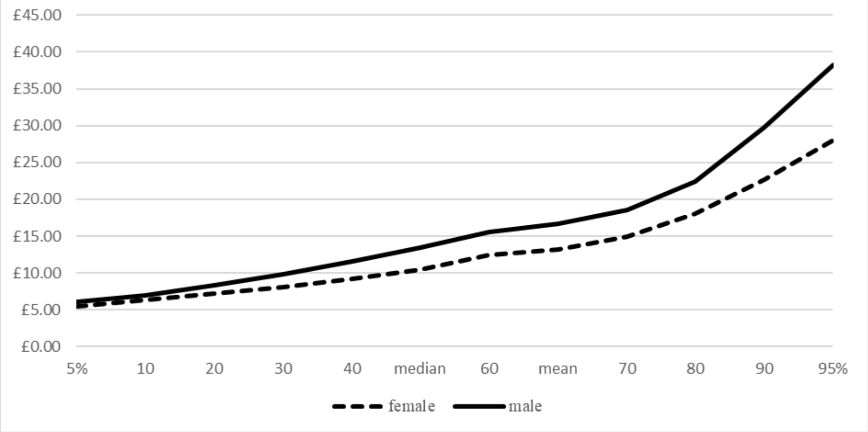
Social statistics away-day May 2024

PAY GAP REPRESENTATION

By Wendy Olsen with thanks to Myong Sook Kim

Figure 1. Point Estimates of Gross Hourly Pay accross the Distribution of Wages, by Sex



A PAY GAP COULD BE AT THE MEAN OR THE MEDIAN.

It is inherently an aggregate statistic.

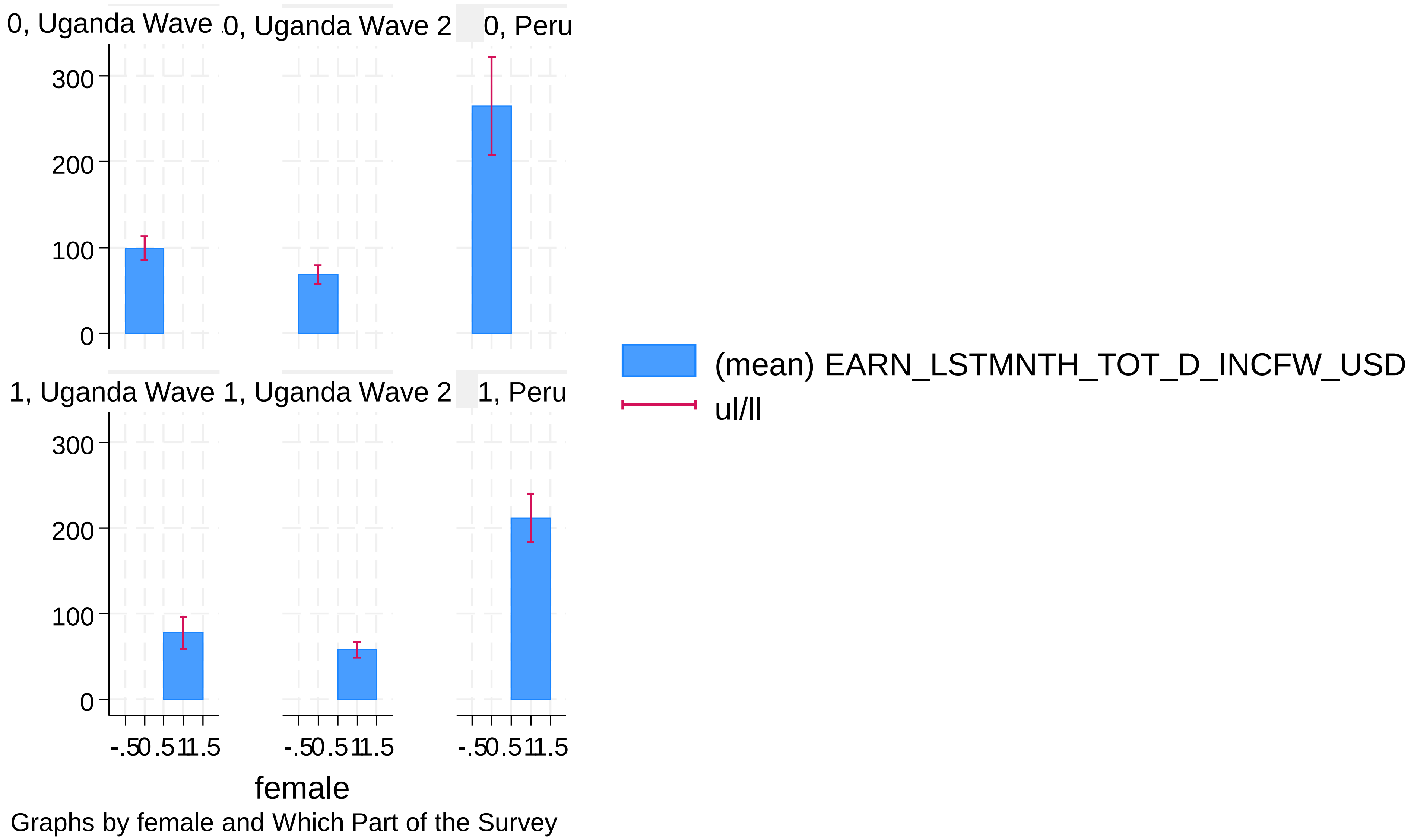
characteristic or coefficient effects. In contrast to the underlying assumptions of st
Oaxaca models, simulation assumes integrated labour markets and specifies a pooled
across the sexes. Here the gender pay gap is calculated as the average difference betwee
and female characteristics as a function of an *undifferentiated slope* (Eq. B).

