Multilevel Models for Applied Social Research

Instructor: Paul Lambert
Univ. Stirling

Co-facilitator: Kate O'Hara
Univ. Stirling & Univ. Edinburgh

0930-1500hrs, 18/Jun/2025

Github resources:
https://github.com/paul-lambert/SGSSS-2025/

0930-1100	Talk 1: Classical perspectives on multilevel modelling Talk 2: Realistic complexity
1115-1230	Lab 1: Implementing selected popular multilevel models
	Talk 3: Case study on effect scores from random effects residuals
1330-1500	Talk 4: MLMs meet econometrics Lab 2: Responding to complex data and to complex analytical options

(1) Recap: The standard random effects multilevel model

- *…refines a statistical model to take account for structural features of datasets which involve clustering of cases (i) into groups (j) …*
- *…the typical motivation is to achieve optimal coefficient and standard error estimates, and gain appropriate insights into statistical patterns amongst the cluster groups….*

```
mixed gcse lrt girl schl_boys schl_girls
mixed gcse lrt girl schl_boys schl_girl ||school:, reml
```

$Y_i = \beta_p X_{pi} + \varepsilon_i$, $\varepsilon_i \sim N(0, \sigma^2_\varepsilon)$

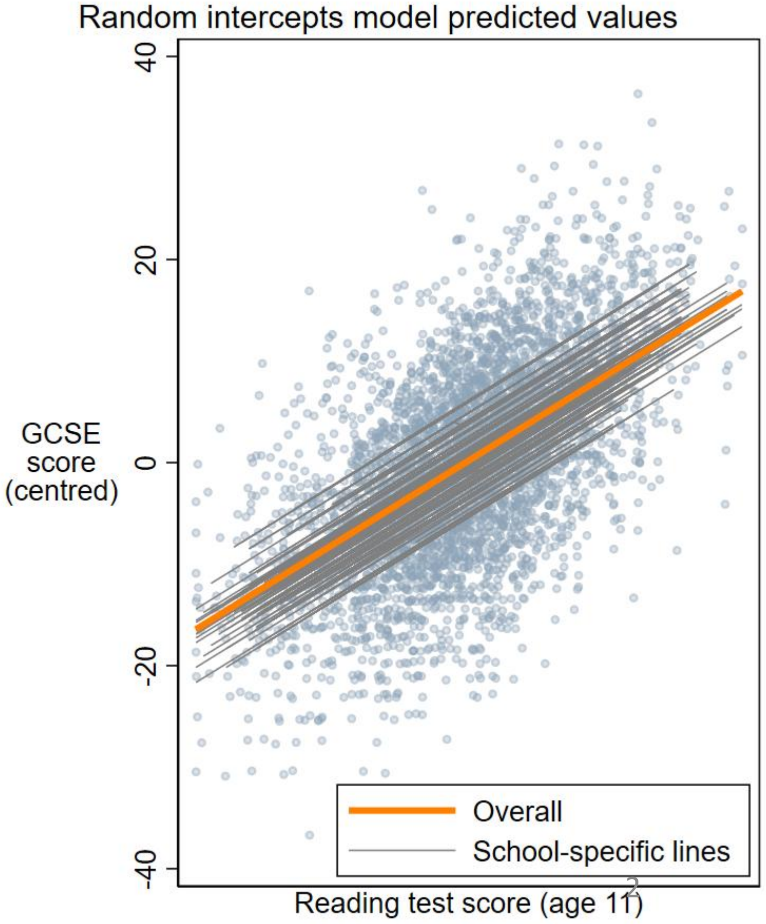
(1) Single level model

$Y_{ij} = \beta_p X_{pij} + \mu_j + \varepsilon_{ij}$,

$\varepsilon_{ij} \sim N(0, \sigma^2_\varepsilon)$

$\mu_j \sim N(0, \sigma^2_\mu)$

(2) Two-level 'random intercepts' multilevel model



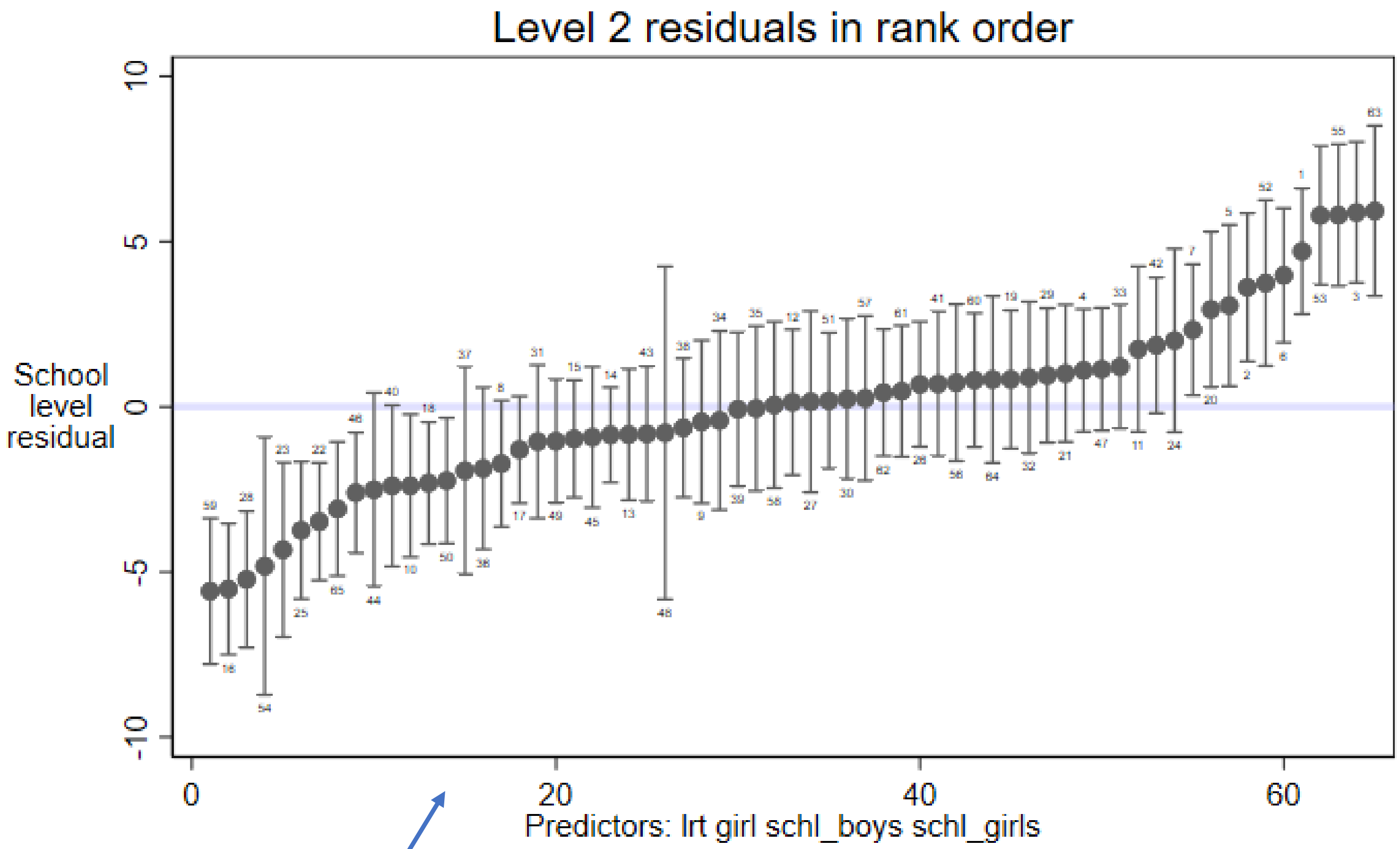
	cons	lrt	girl	schl_boys	schl_girls	N	N_g	ICC	σ_μ	Log-like
(1)	-1.61 (.24)	0.59 (.01)	1.33 (.34)	1.83 (.43)	1.68 (.33)	4059				-14182
(2)	-1.68 (.55)	0.56 (.01)	1.67 (.34)	1.78 (1.13)	1.59 (.89)	4059	65	0.13	2.93	-14011

Regression coefficients (standard errors) & other statistics if predicting exam score at 16 as function of reading test score at age 11, pupil gender, single sex school indicators. Data from Goldstein (2003).

- *Improved model coefficients tell us a fuller story about combination of micro- and macro-level processes*
- *Variation amongst any higher level residuals might be informative about cluster-level processes*

- *Clusters, classically, are ‘too many’, such that it’s not useful to analyse them with dummy variables (‘fixed effects’), but we may want to explore selected cluster-level explanatory variables or inspect model-based cluster-level residuals*

- *In some methodological framings, clusters conceived as a sample from a wider population of clusters*
- *In many applications, clusters are smaller numbers of fixed units (e.g. countries in Europe)*



The popular ‘caterpillar plot’ presentation of cluster level residuals

'Empirical Bayes' residuals..?

