Social statistics away-day May 2024

PAY GAP REPRESENTATION

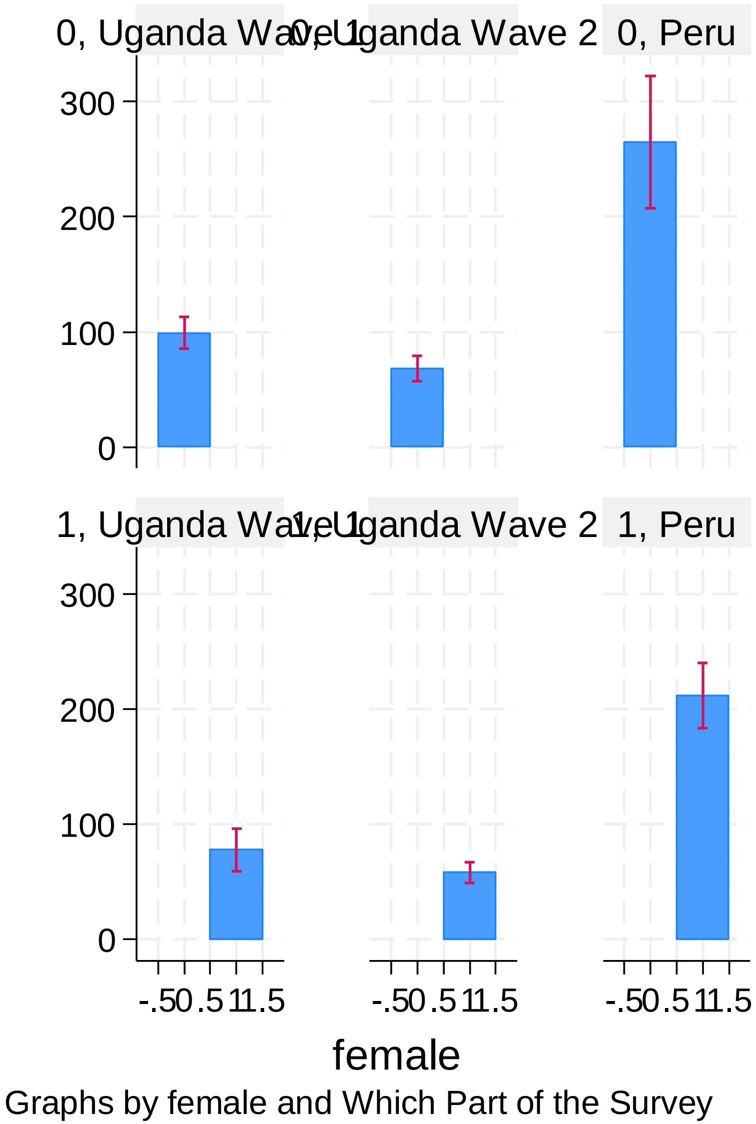
By Wendy Olsen with thanks to Myong Sook Kim

PAY GAP COULD BE AT THE MEAN OR MEDIAN.

It is inherently an aggregate statistic.

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.