$$ln\ y_f - ln\ y_m = \sum(\bar{X}_f - \bar{X}_m)\beta_{overall} + (Sex)\beta_{overall} + (\varepsilon_f - \varepsilon_m) \qquad \text{(Eq. B)}$$

In its pooled estimates of all workers, simulation allows us to include sex *as a me*
covariate and we conceptualise sex as a 'measured residual'. Unlike Oaxaca, ho
simulation permits us to examine *unobserved* residual error $(\varepsilon_f - \varepsilon_m)$ independent
observed measure of sex.

Example: male mean vs female mean, omitting the 0's (non-earners). The distribution of earnings that we use is often "usual" earnings per hour, in the main job.

- This saves confusion and invalid comparisons
- We use the logarithm of pay if everyone has a job
- "Actual earnings" is considered more accurate.



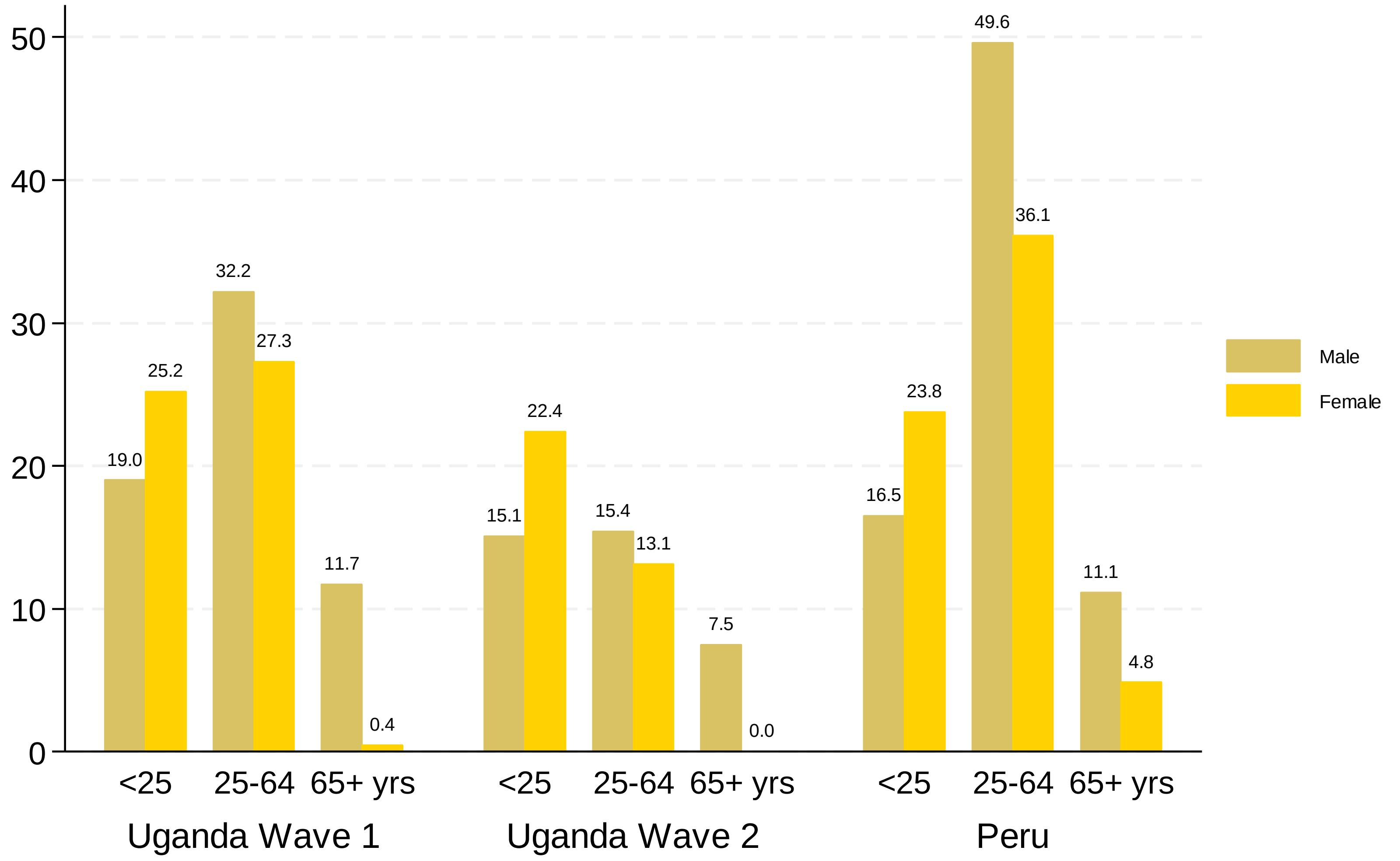
The second panel is the Females
We have waves 1 and 2 in Uganda
At right, Peru has just one wave.

```
))  (rcap ul ll female, by(female wave ))
*Here we can also use the pay median*
graph export "results\comparemeanearnsBysexInwave.wmf",
replace
```

