Alienation & Segmentation Between Latent Classes of American Household Income Distributions, 1977-2015

Gordon Anderson, University of Toronto
& Teng Wah Leo, St. Francis Xavier University

17th May, 2018.

Motivation

- Sociology & political science literatures regarding segmentation and alienation of households in American society → emergence of disenchanted underclass & vanishing middle class with groups emerging, disappearing or reshaping over time (Temin, 2017).
- Parallel expression in economic segmentation & alienation of household groupings in terms of income size distribution.
- Household income inequality has grown steadily in the US. Following Gibrat (1931), classic models of income processes (Freidman, 1957; Hall, 1978; Blundell & Preston, 1998; Batistin, Blundell & Lewbel, 2006) → current income/consumption does not affect expected rate of growth nor the variance of its growth rate.
- But subgroup segmentation/alienation in this context is an open question.

Motivation (Cont'd)

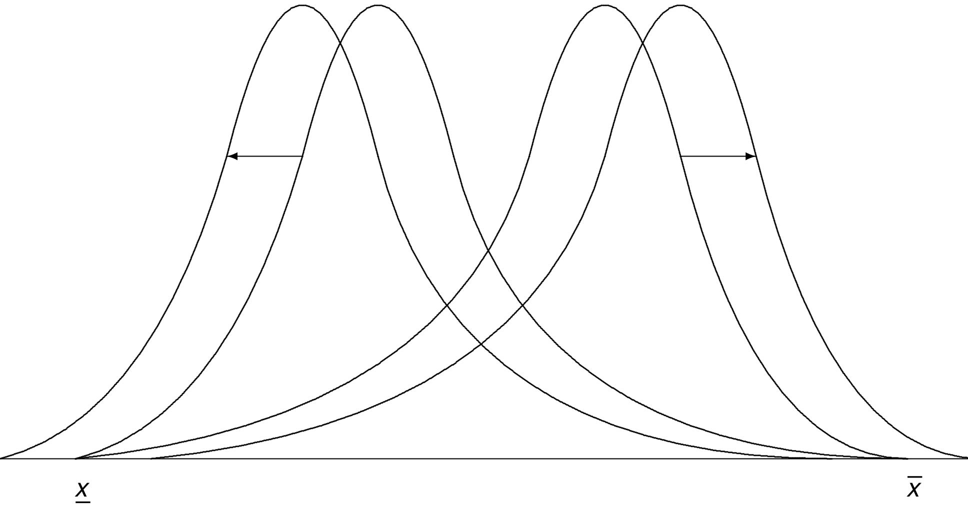
- One problem with studying these phenomena is that a stringent specification of the nature of the subgroups $\longrightarrow$ undue influence on conclusions of the analysis.
- Is this really innocuous?
- Choosing quantiles or cohorts fixes class sizes, while tying class locations to the overall distributional location locks the progress of subgroups into the progress of the overall distribution.
- **Solution:** Semiparametric (Empirical Bayes') analyses of a mixture of subgroup distributions developed, which facilitates the determination of class sizes, their locations and spreads, with the number of classes being independently determined. In addition, correlates of household class membership probabilities can be explored.