The **'Empirical Bayes' residuals** for cluster-based random effects are a 'shrunk' adjustment to the arithmetic residuals for the cluster, with 'shrinkage' towards the population level pattern

- **$EB(\mu_j) = \lambda_j \mu_j$,** where **$\lambda_j = (\sigma^2_\mu / (\sigma^2_\mu + (\sigma^2_\varepsilon / n_j)))$**
- **$EB(\beta_{pj}) = \lambda_j \beta_{pj} + (1 - \lambda_j)\beta_p$.** (where '$\beta_{pj}$' represents $\beta_p + \mu_{pj}$)

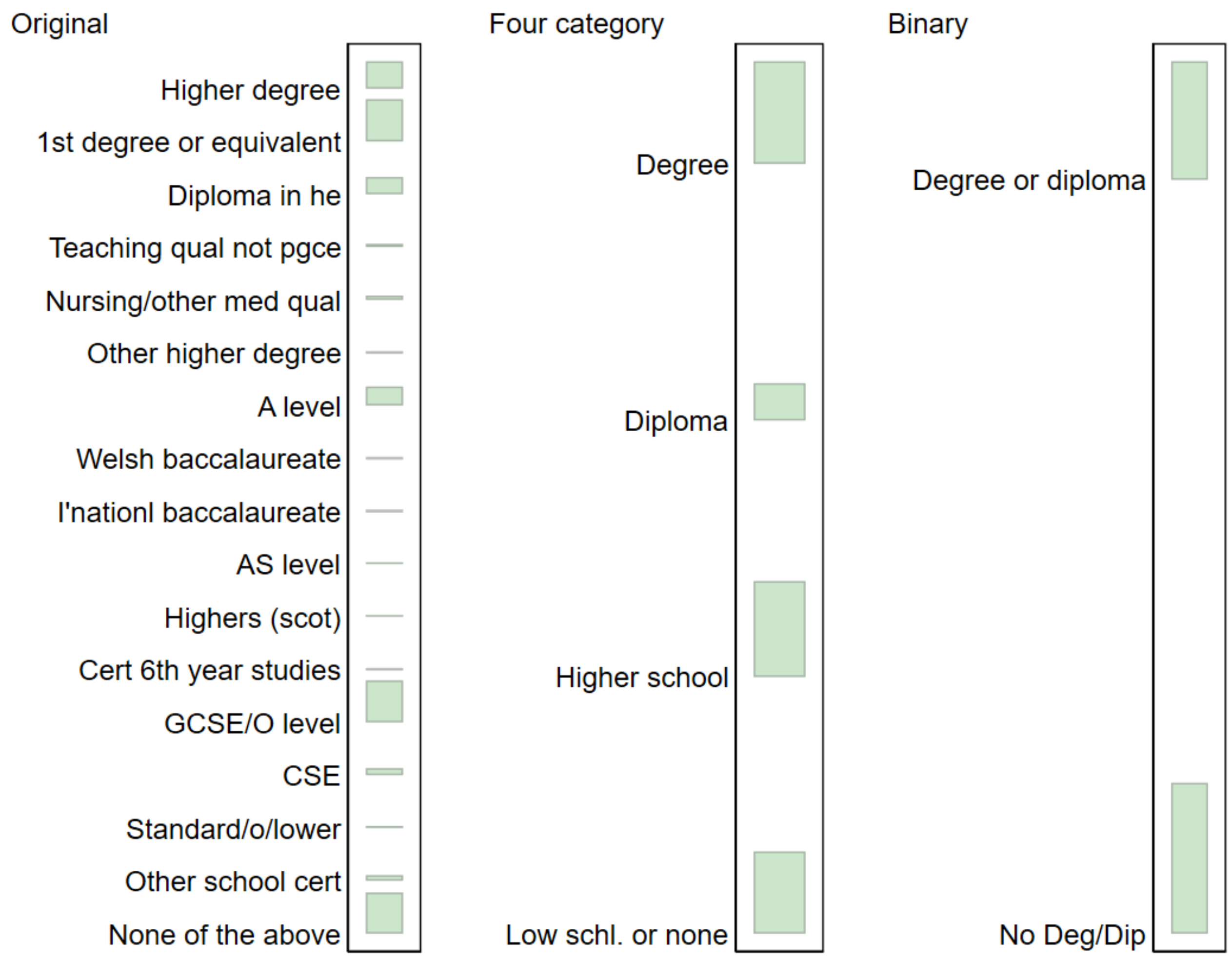
- 'Shrinkage factor' or 'reliability' **λ_j** deflates the impact of cluster specific μ_j when n_j (cluster size) is smaller, and impacts much less when n_j gets larger
- 'Shrinkage' means that $EB(\mu_j)$ is generally a more compelling estimate for the net distinctiveness of cluster j than is its arithmetic value μ_j

- *Research in practice*

'Fixed effects residuals' (the arithmetic average perturbations based on the cases representing the cluster) are often reported when Empirical Bayes random effects residuals would arguably be more appropriate

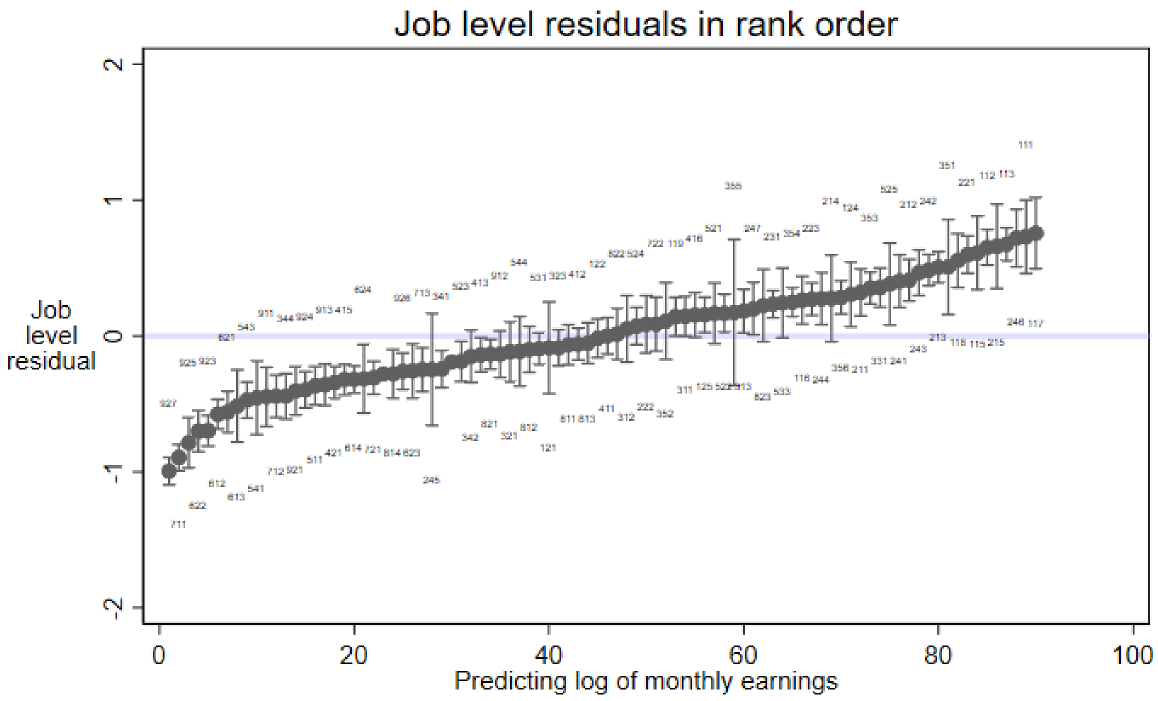
2) Things we could do with categorical data…

- **Routine practice in sociology/education is to simplify a taxonomy considerably, in order to construct a smaller set of contrasts**
- A compelling alternative could be to construct metric dimension score(s) to represent the differences between categories ('scaling categories')
  - Use some information about finer-grained details
  - Easier to fit interaction terms and/or context-based standardisations
  - Might aid theoretical interpretation of differences
  - *Politicised? Americans & Dutch will scale, but Europeans categorise!*
- A 'fixed effects' scaling estimation makes point estimates from empirical data
  - Arithmetic approaches such as 'Effect proportional scaling'
  - Data reduction tools such as correspondence analysis
- A 'random effects' scaling estimation also gives point estimates, but conceives of them as the realisation of random distributions around an average
  - Empirical Bayes residual as weighted average of sample-specific estimate and overall average pattern
  - Mills et al (2007) example for job quality scale