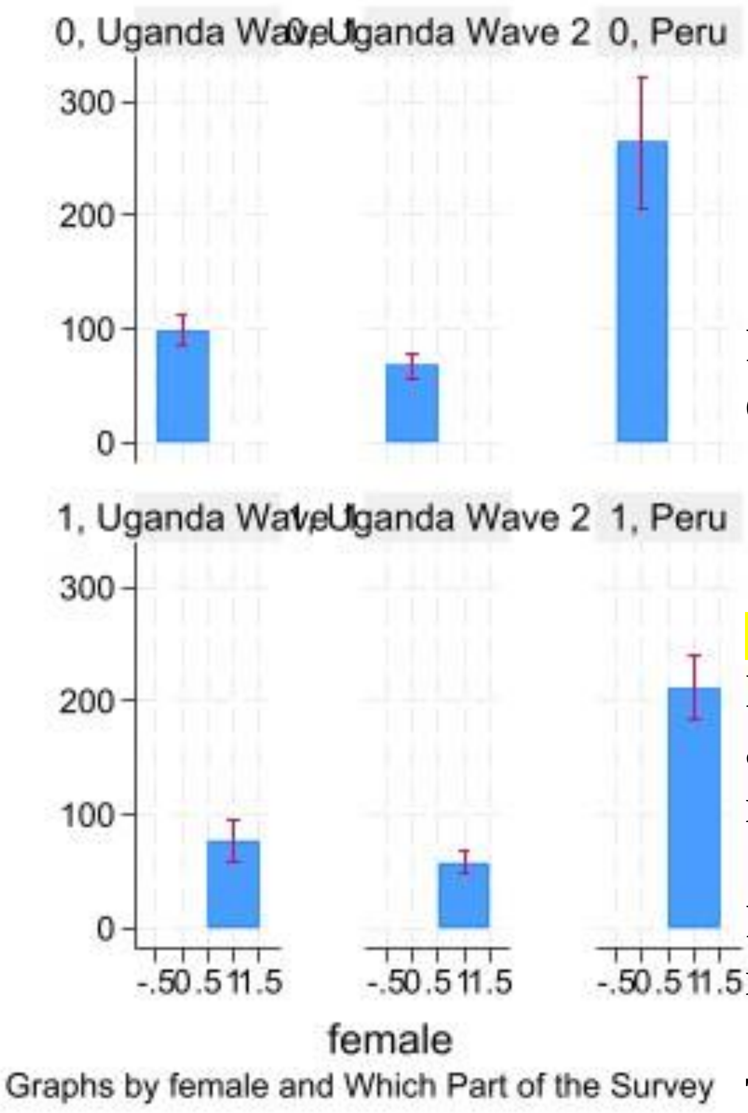
HOW TO ADJUST COLOR AND LABELS ON BARS

```
graph bar
EARN_LSTMNTH_TOT_D_INCFW_USD
[pweight=WEIGHT_FINAL]  if
inlist(ICSE18_MJJ,3,4,5) , over(sex)
over(agegroup3) over(wave) bar(1,
color(sand)) bar(2,color(gold))
legend(size(vsmall))  blabel(bar,
format(%9.1f) size(2))
ytitle("USD/Month, Females Yellow,
Males Darker")
```

So after you collapse, you will need to adjust the twoway graph to use a **categorical variable**, so that **over** can refer to **Bar 1, Bar 2**, etc. This is awkward in twoway grapsh but it can be manipulated.



50
40
30
20
10
0
19.0
25.2
32.2
27.3
11.7
0.4
15.1
22.4
15.4
13.1
7.5
0.0
16.5
23.8
49.6
36.1
11.1
4.8
<25
25-64
65+ yrs
Uganda Wave 1
Uganda Wave 2
Peru
Male
Female



Source: International Labour Organisation

Primary survey data, experimental research on questionnaire design, 2022.

TRICKSY NOT AN ADEQUATE PAYGAP TILL PER-HOUR

Here, pay is per month, estimated from the last paypacket, among only those who had pay. It refers only to adults in the main job.

Furthermore it includes the in-kind pay, converted by respondent to local currency.

The currency is US $ per person per month.

The pay is per month, estimated from the last paypacket, among only those who had pay. It refers only to adults in the main job.

