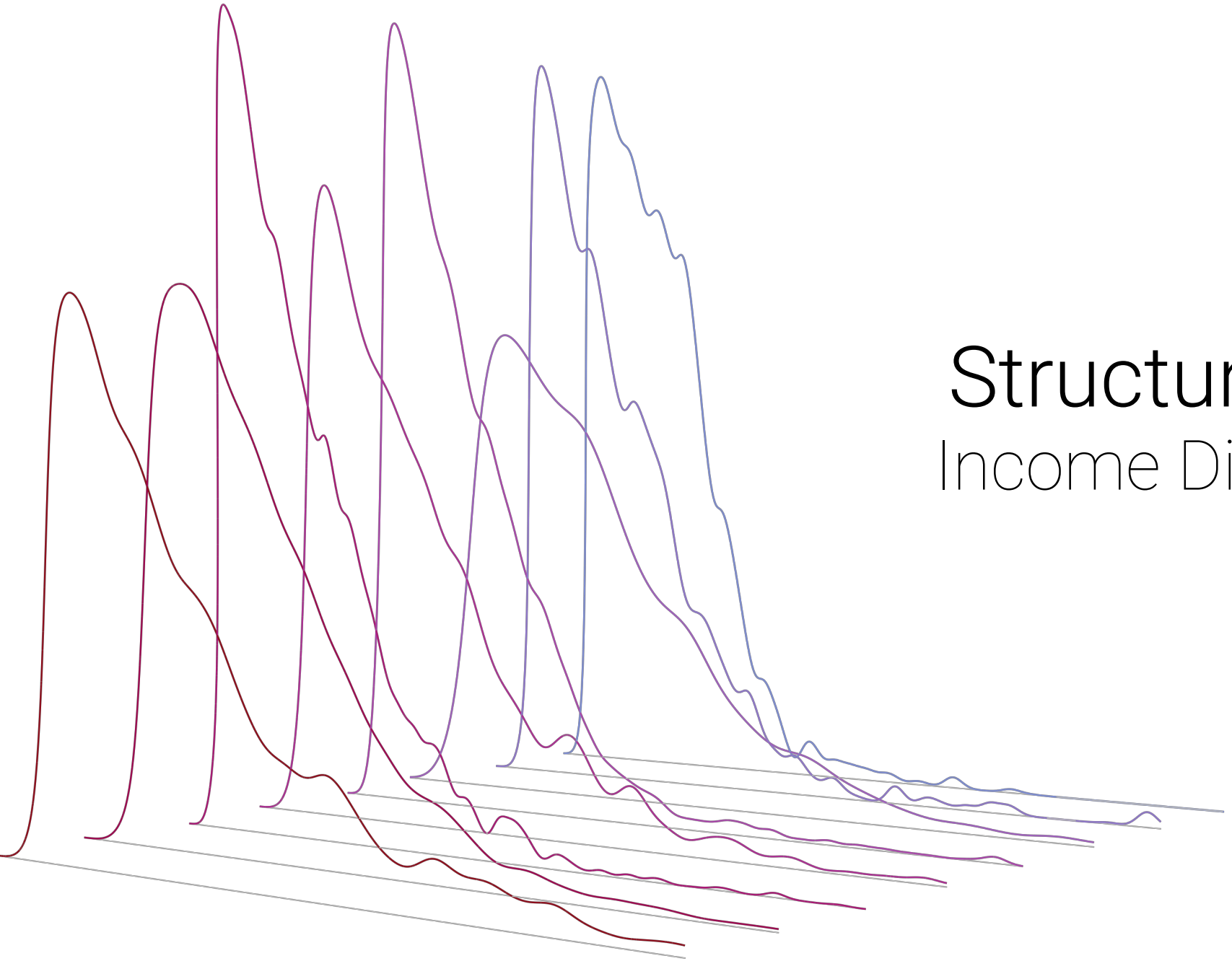
HIGHER SCHOOL OF ECONOMICS
NATIONAL RESEARCH UNIVERSITY



Structure of Household
Income Distribution in Russia

Alan Nartikoev
Anatoly Peresetsky

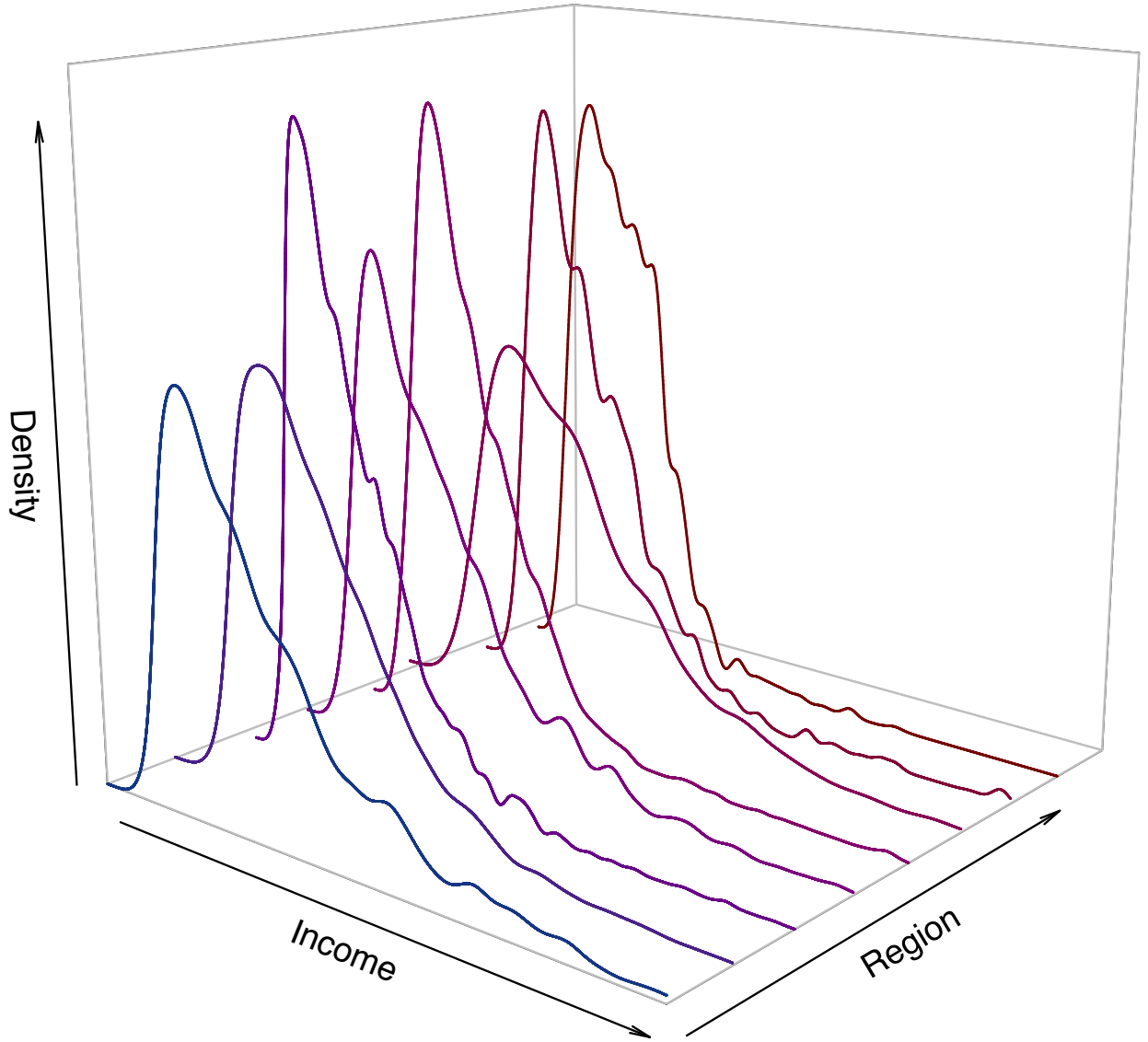
Premises

- Income distribution is heterogeneous: it inevitably includes different groups of population;
- It is possible to reveal relatively homogeneous groups inside the overall distribution;
- Group membership is determined by individual's characteristics;
- Income distribution also exhibits spatial heterogeneity across different regions;

Data

- Statistical Survey of Income and Participation in Social Programs 2017 (Federal State Statistics Service)
- 160 thousand observed households across Russia;
- Data is weighted with an inverse probability of sample inclusion;
- OECD equivalence scale is applied to household disposable income