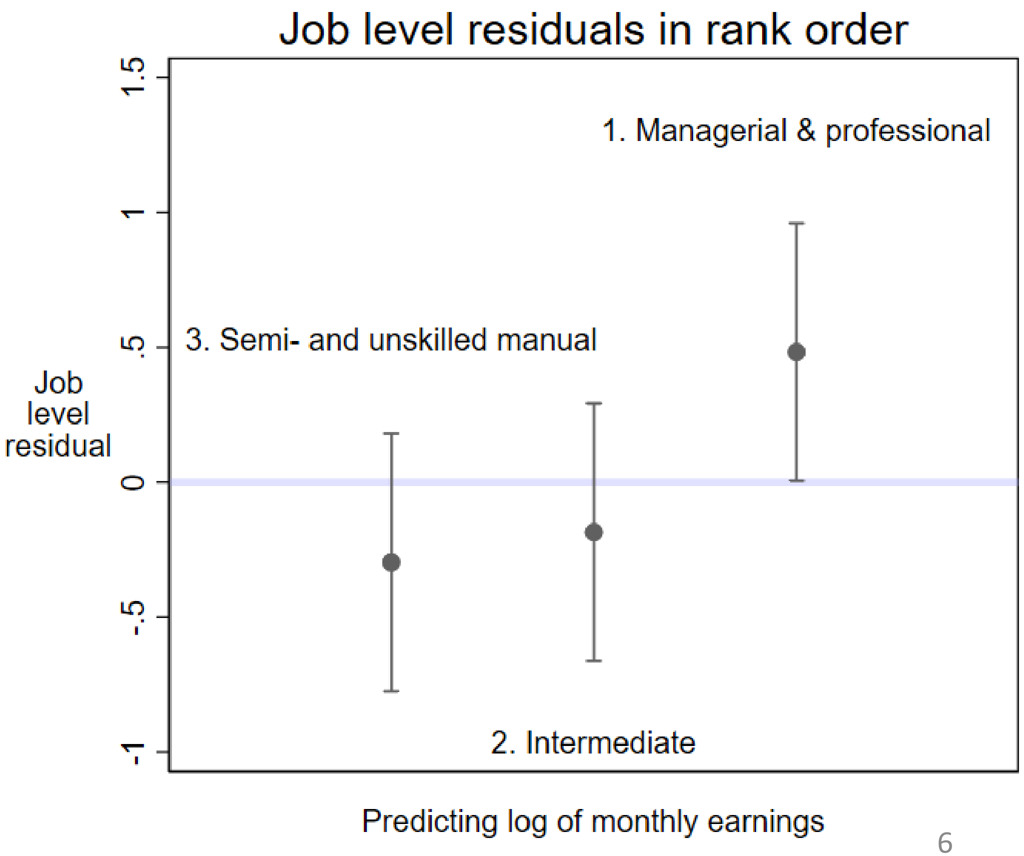
UKHLS wave M: Distribution of highest educational qualifications amongst 25-80 year olds.

Can use random effects ('ESRES') approach for few or many categories...



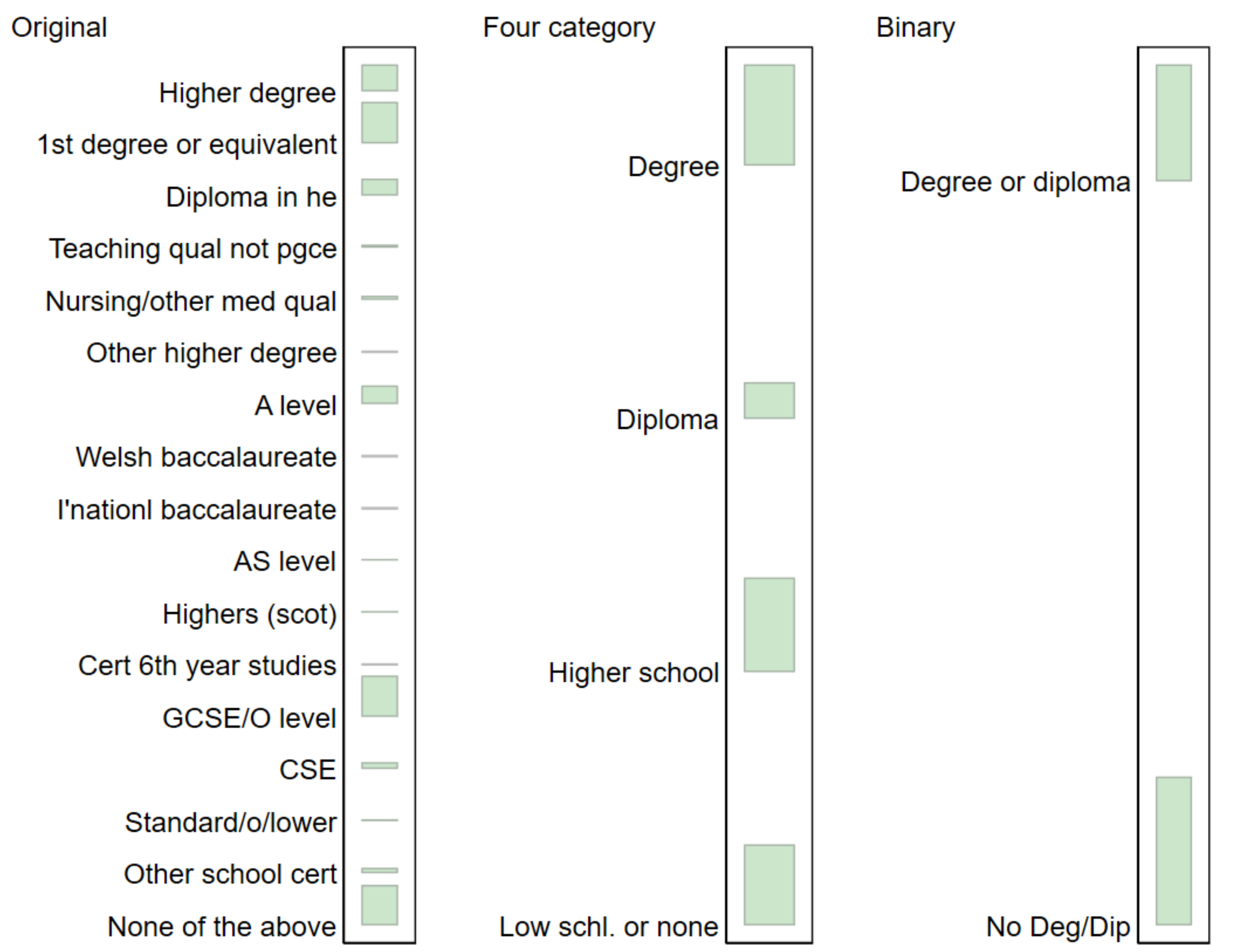
- Many categories
  - difference between EPS and ESRES concentrated on smaller categories and use of control variables; more space to explore and disentangle multiple dimensions; more opportunity to improve upon dummy variable treatments
  - 'Many categories' includes the interaction of several different categorical measures (e.g. academic + vocational training categories)

- Few categories
  - gain insight from scores particularly if using control vars; difference between EPS and ESRES minimal; dummy variable treatment remains more tractable



3) Example: Education scores..

```
* 'Effect proportional scaling'.
egen ln_mean=mean(ln_pay) if qfhigh2 > 0 & dvage > 25 & dvage <= 80, by(qfhigh2)
replace ln_mean = exp(ln_mean)
```



UKHLS wave M: Distribution of highest educational qualifications amongst 25-80 year olds.

```
* 'Effect scores for random effects residuals'.
mixed ln_pay if qfhigh2 > 0 & dvage > 25 & dvage <= 80 || qfhigh2:, reml
predict ln_ref if qfhigh2 > 0 & dvage > 25 & dvage <= 80, reffects
replace ln_ref=ln_ref + _b[_cons]
replace ln_ref = exp(ln_ref)
```

EPS	ESRES	N
2467	2457	2806
2031	2028	4352
1675	1673	1761
1759	1744	341
1548	1549	440
2146	1901	44
1538	1538	1923
1670	1602	20
1547	1550	14
1390	1412	159
1476	1486	167
1579	1571	58
1341	1342	4210
1181	1193	688
1413	1439	205
1127	1143	562
1246	1248	4192
1650	1649	21950

i.e. scale scores for education categories (here based on geometric means of earnings)