Furthermore in includes the in-kind pay, converted by respondent to local currency.

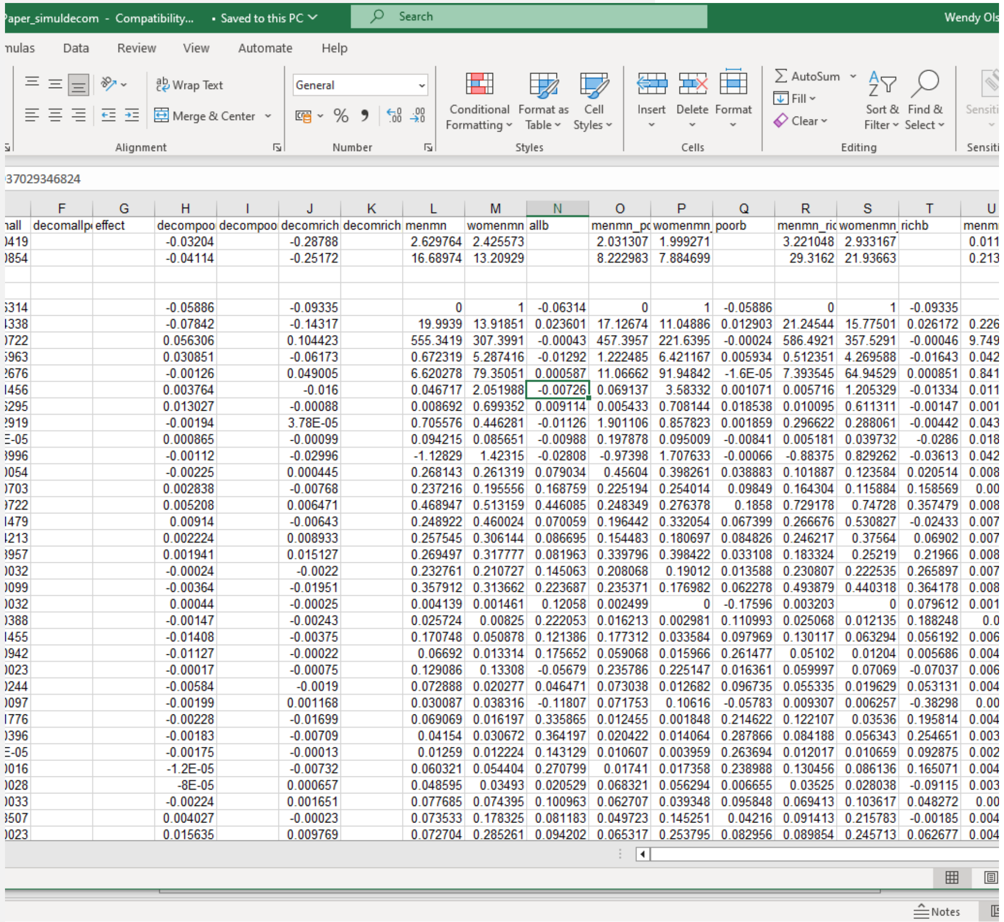
TRICK: If the pay is per month, estimated from the last paypacket, something has been left out of the analysis.

You do not see significant gender difference (red marks the 95% Wald confidence interval) – WHY NOT!!?? There IS a large and significant pay gap in Uganda and Peru - - why not shown?

+ ALSO, NEED TO DEAL WITH MISSING DATA

What about the inactive people with no job, and the unemployed?

- You can possibly impute to them a £1 or $1 payment (per hour).
- This safely places them at the far left of the distribution.
- Regression calculations are sometimes done this way.
- A hurdle regression has two steps. First, what predicts them being in the zero group, and second, what predicts the value of the Y variable for non-zero cases.