Outline

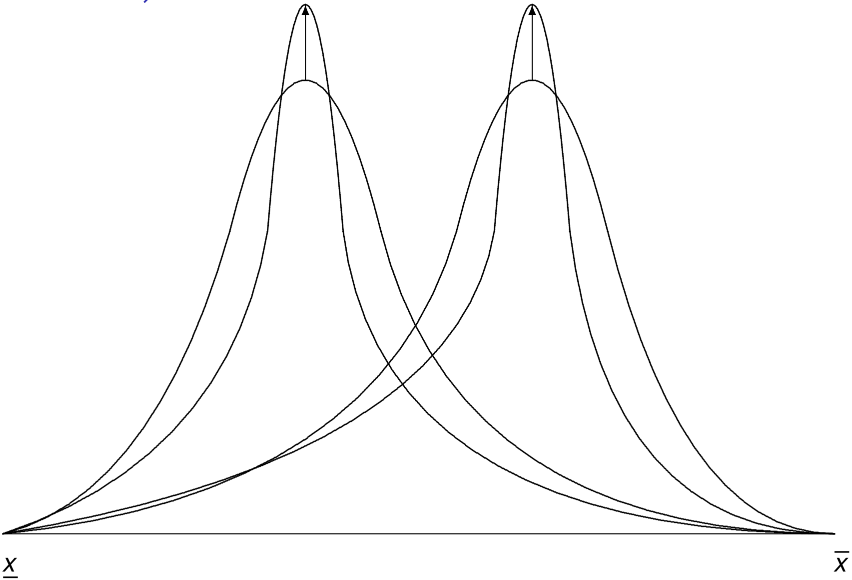
Intricacies in Analysis

- $\nearrow$ overall inequality $\Rightarrow$ nothing regarding subgroup inequalities.
- Overall inequality can $\nearrow$ with $\nearrow$ within subgroup inequalities, while at the same time between subgroup inequality is $\searrow$ diminishing.
- Similarly overall inequality can $\nearrow$, with $\searrow$ within subgroup inequalities, while at the same time between subgroup inequality $\nearrow$.
- Essence of distinction between polarization & inequality (Duclos, Esteban & Ray, 2004) which allows those measures to go in different directions.

Intuition: Location Shift



Intuition: $\nearrow$ Identification



Theoretical Basis: Gibrat's Law

- Right tail of distribution of wealth, income, firm size, or city size, conforms well to a Pareto distribution.
- Essentially models income as accumulation of random multiplicative shocks.
- Based on astronomer Jacobus Kapteyn → assume that underlying such distributions was a simple Gaussian process: a large number of small additive influences, operating independently of each other, would generate a normally distributed variate ("Law of Laplace").
- Gibrat postulated the simplest such process be the logarithm of the variate.
- Champernowne (1953) & Simon (1955) showed that Pareto distributions arise naturally if the time series behavior of variable in question satisfies Gibrat's law: **current position of variable does not influence its expected rate of growth, nor variance of growth rate.**

Theoretical Basis: Gibrat's Law, Cont'd

- Battistin, Blundell & Lewbel (2006) noted significant departures from ln-normality in income data → violation of Gibrat's law. Showed empirically that the within cohort distribution of consumption expenditures across households is closer to ln normalilty than income.
- Perhaps there are no violations, or that there is more to it than meets the eye?

Theoretical Basis: Income Size Distribution

- Denote household income variate as X with an overall distribution $f(X)$.
- Class is considered **latent** with undetermined number of classes indexed $k = 1, .., K \rightarrow$ common income process within household class.
- Following Modigliani & Brumberg (1954), Friedman (1957), Blundell & Preston (1998), & Batistin, Blundell & Lewbel (2006), agents from each class with "no bequest" preferences are assumed to maximize the present value of lifetime happiness,

$$\int_0^T U_k(C_t) e^{-r_k^* t} dt$$

subject to the present value of lifetime wealth

$$\int_0^T X_t e^{-rt} dt$$

Theoretical Basis: Income Size Distribution, Cont'd

- where $U_k(.)$ is an instantaneous consumption utility function, r_k^* is the class rate of time preference, r is the market lending rate.
- This all leads to a consumption smoothing model of the form: $C_t = e^{g_k t} C_0$ with $g_k = \frac{r - r_k}{\gamma_k}$ (where γ_k is a class CRRA parameter), & a one period process $C_t = e^{g_k} C_{t-1}$ or $\ln(C_t) = g_k + \ln(C_{t-1})$ reflective of permanent income process for each class of a form based upon the law of proportionate effects (Gibrat's Law) where its income in period t has form: $X_{k,t} = (1 + \delta_{k,t}) X_{k,t-1}$, where $\delta_{k,t}$ is a random variable with mean δ_k, & variance σ_k^2.

Significance of Gibrat's Law

- Following Gibrat (1931), if the income process has a starting value $x_{k,0}$ $\rightarrow$ letting $X = \ln(x)$, after T periods the variate $x_{k,T}$ will be such that:

$$\begin{aligned} X_{k,T} &= \ln(x_{k,T}) \\ &\sim \mathcal{N}\left(\ln(x_{k,0}) + T\left(\delta_k + 0.5\sigma_k^2\right), T\sigma_k^2\right) = \mathcal{N}\left(\mu_{k,T}, \sigma_{k,T}^2\right) \end{aligned}$$

  where $\delta_{k,t}$ corresponds to period t income growth rate of the k^{th} class, with $\sigma_{k,T}^2$ the T^{th} period variance.