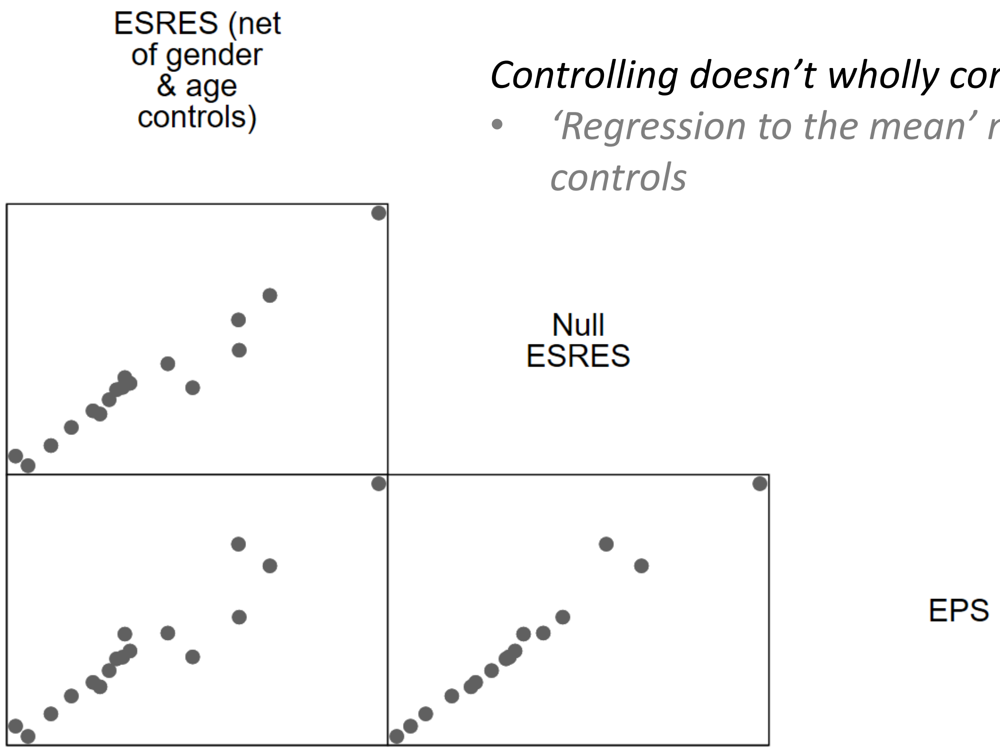
An ‘ESRES’ measure might also be calculated net of control variables…

EPS	ESRES (null)	ESRES (controls)	N
2467	2457	2666	2806
2031	2028	2252	4352
1675	1673	1863	1761
1759	1744	2135	341
1548	1549	1958	440
2146	1901	2132	44
1538	1538	1669	1923
1670	1602	1700	20
1547	1550	1692	14
1390	1412	1606	159
1476	1486	1640	167
1579	1571	1720	58
1341	1342	1497	4210
1181	1193	1285	688
1413	1439	1579	205
1127	1143	1332	562
1246	1248	1419	4192
1650	1649	1835	21950

```
mixed ln_pay fem cen_age cen_age2 age_fem age2_fem ||qfhigh2:, reml
predict ln_ref2 if qfhigh2 > 0 & dvage > 25 & dvage <= 80, reffects /* qual specific resid */
replace ln_ref2=ln_ref2 + _b[_cons]  /* qual specific predicted value (for male aged 30) */
replace ln_ref2 = exp(ln_ref2)
```

(Choices over the scaling scale)

- *here we’ve shown predicted values at the intercept, which are arguably easier to interpret*
- *mean standardised disturbances are probably more consist/comparable*



Controlling doesn’t wholly control, though...

- *‘Regression to the mean’ mollifies impact of controls*

Example: Logistic regression coefficients predicting if respondent reports gambling in the last week (UKHLS wave M)

(via lottery, scratchcards, private betting, spread betting, betting exchange, football pools)

- The scaled education scores are a robust and parsimonious depiction of education pattern (a higher education level reduces gambling, and dampens the age pattern in gambling)

- 'ln_ref{2}' are mean-standardised ESRES scales for 'educational advantage'. 'esres1' uses scale from model with no controls, 'esres2' use gender and age controls; 'a' uses main effects only, 'b' allows interaction with linear component of age effect.
- Some argue (e.g. Snijders & Bosker 2012) that models with scale scores should also include random effects for the categories (not used here)

Variable	base	esres1a	esres2a	esres1b	esres2b
fem	-.4678***	-.4771***	-.4724***	-.4818***	-.4764***
cen_age	.1087***	.1055***	.1051***	.1119***	.111***
cen_age2	-.000988***	-.000988***	-.000983***	-.001051***	-.001041***
ln_ref		-.2127***		-.1151***	
ln_ref2			-.2117***		-.1171***
c.ln_ref# c.cen_age				-.004345***	
c.ln_ref2# c.cen_age					-.004175***
_cons	-1.014***	-.9414***	-.9457***	-.9657***	-.9685***
N	22682	22682	22682	22682	22682
ll	-13609	-13513	-13513	-13506	-13506
bic	27258	27076	27076	27072	27073
r2_p	.0168	.02375	.02375	.02426	.02422

Legend: * p<0.05; ** p<0.01; *** p<0.001

(Summary statistics confirm parsimony compared to models using 2, 4 or 19 dummy variables)

	base	esres1a	esres1b	esres2a	esres2b	eps1a	eps1b	Bin-a	Bin-b	Quart-a	Quart-b	Full-a	Full-b
BIC	27258	27076	27072	27076	27073	27077	27072	27139	27134	27106	27100	27202	27307
$R2*10^3$	17	24	24	24	24	24	24	21	22	23	25	25	27

4) Example: Using ESRES in CAMSIS scaling

[author's work in progress, cf. Lambert & Griffiths 2018]

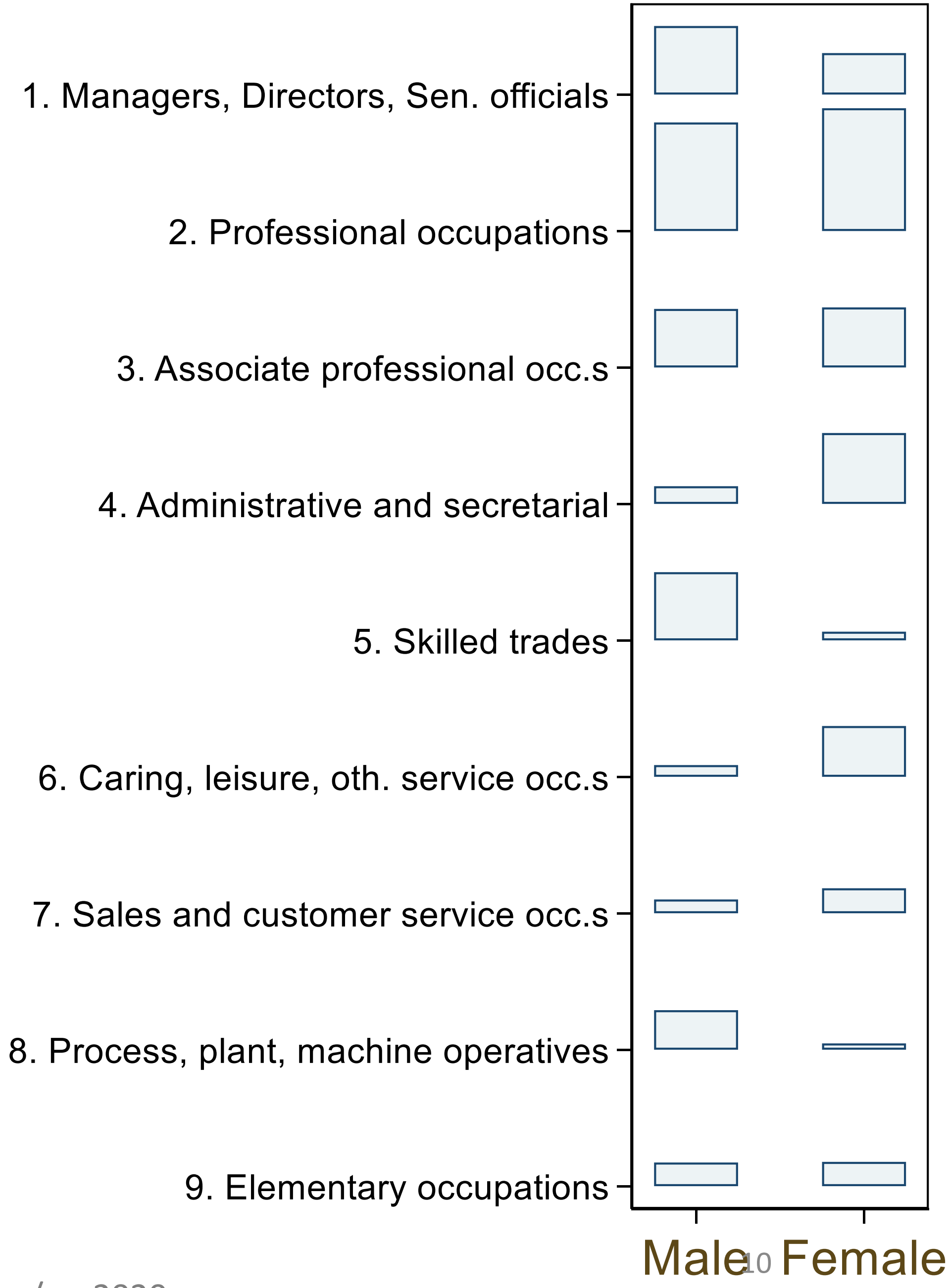
- SOC2020 used on UK surveys from 2021 (3- or 4-digit version)
- Taxonomy of 412 4-digit units (valid range 1111-9269)
- Small modifications from SOC2000 & SOC2010 (more cases in major gp2)