Ll is lower limit of the 95% Wald confidence interval

Ul is the upper limit of the 95% Wald confidence interval

The pay is per person per month

The first panel is the Males

The second panel is the Females

We have waves 1 and 2 in Uganda

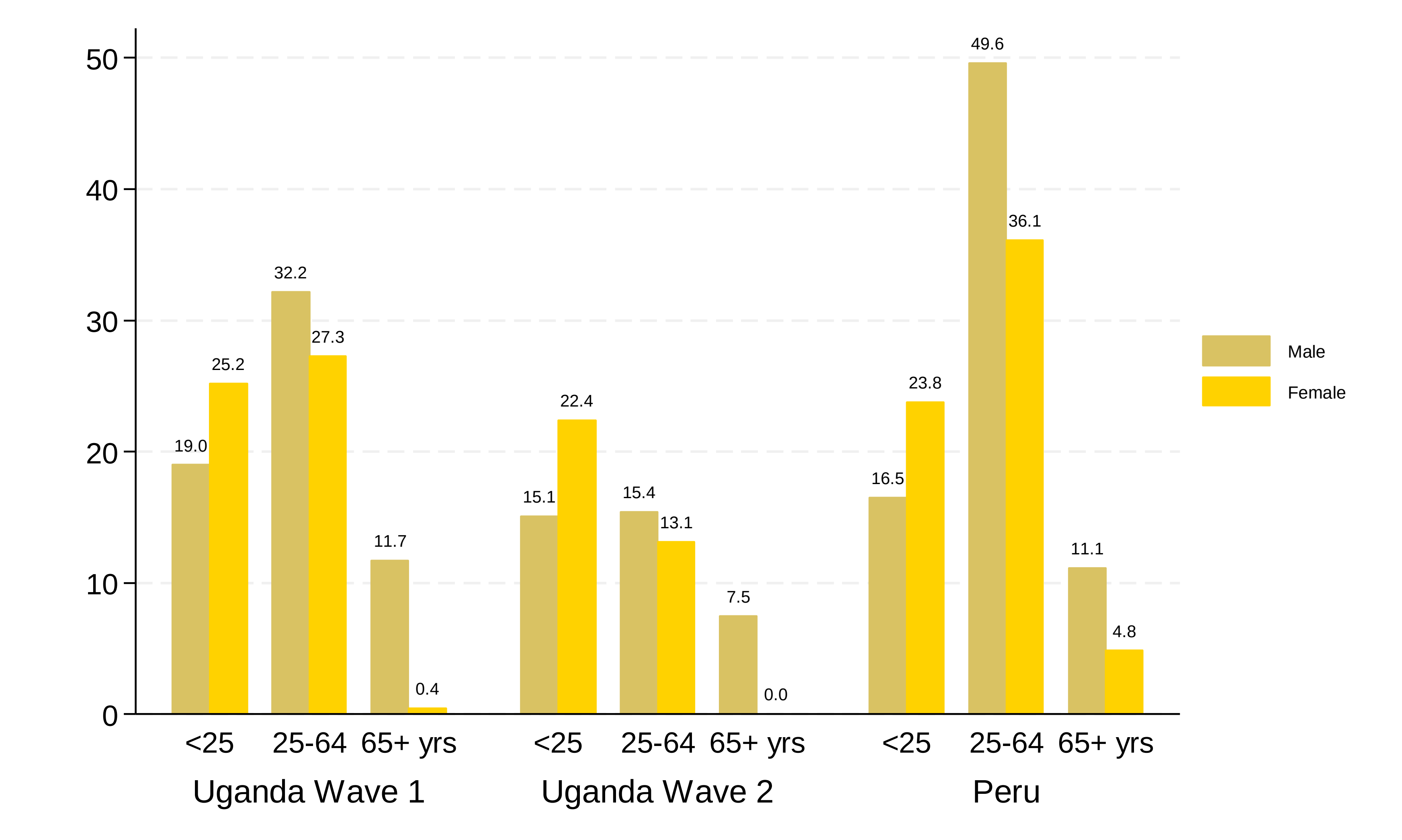
At right, Peru has just one wave.