- ln-normality holds an added convenience here since, following Aitchison and Brown (1957), when X is ln-normally distributed with $\mathbf{E}(X) = e^{\theta + 0.5\sigma^2}$, where θ & σ^2 are respectively the mean & variance of $\ln(X)$, & Gini Coefficient is given by $\left[2\Phi\left(\frac{\sigma_{k,T}}{2}\right) - 1\right]$ $\rightarrow$ subgroup Gini's grow over time since $\sigma_{k,T}^2$ grows over time.

Mixture Distribution & Income Classes

- Given K classes where the proportion of the population in class k at time T is $w_{k,T}$, $f(X)$ the distribution of income at time T is:

$$f(X_T) = \sum_{k=1}^{K} w_{k,T} f(X_T; \mu_{k,T}, \sigma^2_{k,T}) \quad (1)$$

where $\sum_{k=1}^{K} w_{k,T} = 1$.

- Assume $k > j \Leftrightarrow \mu_{k,T} > \mu_{j,T}$. For a given K, & independent random sample X_i, $i = 1, \ldots, N$, the $3K - 1$ parameters of the semi-parametric mixture distribution can be estimated by psuedo maximum likelihood methods (Anderson et. al., 2016 & 2018) $\rightarrow$ EM Algorithm.

Mixture Distribution & Income Classes, Cont'd

- This approach does not depend upon arbitrarily determined boundaries, since class boundaries are fuzzy $\rightarrow$ all that can be determined is $\Pr(X,k)$, the probability that an agent with x is in class k, which is given by:

$$\Pr(X_T,K) = \frac{w_{kT} f\left(X_T; \mu_{kT}, \sigma^2_{kT}\right)}{\sum_{j=1}^{K} w_{jT} f\left(X_T; \mu_{jT}, \sigma^2_{jT}\right)}$$

- With this information, subgroup decomposition of inequality measures can be explored, & polarization & segmentation measures developed.

Mixture Distribution & Income Classes

- Class membership correlates are examined by modifying (1) as

$$f(X_{i,T}) = \sum_{k=1}^{K} w_{i,k,T} f(X_{i,T}; \mu_{k,T}, \sigma^2_{k,T})$$

where $w_{i,k,T}$ is prior probability of agent i with income x_i belonging to class k, given correlates vector $\mathbf{z}_{i,T}$, it may be modelled as:

$$w_{i,k,T} = \Pr(\mathbb{1}(i \in C_k) = 1) = g(\beta'_k \mathbf{z}_{i,t}) \quad (2)$$

for $k = 2, \ldots, K$, and

$$w_{i,1,T} = 1 - \sum_{k=2}^{K} g(\beta'_k \mathbf{z}_{i,T})$$

where $g(.)$ is the link function, eg. a logistic function.

Decomposing GINI

- $$GINI = \frac{1}{2\,\mathbf{E}(x)} \int_0^{\infty} \int_0^{\infty} f(y)[f(x)|x-y|dx]dy$$

  where $\mathbf{E}(x) = \mu = \sum_{k=1}^{K} w_k \mu_k$, unlike some other inequality measures, is generally not subgroup decomposable (Bourguignon 1971).
- This is advantageous when $f(.)$ construed as mixture distribution.

Decomposing GINI, Cont'd

- When class income distributions have compact, mutually exclusive support (i.e. segmented) → GINI can be decomposed into weighted sum of within subgroup *GINI*'s ($\mathcal{G}_k$) plus a between group *GINI* ($\mathcal{B}$) (Mookerjee & Shorrocks, 1982) in the form:

$$
\begin{aligned}
GINI &= \sum_{k=1}^{k} w_k^2 \frac{\mu_k}{\mu} \mathcal{G}_k + \sum_{k=1}^{K} \sum_{h=1}^{K} \frac{w_k w_h}{2\mu} |\mu_k - \mu_h| \\
&= \sum_{k=1}^{K} w_k^2 \frac{\mu_k}{\mu} \mathcal{G}_k + \mathcal{B}
\end{aligned}
$$

Decomposing GINI, Cont'd

- When constituent classes are not segmented → *GINI* may be more generally written as:

$$
\begin{aligned}
GINI &= \sum_{k=1}^{K} w_k^2 \frac{\mu_k}{\mu} \mathcal{G}_k + \mathcal{B} \\
&\quad + \frac{2}{\mu} \sum_{k=2}^{K} \sum_{j=1}^{k-1} w_k w_j \int_0^{\infty} f_k(y) \left[\int_y^{\infty} f_j(x)(x-y)dx \right] dy \\
&= \sum_{k=1}^{K} w_k^2 \frac{\mu_k}{\mu} \mathcal{G}_k + \mathcal{B} + \mathcal{F}
\end{aligned}
$$

where $\mathcal{F}$ is the Non-Segmentation Factor.