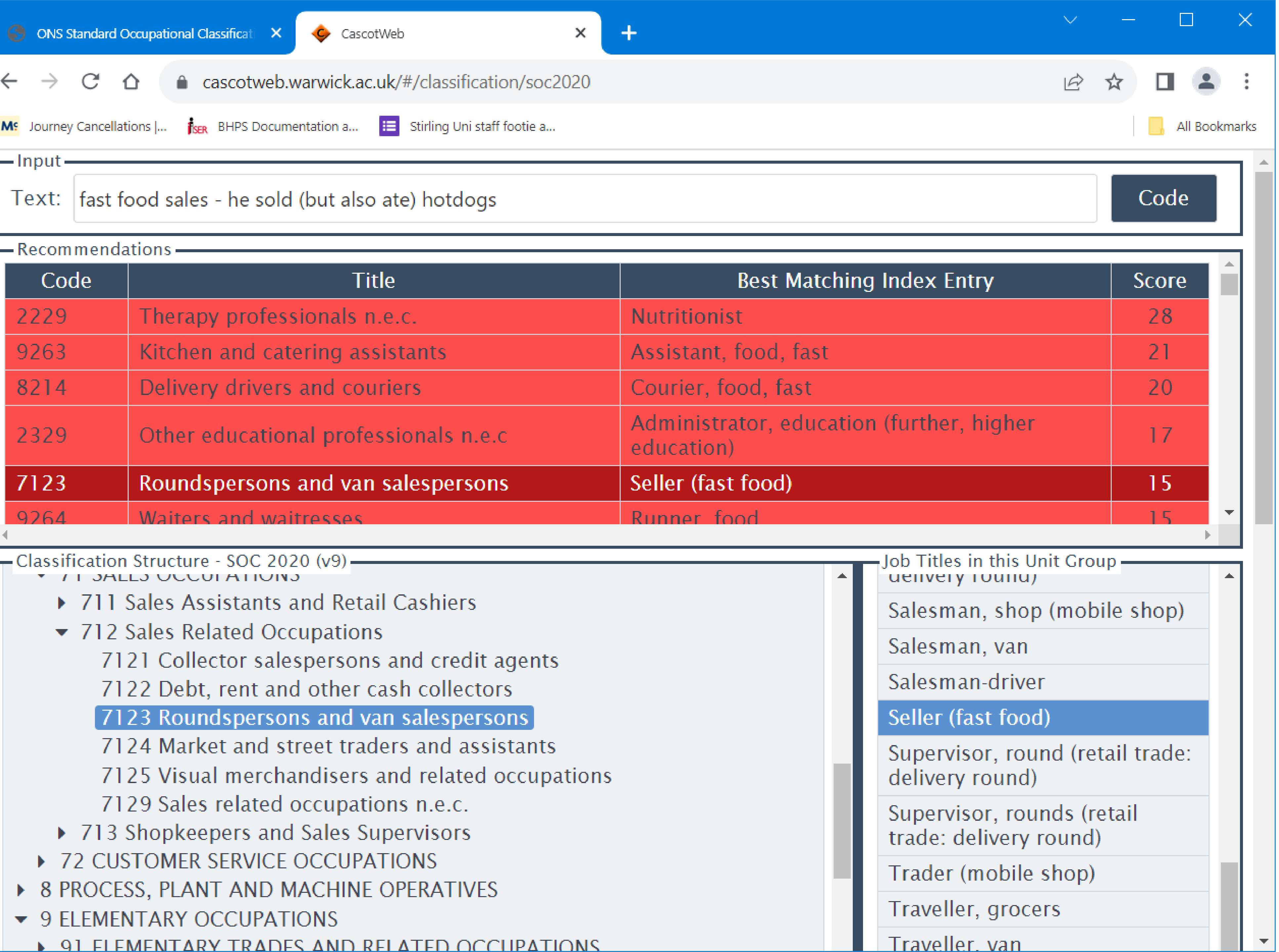
	SOC2020 Major Group	SOC2020 Sub-Major Group	SOC2020 Minor Group	SOC2020 Unit Group	SOC2020 Group Title	Ch
172				2254	Medical radiographers	
173				2255	Paramedics	
174				2256	Podiatrists	
175				2259	Other health professionals n.e.c.	
176						
177		23			**TEACHING AND OTHER EDUCATIONAL PROFESSIONALS**	
178						
179			231		**Teaching Professionals**	
180						
181				2311	Higher education teaching professionals	
182				2312	Further education teaching professionals	
183				2313	Secondary education teaching professionals	
184				2314	Primary education teaching professionals	
185				2315	Nursery education teaching professionals	
186				2316	Special and additional needs education teaching professionals	Un tea
187				2317	Teachers of English as a foreign language	
188				2319	Teaching professionals n.e.c.	
189						
190			232		**Other Educational Professionals**	
191						
192				2321	Head teachers and principals	

SOC2020 descriptions | SOC2020 Framework

https://warwick.ac.uk/fac/soc/ier/software/cascot/choose_classificatio/
https://www.ons.gov.uk/methodology/classificationsandstandards/standardoccupationalclassificationsoc/soc2020

SOC 2020, 2021-23





Occupational information reminders...

- Occupational unit group codes calculated by text coding
- Designed for 4-digit detail though many analysts use 3-, 2- or 1-digit versions
- Funded surveys also collect employment status data (self-employed with lots, few or no employees, manager of lots or of few, supervisor, employee) which may or may not feature in social classifications
- Metadata available to link SOC codes to stratification measures

A CAMSIS version incorporating additional refinements

- **www.camsis.stir.ac.uk/downloads/gb/gb_soc2020_v1.dta**

➢ **Uses diagonals and pseudo-diagonals, but with more specification than previously common**

Blanks out diagonals only in major groups 1 & 2, all H 1 to W 4 major groups, and psds for farmers, medics, catering, journalism, police, arts & housekeeping

➢ **Uses a 'smoothing' strategy using random effects residuals to 'shrink' small group estimates towards the mean:**

ca hocc wocc if psd4==0
predict hscores, rowscore(1)
predict wscores, colscore(1)
mixed wscores if psd4b==0 ||hocc:,
predict hscores2, reffects
mixed hscores if psd4b==0 ||wocc:,
predict wscores2, reffects

➢ I also experimented with a weighting approach in which recommended scores were weighted averages of derived scores at different levels of occ. detail, but after exploration this didn't lead to compelling improvements in measures

	soc2020	mcamsis	fcamsis	isei	siops	icam
173.	1251. Property, housing and estate managers	64.08	63.21	62.39	49	59.95
174.	1252. Garage managers and proprietors	40.29	39.68	51.01	47	56.18
175.	1253. Hairdressing and beauty salon managers and proprietors	56.68	46.84	51.01	47	56.18
176.	1254. Waste disposal and environmental services managers	47.44	50.5	51.01	47	56.18
177.	1255. Managers and directors in the creative industries	73.39	79.12	65.01	75	60.13
178.	1256. Betting shop and gambling establishment managers	56.68	47.85	51.01	47	56.18
179.	1257. Hire services managers and proprietors	47.23	47.85	51.01	47	56.18
180.	1258. Directors in consultancy services	77.41	81.62	51.01	47	56.18
181.	1259. Managers and proprietors in other services n.e.c.	56.68	47.85	51.01	47	56.18
182.	2111. Chemical scientists	70.54	71.42	83.5	69	80.22
183.	2112. Biological scientists	69.74	71.42	80.46	62.66	68.98
184.	2113. Biochemists and biomedical scientists	77.52	71.42	80.46	62.66	68.98
185.	2114. Physical scientists	76.9	83.2	86.81	67	80.22
186.	2115. Social and humanities scientists	73	74.43	83.09	68.51	76.83
187.	2119. Natural and social science professionals n.e.c.	74.88	71.42	80.46	62.66	68.98
188.	2121. Civil engineers	60.38	59.75	81.4	70	73
189.	2122. Mechanical engineers	60.03	59.75	77.1	66	73
190.	2123. Electrical engineers	58.83	59.75	80.78	65	73
191.	2124. Electronics engineers	55.29	59.75	80.75	65	73
192.	2125. Production and process engineers	44.6	59.75	79.05	54	73
193.	2126. Aerospace engineers	61.58	59.75	77.1	66	73
194.	2127. Engineering project managers and project engineers	59.74	59.75	78.69	55	73
195.	2129. Engineering professionals n.e.c.	52.24	59.75	78.69	55	73
196.	2131. IT project managers	55.56	56.34	78.86	61.15	67.11
197.	2132. IT managers	58.06	59.96	78.86	61.15	67.11
198.	2133. IT business analysts, architects and systems designers	61.69	59.44	74.66	51	75.39
199.	2134. Programmers and software development professionals	64.06	70.97	74.66	51	75.39
200.	2135. Cyber security professionals	62.08	67.14	75.13	51	75.39
201.	2136. IT quality and testing professionals	58.39	67.14	74.7	51	75.39