	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
nall	decomallp	effect	decompoo	decompoo	decomrich	decomrich	menmn	womenmn	allb	menmn_pc	womenmn_	poorb	menmn_ric	womenmn_	richb	menm
0419			-0.03204		-0.28788		2.629764	2.425573		2.031307	1.999271		3.221048	2.933167		0.011
0854			-0.04114		-0.25172		16.68974	13.20929		8.222983	7.884699		29.3162	21.93663		0.213
5314			-0.05886		-0.09335		0	1	-0.06314	0	1	-0.05886	0	1	-0.09335	
1338			-0.07842		-0.14317		19.9939	13.91851	0.023601	17.12674	11.04886	0.012903	21.24544	15.77501	0.026172	0.226
0722			0.056306		0.104423		555.3419	307.3991	-0.00043	457.3957	221.6395	-0.00024	586.4921	357.5291	-0.00046	9.749
5963			0.030851		-0.06173		0.672319	5.287416	-0.01292	1.222485	6.421167	0.005934	0.512351	4.269588	-0.01643	0.042
2676			-0.00126		0.049005		6.620278	79.35051	0.000587	11.06662	91.94842	-1.6E-05	7.393545	64.94529	0.000851	0.841
1456			0.003764		-0.016		0.046717	2.051988	-0.00726	0.069137	3.58332	0.001071	0.005716	1.205329	-0.01334	0.011
5295			0.013027		-0.00088		0.008692	0.699352	0.009114	0.005433	0.708144	0.018538	0.010095	0.611311	-0.00147	0.001
2919			-0.00194		3.78E-05		0.705576	0.446281	-0.01126	1.901106	0.857823	0.001859	0.296622	0.288061	-0.00442	0.043
E-05			0.000865		-0.00099		0.094215	0.085651	-0.00988	0.197878	0.095009	-0.00841	0.005181	0.039732	-0.0286	0.018
3996			-0.00112		-0.02996		-1.12829	1.42315	-0.02808	-0.97398	1.707633	-0.00066	-0.88375	0.829262	-0.03613	0.042
0054			-0.00225		0.000445		0.268143	0.261319	0.079034	0.45604	0.398261	0.038883	0.101887	0.123584	0.020514	0.008
0703			0.002838		-0.00768		0.237216	0.195556	0.168759	0.225194	0.254014	0.09849	0.164304	0.115884	0.158569	0.00
0722			0.005208		0.006471		0.468947	0.513159	0.446085	0.248349	0.276378	0.1858	0.729178	0.74728	0.357479	0.008
1479			0.00914		-0.00643		0.248922	0.460024	0.070059	0.196442	0.332054	0.067399	0.266676	0.530827	-0.02433	0.007
4213			0.002224		0.008933		0.257545	0.306144	0.086695	0.154483	0.180697	0.084826	0.246217	0.37564	0.06902	0.007
3957			0.001941		0.015127		0.269497	0.317777	0.081963	0.339796	0.398422	0.033108	0.183324	0.25219	0.21966	0.008
0032			-0.00024		-0.0022		0.232761	0.210727	0.145063	0.208068	0.19012	0.013588	0.230807	0.222535	0.265897	0.007
0099			-0.00364		-0.01951		0.357912	0.313662	0.223687	0.235371	0.176982	0.062278	0.493879	0.440318	0.364178	0.008
0032			0.00044		-0.00025		0.004139	0.001461	0.12058	0.002499	0	-0.17596	0.003203	0	0.079612	0.001
0388			-0.00147		-0.00243		0.025724	0.00825	0.222053	0.016213	0.002981	0.110993	0.025068	0.012135	0.188248	0.0
1455			-0.01408		-0.00375		0.170748	0.050878	0.121386	0.177312	0.033584	0.097969	0.130117	0.063294	0.056192	0.006
0942			-0.01127		-0.00022		0.06692	0.013314	0.175652	0.059068	0.015966	0.261477	0.05102	0.01204	0.005686	0.004
0023			-0.00017		-0.00075		0.129086	0.13308	-0.05679	0.235786	0.225147	0.016361	0.059997	0.07069	-0.07037	0.006
0244			-0.00584		-0.0019		0.072888	0.020277	0.046471	0.073038	0.012682	0.096735	0.055335	0.019629	0.053131	0.004
0097			-0.00199		0.001168		0.030087	0.038316	-0.11807	0.071753	0.10616	-0.05783	0.009307	0.006257	-0.38298	0.00
1776			-0.00228		-0.01699		0.069069	0.016197	0.335865	0.012455	0.001848	0.214622	0.122107	0.03536	0.195814	0.004
0396			-0.00183		-0.00709		0.04154	0.030672	0.364197	0.020422	0.014064	0.287866	0.084188	0.056343	0.254651	0.003
E-05			-0.00175		-0.00013		0.01259	0.012224	0.143129	0.010607	0.003959	0.263694	0.012017	0.010659	0.092875	0.002
0016			-1.2E-05		-0.00732		0.060321	0.054404	0.270799	0.01741	0.017358	0.238988	0.130456	0.086136	0.165071	0.004
0028			-8E-05		0.000657		0.048595	0.03493	0.020529	0.068321	0.056294	0.006655	0.03525	0.028038	-0.09115	0.003
0033			-0.00224		0.001651		0.077685	0.074395	0.100963	0.062707	0.039348	0.095848	0.069413	0.103617	0.048272	0.00
3507			0.004027		-0.00023		0.073533	0.178325	0.081183	0.049723	0.145251	0.04216	0.091413	0.215783	-0.00185	0.004
0023			0.015635		0.009769		0.072704	0.285261	0.094202	0.065317	0.253795	0.082956	0.089854	0.245713	0.062677	0.004

COMPARISON USING RCAP IN STATA

A simple comparison of the mean wage-rate by sex gives the 'gender pay gap'

The pay gap at the mean

- In Stata, you will:
- Preserve
- Collapse
- Twoway graph …(Hbar) ..(rcap…)
- The same categorical variable but different continuous variables. I use llmen and ulmen, and llwomen and ulwomen as variable names in the collapse command
- You need to use your Pweights or Aweights in the collapse command
- Then graph … (rcap ulmen llmen) and so on.