Kernel Density Estimation



Descriptive Statistics

	n obs.	Min.	1st Quartile	Median	Mean	3rd Quartile	Max.
Central	40,560	820	185,400	271,600	365,000	459,800	4,688,000
Northwestern	17,448	2,000	215,600	311,600	367,600	459,800	3,110,000
Volga	31,536	780	155,900	210,300	238,000	287,800	2,677,000
Ural	13,152	1,200	169,200	240,900	300,100	363,400	3,931,000
Siberian	21,936	1,150	151,600	213,100	250,300	304,200	5,513,000
Far Eastern	10,200	800	215,600	322,700	397,300	502,400	3,500,000
Southern	16,584	1,800	151,700	206,500	235,100	284,600	8,655,000
North Caucasian	8,592	521,700	116,300	169,200	188,600	234,200	3,560,000

Model Description (Flachaire and Nuñez, 2007)

Finite mixture of distributions (generalized linear model):

$$f(y;\boldsymbol{\beta},\boldsymbol{\theta}) = \sum_{k=1}^{K} \pi_k(x;\boldsymbol{\beta}_k) f_k(y;\boldsymbol{\theta_k})$$

Individual's $\boldsymbol{i}$ probability of belonging to group $\boldsymbol{k}, \boldsymbol{\pi_{ik}(\cdot)}$, is a *concomitant variable*, i.e.

$$\pi_{ik} = F_k(\boldsymbol{x}_i;\boldsymbol{\beta}_k),$$

Where $F_k(x_i;\beta_k)$ is a distribution function depending on the parameter vector β_k, such that

$$\sum_{k=1}^{K} F_k(\boldsymbol{x}_i) = 1$$

Model Description (Flachaire and Nuñez, 2007)

We choose $k = 3$ and f_k being LogNormal, i.e.

$$f(y; \boldsymbol{\beta}, \boldsymbol{\theta}) = \sum_{k=1}^{3} \pi_k(x; \boldsymbol{\beta}_k) \frac{1}{y\sigma_k\sqrt{2\pi}} \exp\left(\frac{-(\ln y - \mu_k)^2}{2\sigma_k^2}\right),$$

where parameter vector $\boldsymbol{\theta}$ is given by $(\mu_1, \mu_2, \mu_3, \sigma_1, \sigma_2, \sigma_3)'$. We also choose $\boldsymbol{\pi_k}$ to be modelled via multinomial logistic regression:

$$\pi_{ik} = \pi_k(\boldsymbol{x_i}; \boldsymbol{\beta}) = \frac{e^{\boldsymbol{x}_i^\intercal \boldsymbol{\beta}_k}}{\sum_{l=1}^{K} e^{\boldsymbol{x}_i^\intercal \boldsymbol{\beta}_l}},$$

Estimation via EM-algorithm

j-th E-step:

$$\hat{p}_{ik}^{(j)} = \frac{\pi_k(\boldsymbol{x}_i; \boldsymbol{\beta}_k^{(j)}) f(y_i; \boldsymbol{\theta}_k^{(j)})}{\sum_{l=1}^{K} \pi_l(\boldsymbol{x}_i; \boldsymbol{\beta}_l^{(j)}) f(y_i; \boldsymbol{\theta}_l^{(j)})},$$