- $\therefore \mathcal{F}$ is a function of extent of distributional overlap → measure of degree to which class distributions are not segmented.

Derived Measures

- $\therefore$ A segmentation index ($\mathcal{S}$) is given by:

$$\mathcal{S} = 1 - \frac{\mathcal{F}}{GINI} \tag{3}$$

  & $\mathcal{S} \in [0, 1]$.

- As noted, Polarization has as much to do with Segmentation as the distance between groups (Alienation) $\rightarrow$ net effect of the two forces. The index in (3) only reflects the former.
- Joint effect of Segmentation & Alienation $\rightarrow$ *GINI* based Polarization Index ($\mathcal{P}$) can be constructed in the form:

$$\mathcal{P} = \mathcal{S}^{\alpha}\mathcal{B}^{1-\alpha} \tag{4}$$

Data

- Household income & covariates of latent class drawn from US March Current Population Survey between 1976-2016 from IPUMS, 4 decades.
- Used odd years, 1977-2015.
- Income is adult equivalized using the square root rule.
- Household incomes below $1,200 were dropped $\rightarrow 498,333$ observations for odd years.
- Covariates include: Age of Head & Spouse, Educational Attainment, Family Size, Degree of Assortative Matching, Occupation of Head & Spouse, Metro, Region, & Trend.

Summary Statistics

Class Covariates (4 Component Model)

Component	2^{nd}	3^{rd}	4^{th}		2^{nd}	3^{rd}	4^{th}
Age	0.1176	0.2979	0.5505	$<$ Grade 9,	-1.3810	-2.7235	-4.0376
	(0.0038)	(0.0042)	(0.0077)	Husband	(0.0168)	(0.0237)	(0.0812)
Age^2	-0.0013	-0.0031	-0.0052	$\geq$ Grade 9, w/o	-0.6796	-1.1756	-1.9209
	(0.0000)	(0.0000)	(0.0001)	High Sch, Husbad	(0.0101)	(0.0112)	(0.0251)
Age of Spouse	0.1532	0.3284	0.4501	University Grad,	0.4306	1.0735	2.0875
	(0.0035)	(0.0039)	(0.0071)	Husband	(0.0116)	(0.0095)	(0.0131)
Age^2 of Spouse	-0.0013	-0.0029	-0.0042	$<$ Grade 9, Wife	-1.2322	-2.4136	-3.5193
	(0.0000)	(0.0000)	(0.0001)		(0.0189)	(0.0290)	(0.0937)
Family Size	-0.0640	-1.3331	-2.5279	$\geq$ Grade 9, w/o	-0.5069	-0.8396	-1.3860
	(0.0204)	(0.0237)	(0.0457)	High Sch, Wife	(0.0103)	(0.0110)	(0.0214)
Family $Size^2$	-0.0286	0.0508	0.0892	University Grad,	0.3789	1.0644	1.7381
	(0.0023)	(0.0029)	(0.0062)	Wife	(0.0120)	(0.0099)	(0.0135)
Family Size $\times$ Trend	-0.0137	0.0174	0.0075				
	(0.0018)	(0.0021)	(0.0038)				
Family $Size^2$ $\times$ Trend	0.0010	-0.0024	0.0021				
	(0.0002)	(0.0003)	(0.0005)				
Black	-0.5189	-0.9246	-1.5912				
	(0.0133)	(0.0148)	(0.0280)				
Not in Metro	-0.1873	-1.0027	-1.9584				
	(0.0078)	(0.0082)	(0.0154)				
Trend	-0.0957	-0.2733	-0.3462				
	(0.0037)	(0.0040)	(0.0066)				

Standard errors are in parenthesis. Educational Attainment is compared against high school graduate for both head & spouse.