Variant Pay-Gaps

- Most obviously, you can do the pay gap at the median wage-rate.
  - Monthly earnings is not a good idea because it ignores the part-time working hours. Therefore, it will exaggerate the gender pay gap.
- You also want to create a sexual-orientation pay gap? Make sure you give the confidence interval for the smaller group, especially if N<30 for the calculation of one of the means.
- Ethnicity Pay Gaps
- Disability Pay Gaps … etc.
- All part of inequalities research.

FIGURE 1 LOG PAY GAP REGRESSION & DECOMPOSITION

CJE article – version before the revisions – The Poor group is lowest 12%.

Table 3. Decomposition of the Gender Pay Gap by Household Type. (OPTION B)

	Analysis at the Mean	Poor Households	Wealthy Households
	% Contributor of total	% Contributor of total	% Contributor of total
Cumulative Work History in Years			
Full-time work	18%	69%	13%
Part-time work	8%	-92%	4%
Unpaid Family Care	7%	-12%	6%
Parental Leave	-3%	-41%	0%
Unemployment	-1%	6%	0%
Illness	0%	-3%	0%
Key Work Indicators			
Occupational sex-segregation	20%	3%	10%
Public Sector	-7%	-29%	2%
Union member	-2%	-7%	-3%
Bonus receipt	6%	-13%	8%
Residuals			
Measured Female Residual	31%	184%	32%
Unobserved Residuals	16%	2%	12%
Total Coefficient Effects	-0.2	-0.03	-0.29

The working poor are defined as workers living in households with equivalised income <60% of median earned income (£1,335 or less). Here, it is the bottom 12% of households. We apply a similar cut-off on the right-hand side, >160% of median, N=10,275