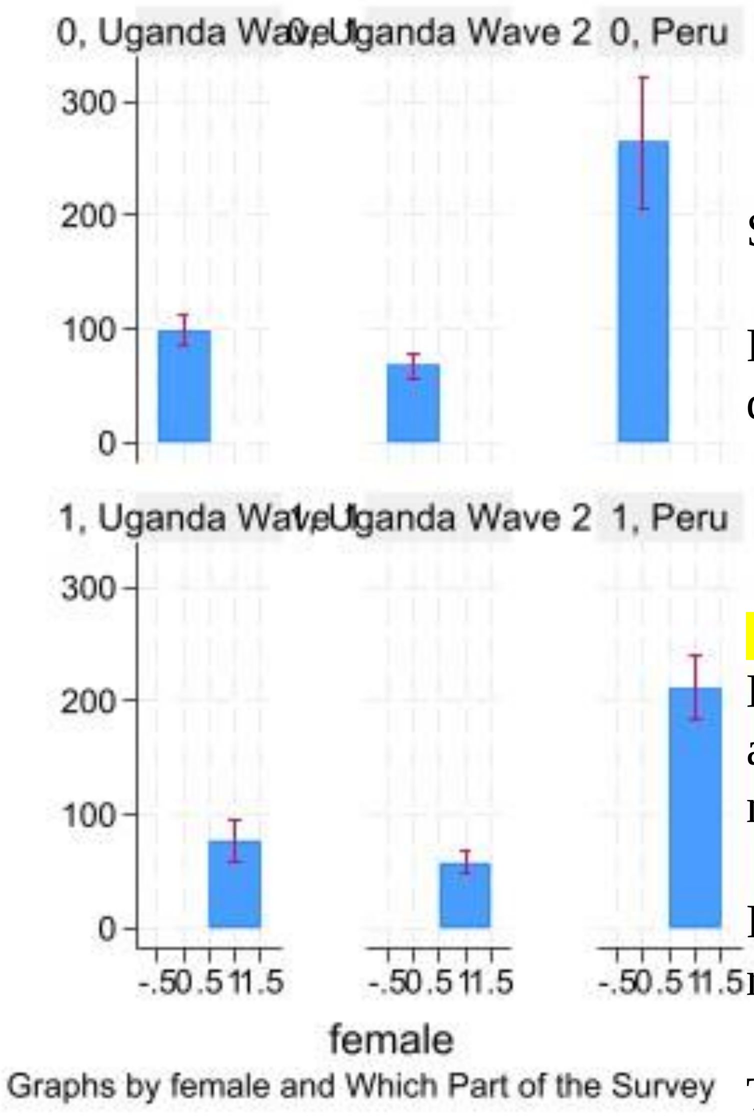
```
collapse (mean) meanearnmjincusd =
EARN_LSTMNTH_TOT_D_INCFW_USD (median) medianearnmjincusd =
EARN_LSTMNTH_TOT_D_INCFW_USD (semean) semeanearnmjincusd =
EARN_LSTMNTH_TOT_D_INCFW_USD [aweight = WEIGHT_FINAL],
by(female wave)
lab var female "Gender of Respondent"
lab def female 0 "Male" 1 "Female", modify
lab var female female
list
gen ll=meanearnmjincusd-1.96*semeanearnmjincusd
gen ul=meanearnmjincusd+1.96*semeanearnmjincusd
gen llmed=medianearnmjincusd-1.96*semeanearnmjincusd
gen ulmed=medianearnmjincusd+1.96*semeanearnmjincusd