Class Covariates (4 Component Model), Cont'd

Component	2nd	3rd	4th		2nd	3rd	4th
		Husband				Wife	
Technical/	-0.2349	-0.6779	-1.3166	Homework	-1.6002	-2.7451	-2.5813
Sales/Admin	(0.0112)	(0.0096)	(0.0134)		(0.0092)	(0.0099)	(0.0162)
Service	-1.0459	-2.0564	-3.7812	Managerial/	0.1825	0.7303	1.1988
	(0.0142)	(0.0151)	(0.0335)	Professional	(0.0109)	(0.0091)	(0.0135)
Farming/Forestry/	-1.9552	-3.5204	-4.3223	Service	-0.8114	-1.9479	-2.6275
Fishing	(0.0188)	(0.0239)	(0.0445)		(0.0110)	(0.0131)	(0.0315)
Production/	-0.2987	-0.9483	-2.7021	Farming/Forestry/	-1.2557	-2.0447	-1.7143
Craft/Repair	(0.0110)	(0.0106)	(0.0224)	Fishing	(0.0333)	(0.0434)	(0.0751)
Operator/Fabricator/	-0.5611	-1.5929	-3.6602	Production/	-0.1374	-0.4005	-0.4758
Laborer	(0.0114)	(0.0117)	(0.0304)	Craft/Repair	(0.0276)	(0.0288)	(0.0577
				Operator/Fabricator/	-0.3421	-0.9357	-1.6814
				Laborer	(0.0147)	(0.0170)	(0.0502)
New England	0.4280	0.7382	0.7467	East-South	-0.0981	-0.2955	-0.7006
Div	(0.0147)	(0.0137)	(0.0205)	Central Div	(0.0165)	(0.0177)	(0.0307)
Mid Atlantic	0.2736	0.4991	0.6731	West-South	-0.3136	-0.5197	-0.6958
Div	(0.0134)	(0.0128)	(0.0193)	Central Div	(0.0139)	(0.0146)	(0.0232)
East-North	0.4396	0.7237	0.5688	Mountain Div	0.0412	-0.0562	-0.3702
Central Div	(0.0126)	(0.0122)	(0.0192)		(0.0133)	(0.0137)	(0.0221)
West-North	0.3152	0.3196	-0.1142	Pacific Div	-0.0848	0.2287	0.5089
Central Div	(0.0134)	(0.0134)	(0.0219)		(0.0135)	(0.0131)	(0.0192

Standard errors are in parenthesis. Occupation for husband is compared against managerial/professional occupations, while that for the wife is compared against Technical/Sales/Administrative Support. Regional effects are compared against South Atlantic Div.

On Class Covariates

- Education, relative to completed high school, & age (both parents) $\nearrow$ probability of higher class membership (the latter with diminishing effect).
- Family size $\searrow$ probability of higher class membership.
- Black, relative to white & non-metro, relative to metro both $\searrow$ probability of higher class membership.
- For husbands, relative to managerial & professional, all occupational categories $\searrow$ the probability of higher class membership. For wives relative to technical/admin/sales support, all occupational categories $\searrow$ probability of higher class membership, except for managerial & professional occupations which $\nearrow$ probability of higher class membership.
- Regionally, East-South Central & West-South Central have negative effect on higher income class membership. New England, Mid-Atlantic, & East-North Central have positive effect on higher income class membership.