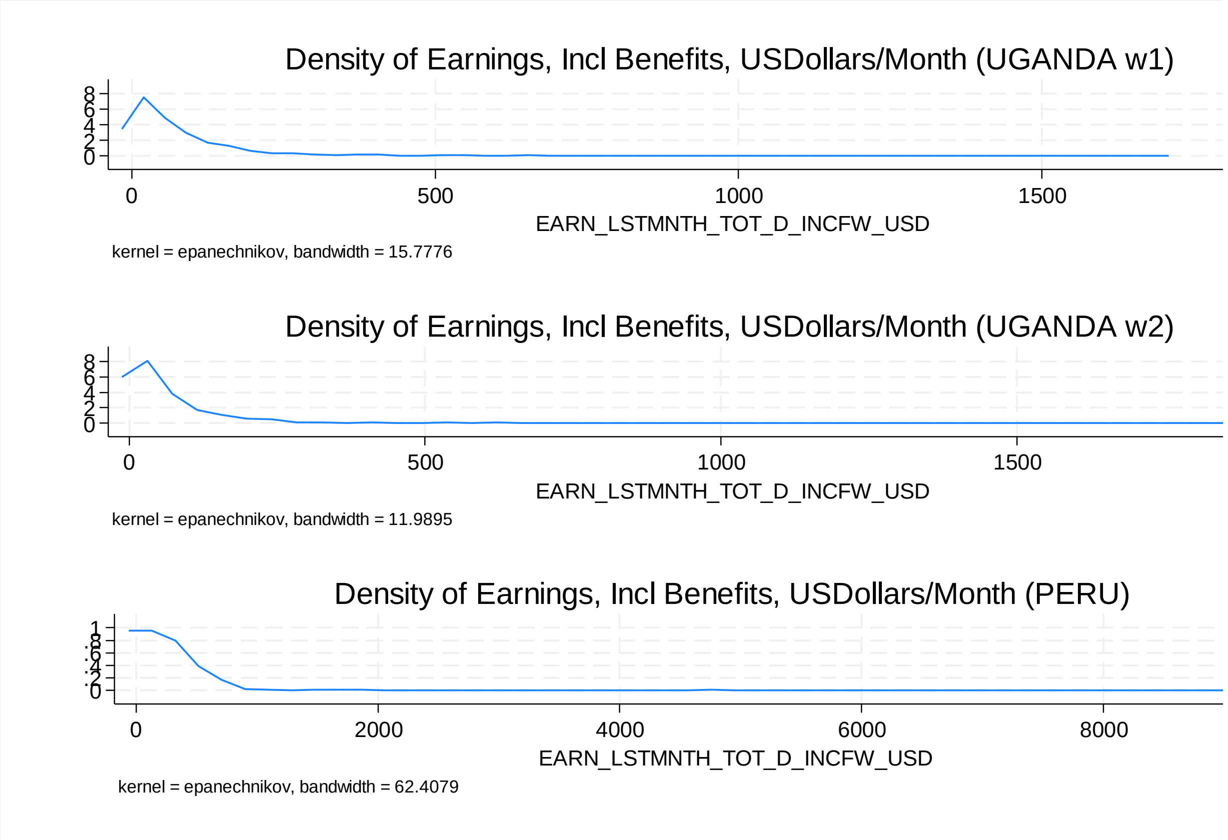
SOOK KIM MADE GREAT CONTRIBUTIONS

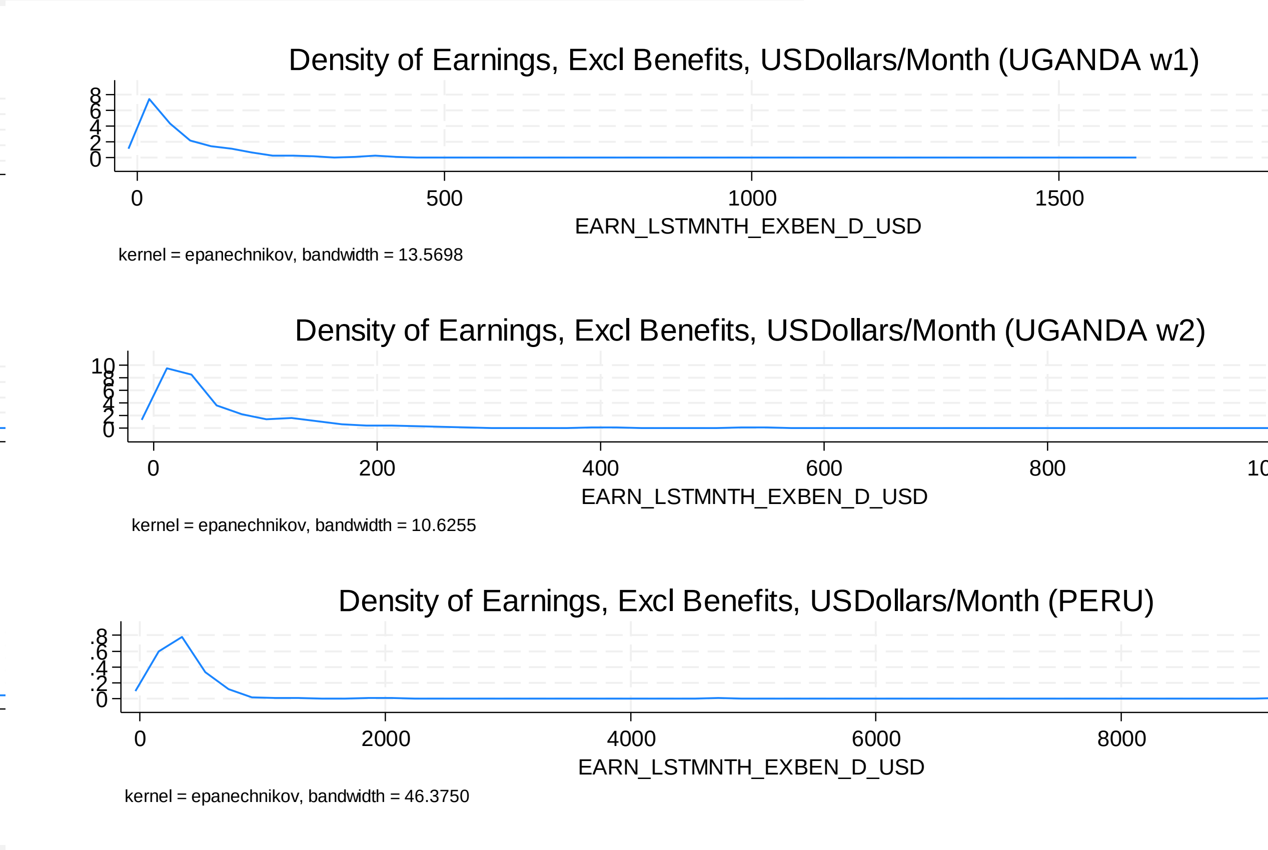
Work histories measured as one-month units on annual recall, overlapping, corrected.

All code reproducible and open-access.

Table 1. Key Descriptive Statistics, weighted by cross-sectional weights

	All Households		Poor Households		Wealthy Households	
	Men	Women	Men	Women	Men	Women
Gross Hourly Pay	£16.67	£13.21	£8.22	£7.90	£29.20	£21.95
Gender Pay Gap	20%		4%		28%	
Cumulative Work History, in years and proportion of time in each status:						
Full-time work	19.98 (0.93)	13.93 (0.62)	17.13 (0.83)	11.06 (0.49)	21.17 (0.96)	15.84 (0.71)
Part-time work	0.67 (0.03)	5.28 (0.23)	1.22 (0.06)	6.39 (0.28)	0.5 (0.02)	4.26 (0.19)
Unemployment	0.71 (0.03)	0.45 (0.02)	1.90 (0.09)	0.86 (0.04)	0.30 (0.01)	0.29 (0.01)
Unpaid family care	0.05 (0.00)	2.05 (0.03)	0.07 (0.00)	3.57 (0.16)	0.01 (0.00)	1.2 (0.03)
Parental Leave	0.01 (0.00)	0.7 (0.03)	0.00(0.00)	0.70 (0.03)	0.01 (0.00)	0.61 (0.03)
Illness	0.09 (0.00)	0.09 (0.00)	0.20 (0.01)	0.10 (0.00)	0.01 (0.00)	0.04 (0.00)





This set of diagrams is not men/women. It is inclusive of in-kind (LEFT) versus excluding in-kind payments (RIGHT).