graph bar meanearnmjincusd female, by(wave female)

graph twoway  (bar meanearnmjincusd female , by(female 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
<25
25-64
65+ yrs
<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.

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

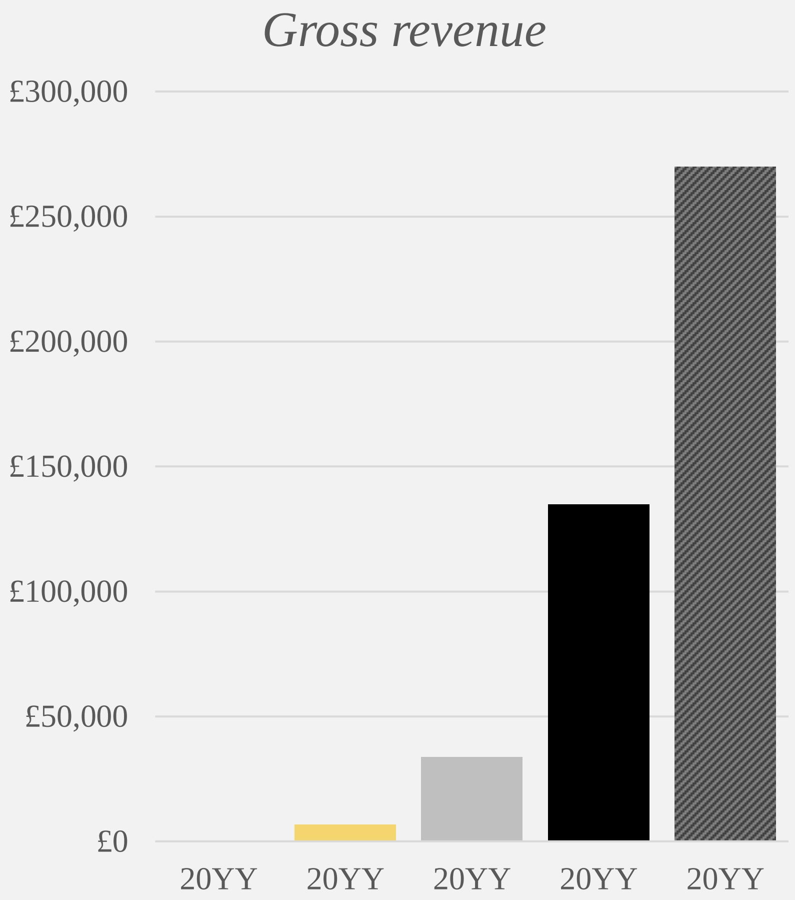
- In Stata, you will:
- Preserve
- Collapse
- Twoway graph …(Hbar) ..(rcap…)
- The same categorical variable but different continuous variables. I use llmen and ulmen, and llwomen and ulwomen as variable names in the collapse command