Figure 1: Class Membership Shares

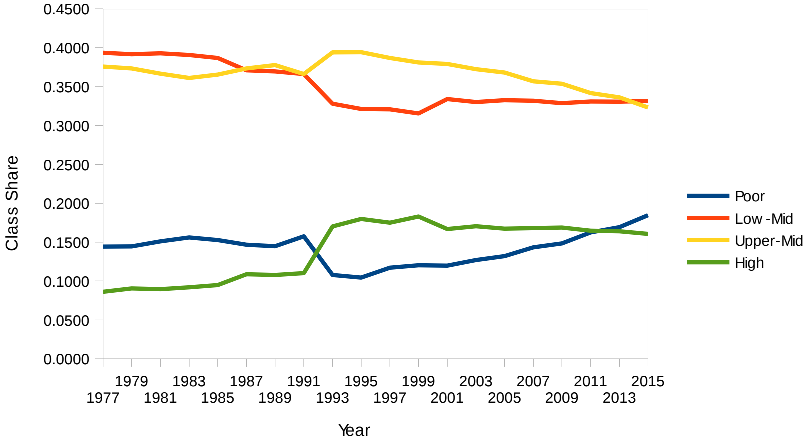
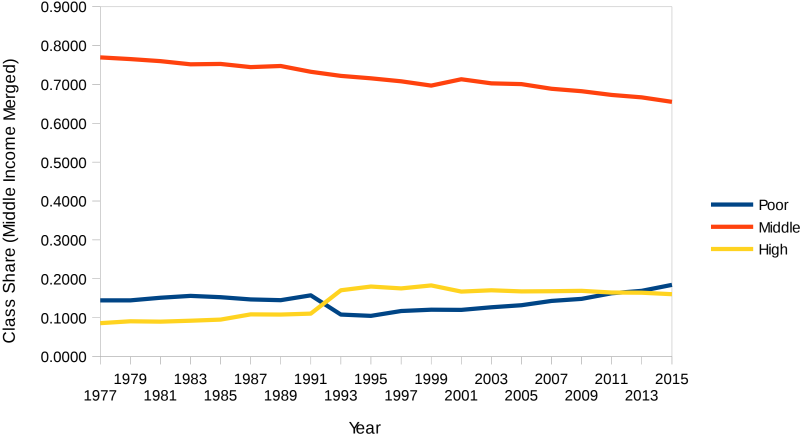


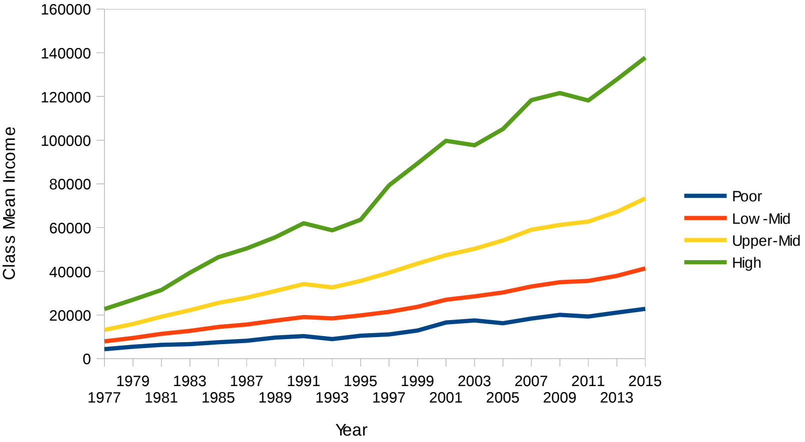
Figure 1a: Class Membership Shares, Combined Middle



On Class Shares

- Steadily $\searrow$ middle class.
- Clear structural break in 1991.
- $\nearrow$ poor & high income groups over the period.

Figure 2: Class Mean Income Processes

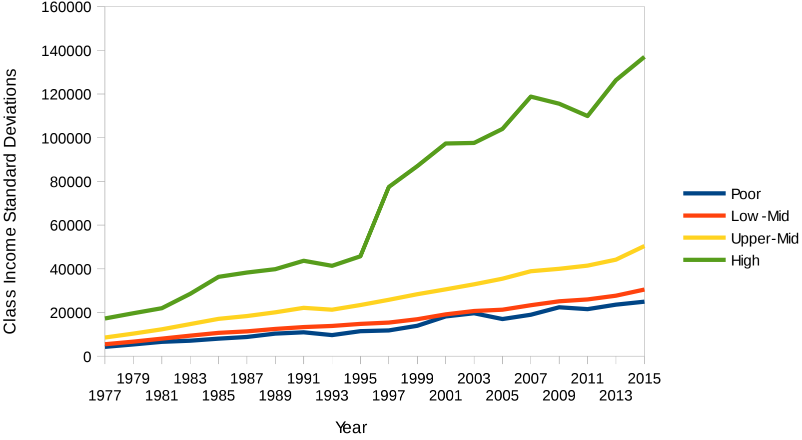


Class Average Mean Income Processes, Cont'd

Table: Growth Regressions Class Membership Probabilities

Class	2 Year Coefficient	Unrestricted		Restricted (χ_6 = 1.4658)	
Low	Δ	0.0023	(0.5418)	0.0032	(1.8197)
	Δ dum(77-91)	-0.0004	(0.0584)	-0.0006	(0.1981)
Lower-Middle	Δ	-0.0029	(0.9191)	-0.0032	(1.8197)
	Δ dum(77-91)	-0.0010	(0.2017)	0.0006	(0.1981)
Upper-Middle	Δ	-0.0036	(1.2970)	-0.0032	(1.8197)
	Δ dum(77-91)	0.0022	(0.4828)	0.0006	(0.1981)
High	Δ	0.0042	(0.9448)	0.0032	(1.8197)
	Δ dum(77-91)	-0.0008	(0.1034)	-0.0006	(0.1981)