DECOMPOSING THE BARRIERS TO EQUAL PAY: EXAMINING DIFFERENTIAL PREDICTORS OF THE GENDER PAY GAP BY SOCIO-ECONOMIC GROUP’

By Vanessa Gash, Sook Kim, Nadine Zweiner, and Wendy Olsen

Submitted to Cambridge Journal of Economics 2022

Revised and resubmitted to CJE 2024

Keywords: gender pay gap, sex-segregation, work-history, working-time.

JEL: B54, Feminist Economics, E24, Employment and Wages, J31, Wage Differentials.

We submitted the paper in 2022. We received 9 pages single spaced editorial & reviewer comments in Sept. 2023. We resubmitted in Feb 2024. We await a response now (May 2024).

The equations in the paper cover the Decomposition of Pay Gaps by the Blinder-Oaxaca two-term method. This method has enduring interest. One reason is that any linearised model can be decomposed, but when we use nonlinear Generalised Linear modelling we often cannot decompose the factors’ influence amounts upon the Y variable. Authors who don’t know GLM use Blinder-Oaxaca. Possibly it is a blind alley because then, we are not investing time in better models.

A hurdle model or a Tobit model was avoided in this paper.

TABLE 1 K-S TEST OF THE DIFFERENCE OF TWO DISTRIBUTIONS (MEN'S VS WOMEN'S)

The non-parametric Kolmogorov-Smirnov test is often used if one of the variables is ordinal. Here the distribution is so awkward it is treated as if it were ordinal; or you can do Spearman's on the ranks.

						Company revenue
Quiz						£0
Sketch the logged wage distribution of men						£1,013
Sketch the same for women						£5,063
Sketch them on the same graph						£20,250
Sketch the difference-distribution						£40,500

GOOD EXPLANATORY POWER ON PAY-GAPS

In the OECD, the bonus culture has created some explosive salary levels. Taking logarithms does not fully solve this problem. In regression, you can offer a bonus binary variable. This gets a high positive coefficient and has a role as a % of the explained variation of pay. This is useful for decompositing two sub-groups Y_a and Y_b.

When the variable is highly skewed, the error term in regression will not be normally distributed. Therefore, treat the variable using some of the three options a, b, and c:

a) *Transform it using logs;*

b) *Add a binary variable to explain a key part of the right-hand skewness;*
   - a) Overtime explains higher wages
   - b) Union membership explains higher wages
   - c) Having a degree explains higher wages
   - d) Usually in regression we get up to 30-35 variables.

     In India, the exclusion of the informal labour relationships creates many, many zeroes for ‘wage’.

c) *Add a hurdle model to explain the zeroes part. A Tobit model, a negative binomial model, or a zero-inflated binomial model could also work. All these are harder/ impossible to decompose.*

After these steps, your Wage Equation residuals may be normally distributed and your explanatory %’s in the decomposition make sense.

One can standardise all the variables. You then have also the same units in all variables. But standardising the binary variables is very much argued about. See Gelman (2008) about standardising binary variables. Good luck to you with research!

REFERENCES ON PAY GAPS PLUS GELMAN AND WARD AND AHLQUIST

- Cooke, L.P. 2014. Gendered parenthood penalties and premiums across the earnings distribution in Australia, the United Kingdom, and the United States, *European Sociological Review,* vol. 30, no. 3, 360-372
- England, P. 1992. From Status Attainment to Segregation and Devaluation, *Contemporary Sociology*, vol. 21, no. 5, 643-647
- England, P. 2010. The gender revolution: Uneven and stalled, *Gender & Society*, vol. 24, no. 2, 149-166
- Gelman, Andrew (2008) Scaling regression inputs by dividing by two standard deviations, *Statistics In Medicine,* 27: 2865–2873, Doi https://stat.columbia.edu/~gelman/research/published/standardizing7.pdf
- Kim, J., Olsen, W., 2023. Harmful Forms of Child Labour in India from a Time-use perspective. *Development in Practice* 33, 190–204. doi:10.1080/09614524.2022.2155620.
- Kim, J., Olsen, W., Wiśniowski, A., 2022. Predicting Child-labour Risks by Norms in India. *Work, Employment and Society* 37. 1605-1626. doi:10.1177/09500170221091886.
- Ward, Michael D., and John S. Ahlquist (2018) *Maximum Likelihood for Social Science Strategies for Analysis,* Cambridge University Press. URL https://www.cambridge.org/core/books/maximum-likelihood-for-social-science/958C893D749B06A2A7D09232F0B9D959 E-book is in UOM Rylands Library.

THANK YOU

Wendy Olsen

+44 7891 266635

Wendy.Olsen@manchester.ac.uk

Social Statistics Department

University of Manchester (click here for the staff of the department)