(a) Component Weight

(b) $\beta$ (Exp)

(c) $\mu$ (lgN, scale)

(d) $\sigma$ (lgN, shape)

Scale parameters adjusted for inflation.
ML estimates: Inequality

Inequality (Gini) in the Mixture

- Overall
- Exp. w/in
- IgN w/in

Gini

1980 1990 2000 2010
Thank you!
# ML estimates: Parameter Interpretations

<table>
<thead>
<tr>
<th>Component</th>
<th>Mean</th>
<th>Median</th>
<th>Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exponential</td>
<td>$\beta$</td>
<td>$\frac{1}{\ln 2} \beta$</td>
<td>0</td>
</tr>
<tr>
<td>Log-Normal</td>
<td>$\exp(\mu + \frac{1}{2} \sigma^2)$</td>
<td>$\exp(\mu)$</td>
<td>$\exp(\mu - \sigma^2)$</td>
</tr>
</tbody>
</table>

![Graph showing data trends over years]
ML estimates: Inequality

<table>
<thead>
<tr>
<th>Component</th>
<th>Mean</th>
<th>Gini</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exponential</td>
<td>$\beta$</td>
<td>0.5</td>
</tr>
<tr>
<td>Log-Normal</td>
<td>$\exp\left(\mu + \frac{1}{2} \sigma^2\right)$</td>
<td>$2\Phi(\sigma/2) - 1$</td>
</tr>
</tbody>
</table>

Gini decomposition: \[ G = \sum p_i \cdot s_i \cdot G_{wi} + (G_b + G_t) \]

- $G_{wi}$:
  1. Gini of exponential component is constant
  2. Gini of log-normal component is *increasing*
  3. Population & income shares of log-normal component *increases*

- $G_b$: between-component inequality
- $G_t$: Transvariation (overlap) between components . . .

---

* $p_i$ is the population share, $s_i$ the income share captured in component $i$
ML estimates: Inequality

<table>
<thead>
<tr>
<th>Component</th>
<th>Mean</th>
<th>Gini</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exponential</td>
<td>$\beta$</td>
<td>0.5</td>
</tr>
<tr>
<td>Log-Normal</td>
<td>$\exp(\mu + \frac{1}{2} \sigma^2)$</td>
<td>$2\Phi(\sigma/2) - 1$</td>
</tr>
</tbody>
</table>

Gini decomposition: $G = \sum p_i \cdot s_i \cdot G_{wi} + (G_b + G_t)$

- $G_{wi}$:
  1. Gini of exponential component is constant
  2. Gini of log-normal component is increasing
  3. Population & income shares of log-normal component increases

- $G_b$: between-component inequality
- $G_t$: Transvariation (overlap) between components ...

$\Rightarrow G_b, G_t$ not calculated separately $\Leftarrow$

---

*p* is the population share, *s* the income share captured in component *i*
Model results

The model with inflation adjusted location parameters.
Camilo Dagum (1977) contrasted three approaches for finding a functional description of the income distribution.

1. Based only on Goodness-of-fit
2. Based on the generation of an income distribution by means of a stochastic process
3. Based on the solutions to differential equations that capture the regularity & permanence observed
Camilo Dagum (1977) contrasted three approaches for finding a functional description of the income distribution.

1. Based only on Goodness-of-fit
2. Based on the generation of an income distribution by means of a stochastic process
3. Based on the solutions to differential equations that capture the regularity & permanence observed

Failure to consider mixture models with a finite number of components ⇒ Approach 1!
Further Work

- Theoretical Model
- Incorporating power-law tail (Pareto component)
- Refining who (type of worker, demographic) is captured in each component (with latent variables)
  - Implication for labor market segmentation & stratification
- Solving the mystery:
  Why is the exponential component apparently shrinking over the past 40 years?