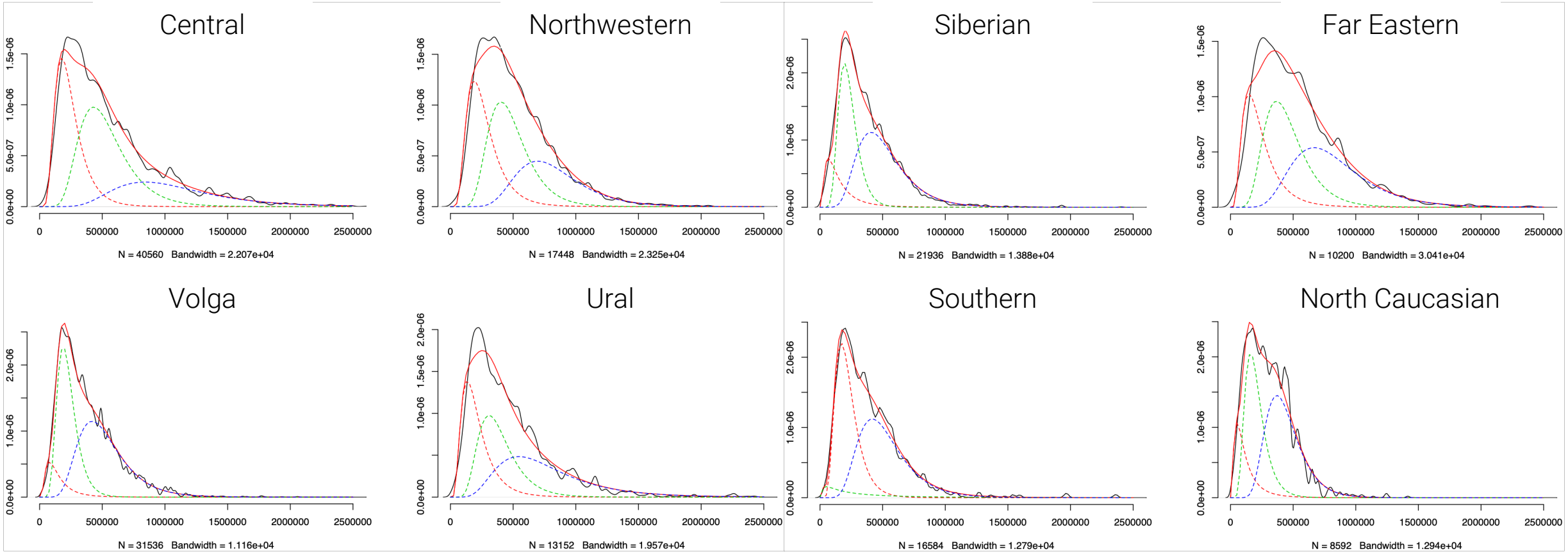
j-th M-step:

$$\hat{\boldsymbol{\Psi}}^{(j+1)} = (\hat{\boldsymbol{\beta}}^{(j+1)}, \hat{\boldsymbol{\theta}}^{(j+1)})^\intercal = \arg\max_{\boldsymbol{\beta}, \boldsymbol{\theta}} Q(\boldsymbol{\beta}^{(j+1)}, \boldsymbol{\theta}^{(j+1)}; \boldsymbol{x}; \boldsymbol{y}),$$

where

$$Q(\boldsymbol{\beta}^{(j+1)}, \boldsymbol{\theta}^{(j+1)}; \boldsymbol{x}; \boldsymbol{y}) = \sum_{i=1}^{n} \sum_{k=1}^{K} \hat{p}_{ik}^{(j)} \ln(\pi_k(\boldsymbol{\beta}_k^{(j+1)}; x_i)) + \sum_{i=1}^{n} \sum_{k=1}^{K} \hat{p}_{ik}^{(j)} \ln(f(\boldsymbol{\theta}_k^{(j+1)}; y_i))$$