A CAMSIS version incorporating additional refinements

- **www.camsis.stir.ac.uk/downloads/gb/gb_soc2020_v1.dta**

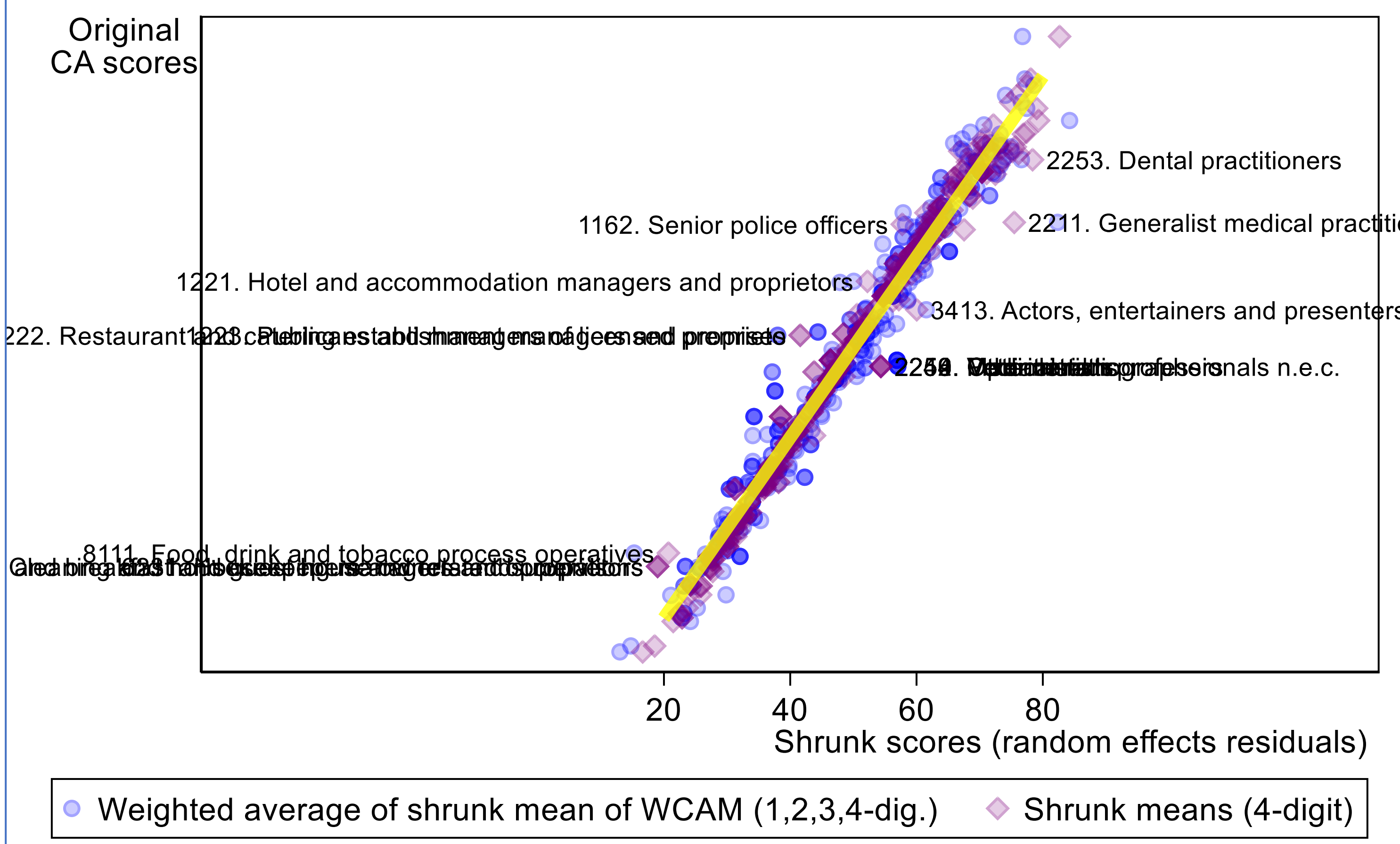
- **Uses diagonals and pseudo-diagonals, but with more specification than previously common**

  Blanks out diagonals only in major groups 1 & 2, all H 1 to W 4 major groups, and psds for farmers, medics, catering, journalism, police, arts & housekeeping

- **Uses a 'smoothing' strategy using random effects residuals to 'shrink' small group estimates towards the mean:**

  ca hocc wocc if psd4==0
  predict hscores, rowscore(1)
  predict wscores, colscore(1)
  mixed wscores if psd4b==0 ||hocc:,
  predict hscores2, reffects
  mixed hscores if psd4b==0 ||wocc:,
  predict wscores2, reffects

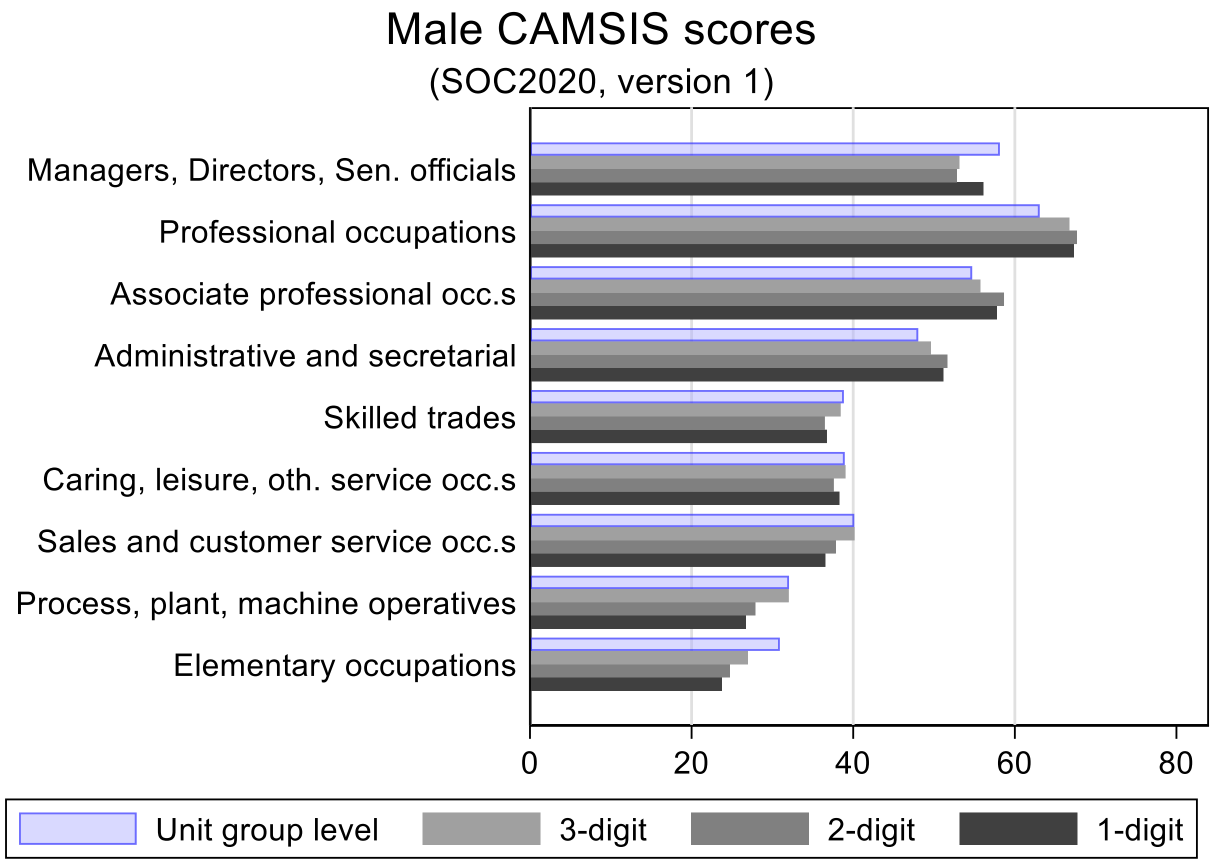
- I also experimented with a weighting approach in which recommended scores were weighted averages of derived scores at different levels of occ. detail, but after exploration this didn't lead to compelling improvements in measures