Figure 3: Class Standard Deviation of Income



On Class Shares

- Steady growth rate in subgroup income means (3.1% implied annual growth rate).
- $\nearrow$ variation over the period in all classes, consistent with Gibrat's law.
- Jump in high income group variation post structural break.

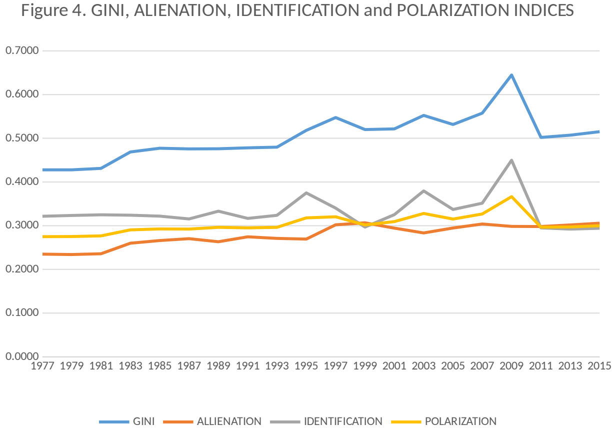


Figure 4. GINI, ALIENATION, IDENTIFICATION and POLARIZATION INDICES
0.7000
0.6000
0.5000
0.4000
0.3000
0.2000
0.1000
0.0000
1977 1979 1981 1983 1985 1987 1989 1991 1993 1995 1997 1999 2001 2003 2005 2007 2009 2011 2013 2015
GINI
ALLIENATION
IDENTIFICATION
POLARIZATION

Change in Alienation/Segmentation & Polarization

- $\nearrow$ Gini.
- $\nearrow$ Alienation (relative group differences).
- Constant Identification (degree of segmentation)
- $\nearrow$ Polarization (Monotonic Increasing function of Identification & Segmentation)

Number of Classes & Structural Break

Table: Choice of Number of Classes

AIC	10961196.2135	10878321.3121	10848468.4075	10837313.7122	10825287.8543
BIC	10961394.8233	10878650.6309	10848928.4354	10837904.4490	10826009.3002
Likelihood	-5480364.1067	-5438772.6560	-5423692.2038	-5417960.8561	-5411793.9272
Components	2	3	4	5	6
# of Parameters	117	194	271	348	425
# of Observations	498333	498333	498333	498333	498333

- Model is suggesting more than 4 components $\rightarrow$ Overfitting? $\Rightarrow$ Currently estimating > 6 components.
- At least one structural break in 1991. Is it significant? $\rightarrow$ Currently estimating ≥ 4 components splitting data into two sets to verify significance of structural break, & whether there is Δ in number of classes.
- Additional classes principally drawn from middle components $\Rightarrow$ more complicated narrative.

Other Issues

- Has the transition matrix changed over the years? Data not panel. → Examine Latent-Markov transition on state/metro level data over same years. → But examines Alienation/Segmentation/Polarization between states.
- How much of these cross time changes result from ruling government at state & federal levels? → including voting pattern at both levels.
- Ideally, need large panel to discern both Latent Class & Latent Markov covariates.

Conclusion

- Need a more robust approach to examining Segmentation/Alienation/Polarization → Use Latent Class model to discern number of components.
- Violation of Gibrat's Law not seen when Latent Class model used.
- When approach applied to US:
  1. Segmentation seems to have been constant over the past 4 decades.
  2. Observed divisiveness may possibly be due to falling middle income shares, increasing poverty share, & ballooning income of highest status class → Polarization
- However, caveat to this agreement with common narative:
  1. There are more than 4 components.
  2. Preliminary results suggest this clean narative is more complicated.