Distribution Estimates



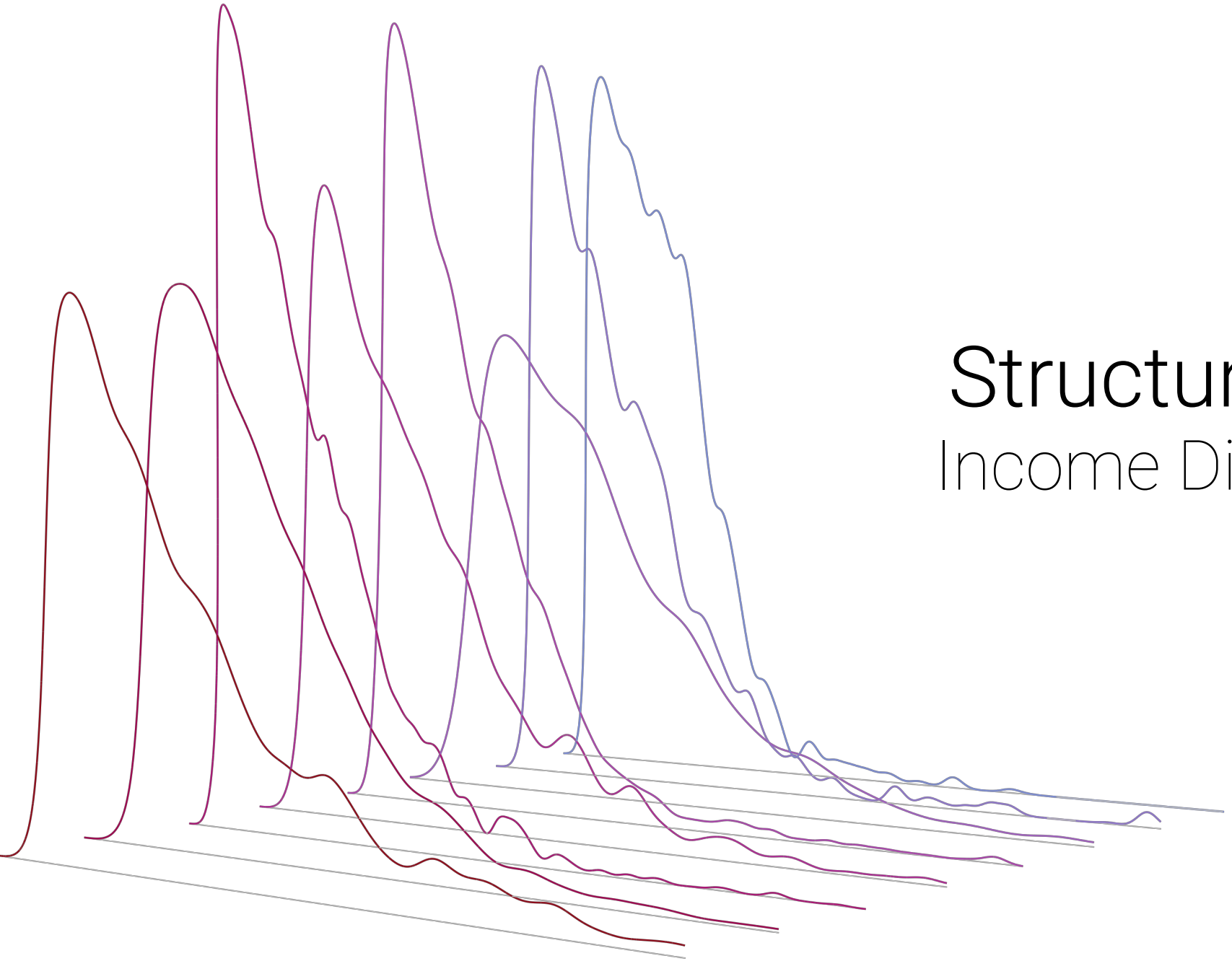
Distribution Estimates

	Central		Northwestern		Volga		Ural	
	II	III	II	III	II	III	II	III
(Intercept)	−3.469	−11.103	−2.875	−7.627	3.088	*-0.004*	−1.838	−5.772
edu = sec.	−1.929	−4.173	−1.642	−3.397	−1.150	−2.937	−2.029	−5.595
edu = prim.	−3.255	−5.863	−3.683	−6.076	−1.331	−4.044	−3.151	−7.887
loc = med.	0.547	0.376	0.457	0.096	0.631	0.874	0.380	0.499
loc = big	2.531	6.107	1.977	3.913	1.082	2.059	1.071	1.263
loc = rural	−1.030	−2.400	−1.246	−2.683	−1.168	−1.322	−1.756	−4.161
employed ratio	8.534	15.421	8.200	13.540	0.288	7.257	8.295	15.800
	Siberian		Far Eastern		Southern		North Caucasian	
	II	III	II	III	II	III	II	III
(Intercept)	2.545	−0.367	−2.022	−5.583	−4.007	−2.986	1.457	−2.237
edu = sec.	−0.984	−2.781	−2.454	−4.856	−0.212	−1.728	−0.413	−2.069
edu = prim.	−1.408	−4.029	−4.234	−7.521	0.055	−2.922	−0.871	−4.559
loc = med.	0.576	1.392	0.870	0.136	1.867	0.228	0.906	1.478
loc = big	1.191	1.764	1.100	0.813	−1.254	0.984	−0.279	0.869
loc = rural	−1.260	−2.178	−1.073	−2.959	1.989	−0.426	−1.253	−1.343
employed ratio	0.127	7.350	11.039	17.997	0.982	7.164	4.372	14.042

Conclusion

- It is possible to distinguish three income classes in the most economically well-off federal districts of Russia;
- The least well-off districts tend to have more leptokurtic income density curves;
- The least well-off districts do not provide evidence of the existence of statistically significant middle class groups;
- Higher education, inhabitancy in huge urban areas and higher ratio of employed household members are positive predictors of probability of belonging to a high-income class

HIGHER SCHOOL OF ECONOMICS
NATIONAL RESEARCH UNIVERSITY



Structure of Household
Income Distribution in Russia

Alan Nartikoev
Anatoly Peresetsky