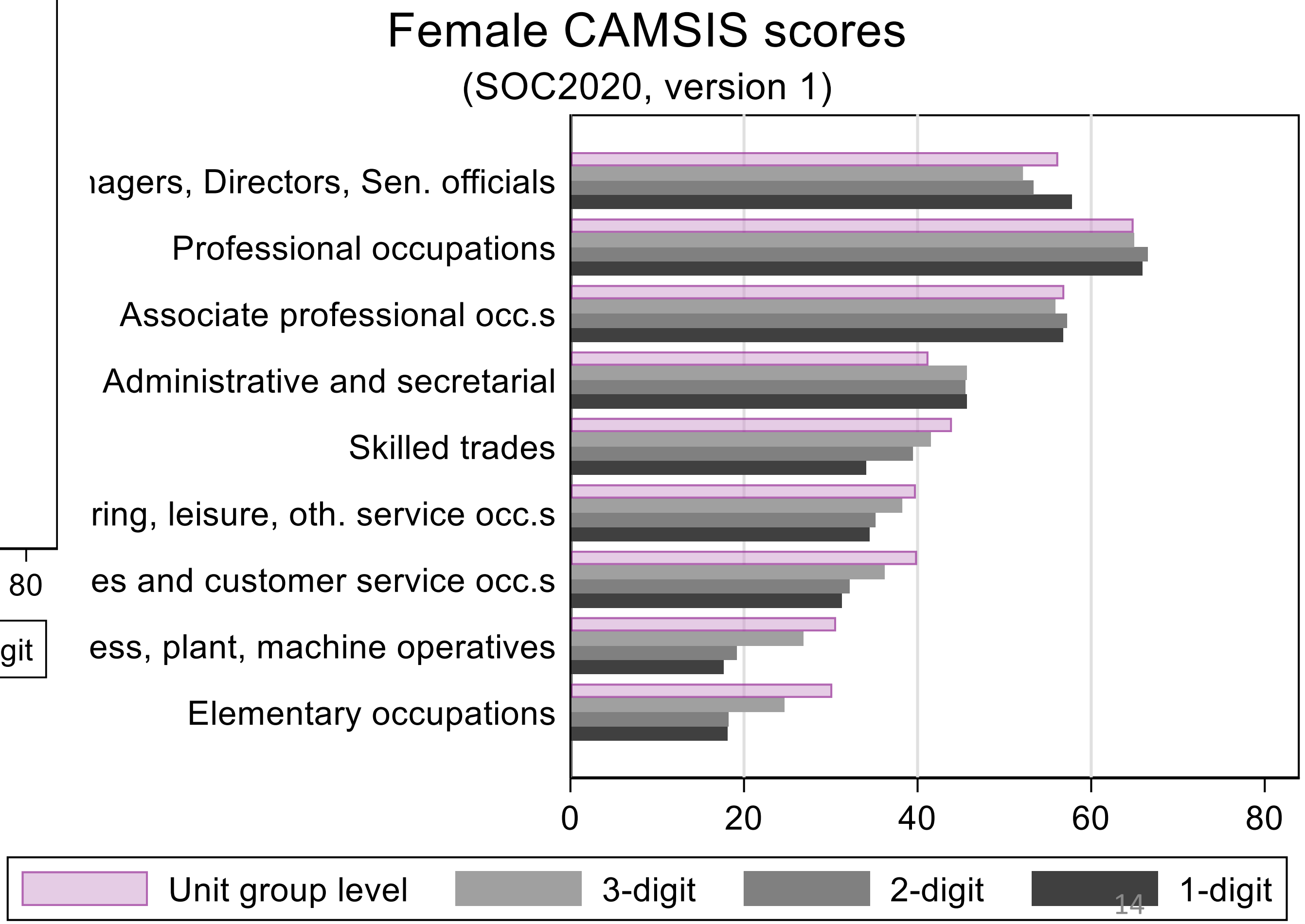


Alternative CAMSIS scalings from LFS 2021-3. Labels shown if gap between original and (2) exceedd 0.33SD

A CAMSIS version incorporating additional refinements

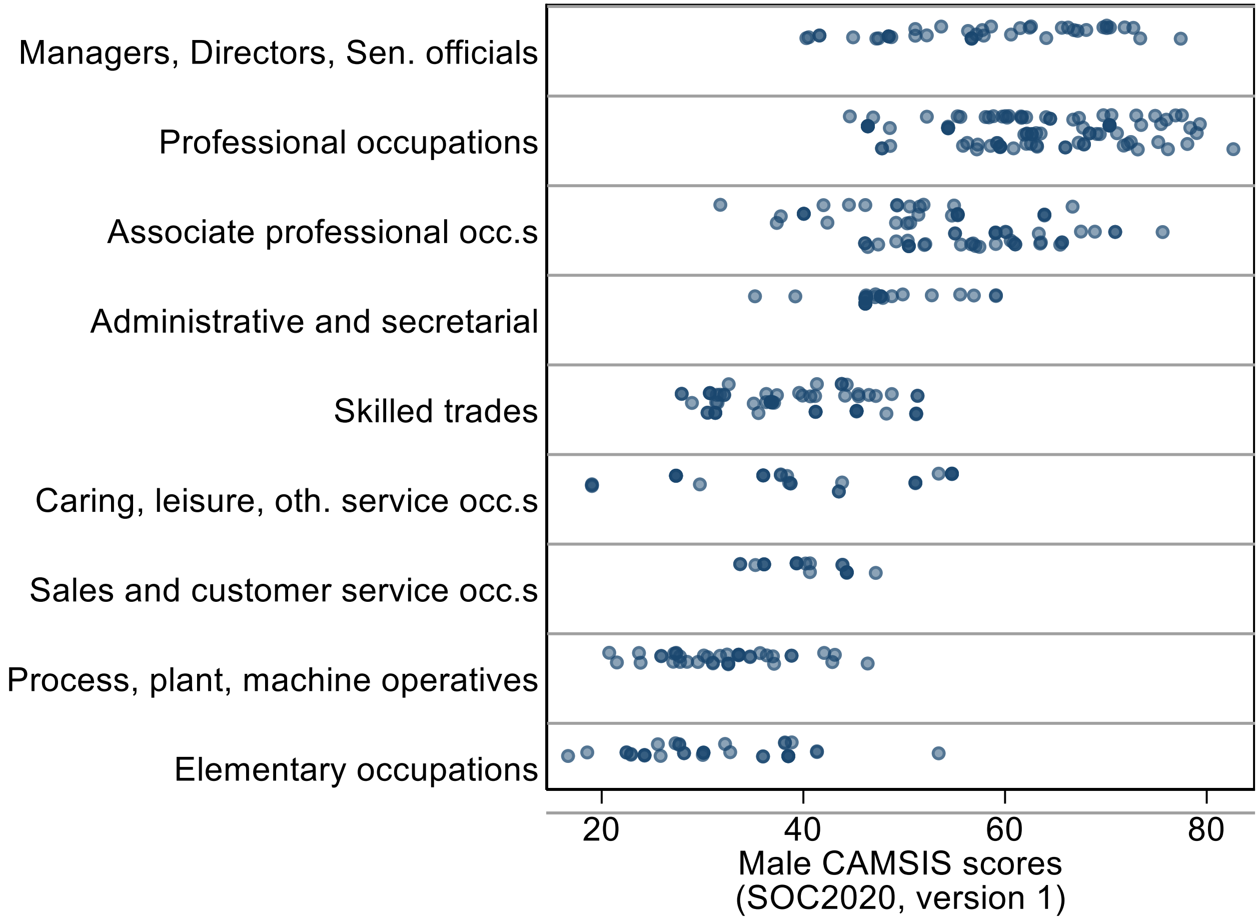
- **www.camsis.stir.ac.uk/downloads/gb/gb_soc2020_v1.dta**



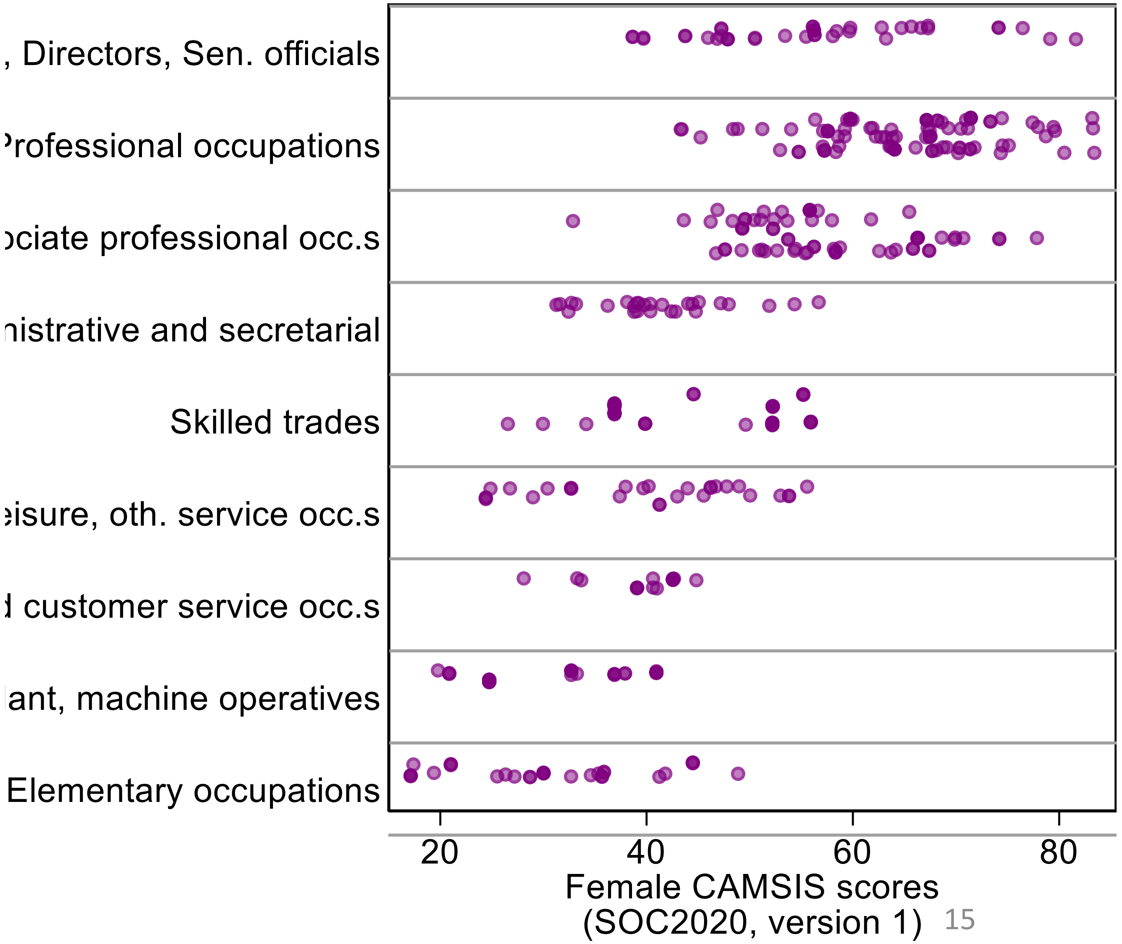


A CAMSIS version incorporating additional refinements

- **www.camsis.stir.ac.uk/downloads/gb/gb_soc2020_v1.dta**

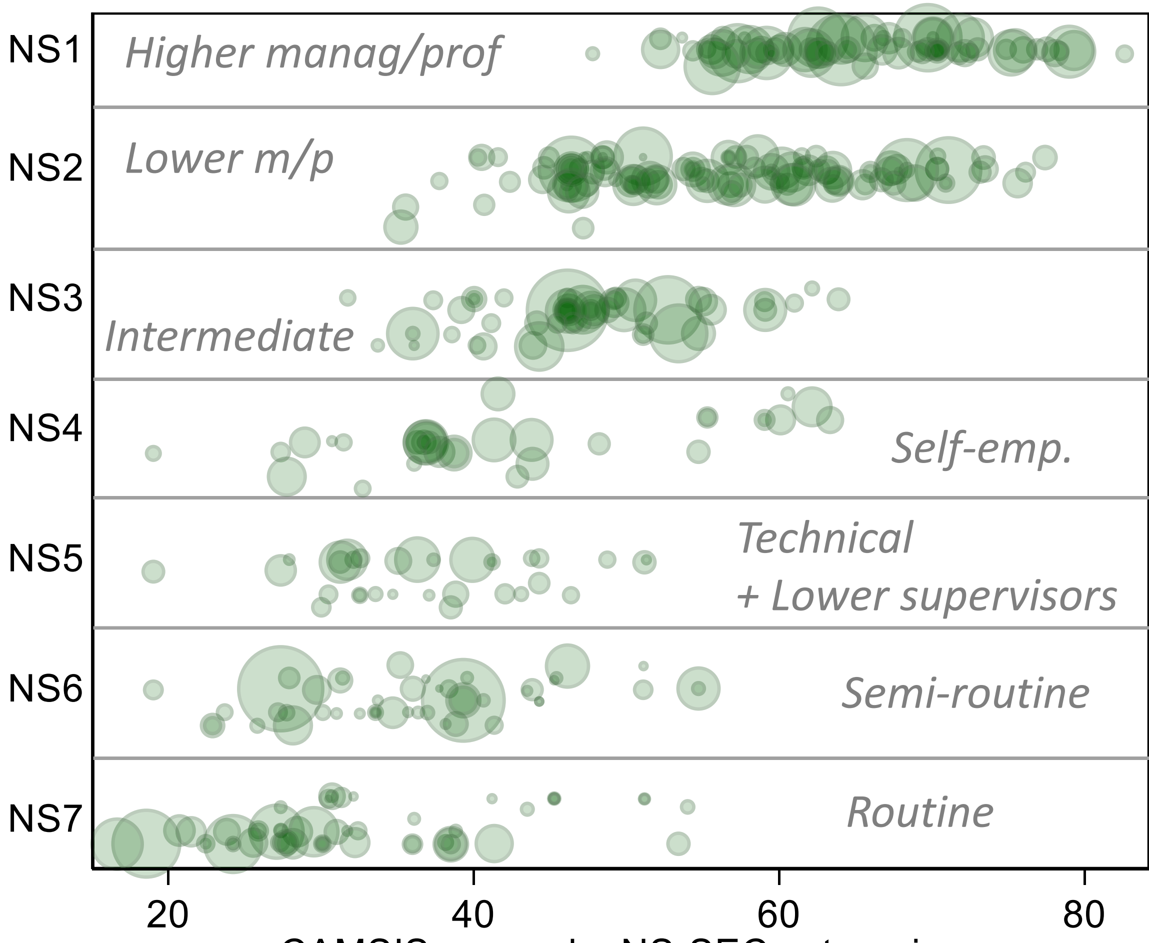


(Scale scores allow use to use heterogeneity in occupations, albeit with cautious 'smoothing')



- *Outliers are generally smaller occupations, but smoothed patterns aren't implausible. This image represents heterogeneity within 'big class' categories.*

	NS1	NS2	NS3	NS4	NS5	NS6	NS7
	*(100*correlation between MCAMSIS and measure)*						
Social housing	6	8	6	13	12	9	10
Health prob.s	0	2	3	0	1	1	1
Disability	1	1	4	1	3	1	3
Log hrly. pay	9	13	0	30	32	10	4
Educ. Quals	27	32	9	41	10	9	19
Educ. lvg. age	16	13	2	20	4	7	10
N for (1)	5845	6732	3818	1873	1179	2395	2590



CAMSIS scores by NS-SEC categories
(weighted by volume of adults in LFS, AJ2023)
(Vertical offsets within NSSEC are based on numeric SOC2020 code)

- *Correlations observed between MCAMSIS and measure within NS-SEC categories (i.e. heterogeneity within NS-SEC)*

5) Summary: When might 'effect scores from random effects residuals' (ESRES's) be useful?

Points in favour…

- Parsimony
  - Coherent interpretation
  - Increases chances of suitably complex specifications, such as interactions, context-based standardisations, and multi-process specifications
- Certain optimal scenarios
  - Large numbers of categories
  - Robust estimates for small groups
  - Unknown or ambiguous underlying structures
- Amenable to extensions
  - Scores in multiple dimensions/outcomes
  - Scores for random intercepts or slopes
  - Analysis using scores could use multiple simulations from range of plausible scores

But on the other hand…

- Scenarios when impact may be muted
  - Smallish number of categories that are fairly well understood and/or well-represented in data
  - Pertaining to measures of limited importance to the system being studied
  - Multicategory specifications likely to support slightly better fit to the data
  - Effect Proportional Scaling is probably just the same, unless sparse groups matter
- Lack of existing applications
  - Layer of abstraction
  - Limited awareness
  - Pressure to explain/document
  - Wild west of derived measures

=> Compelling case for more methodological attention to ESRES's, and more efforts at implementation…

References cited

- Goldstein, H. (2003). *Multilevel Statistical Models, Third Edition. London: Arnold.*
- Lambert, P. S., & Griffiths, D. (2018). *Social Inequalities and Occupational Stratification: Methods and Concepts in the Analysis of Social Distance. Basingstoke: Palgrave*
- Mills, C. (2007). Unequal jobs: Job quality and job satisfaction. In P. McGovern, S. Hill, C. Mills, & M. White (Eds.), *Market, Class and Employment (pp. 233-282). Oxford: Oxford University Press.*
- Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling, 2nd Edition. London: Sage.*
- University of Essex, & Institute for Social and Economic Research. (2023). *Understanding Society: Waves 1-13, 2009-2022 and Harmonised BHPS: Waves 1-18, 1991-2009. [data collection]. 18th Edition. Colchester: UK Data Service, SN: 6614, DOI: http://doi.org/10.5255/UKDA-SN-6614-19.*