- You need to use your Pweights or Aweights in the collapse command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

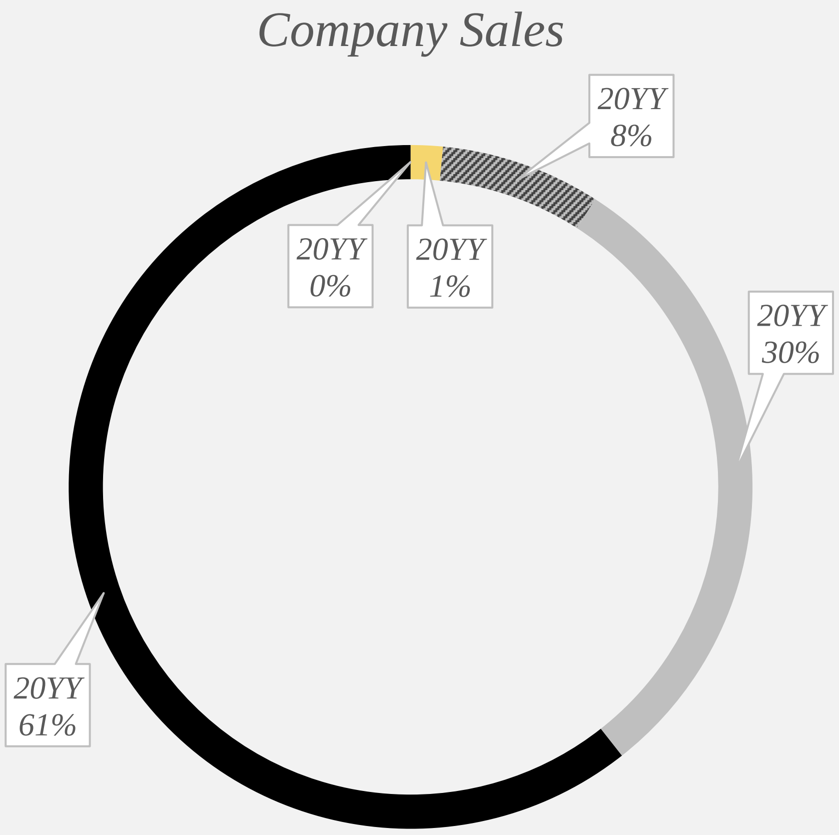
Variant Pay-Gaps

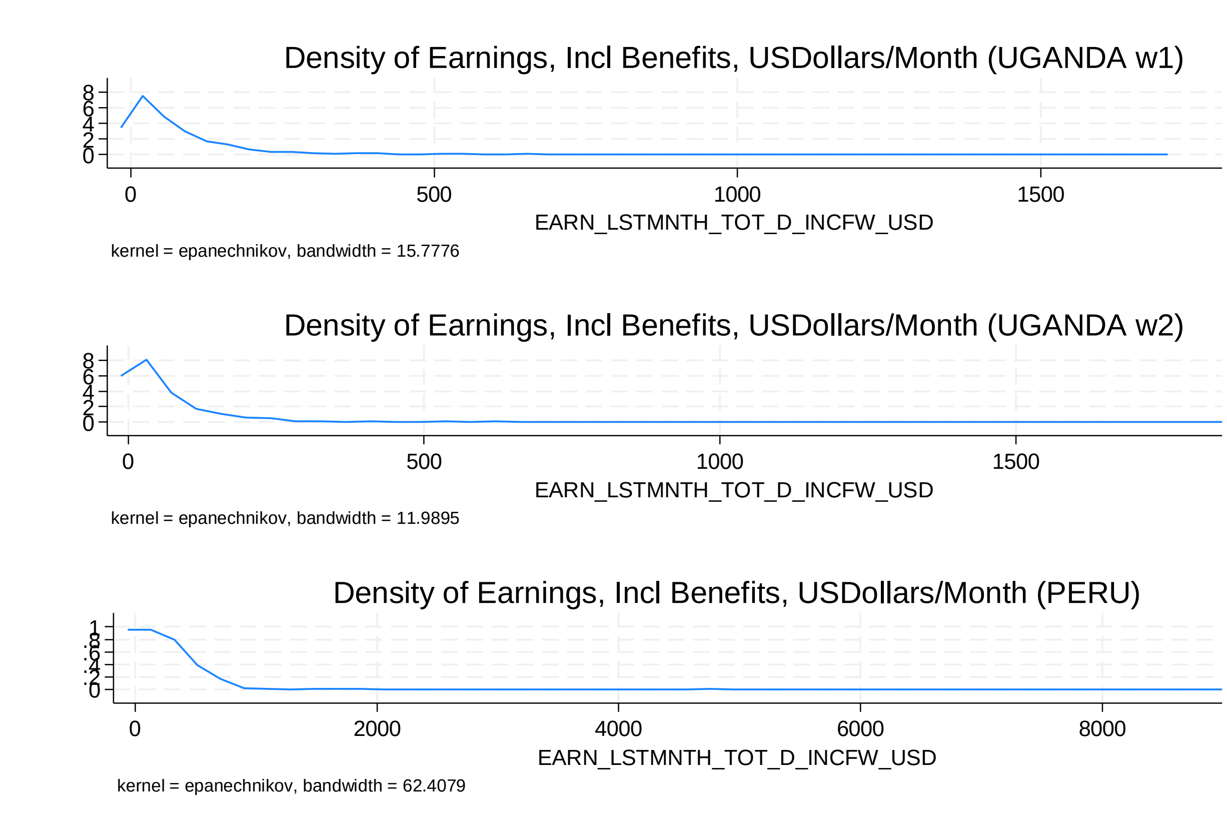
- Most obviously, you can do the pay gap at the median wage-rate.
  - Monthly earnings is not a good idea because it ignores the part-time working hours. Therefore, it will exaggerate the gender pay gap.
- You also want to create a sexual-orientation pay gap? Make sure you give the confidence interval 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

Lorem







Density of Earnings, Incl Benefits, USDollars/Month (UGANDA w1)

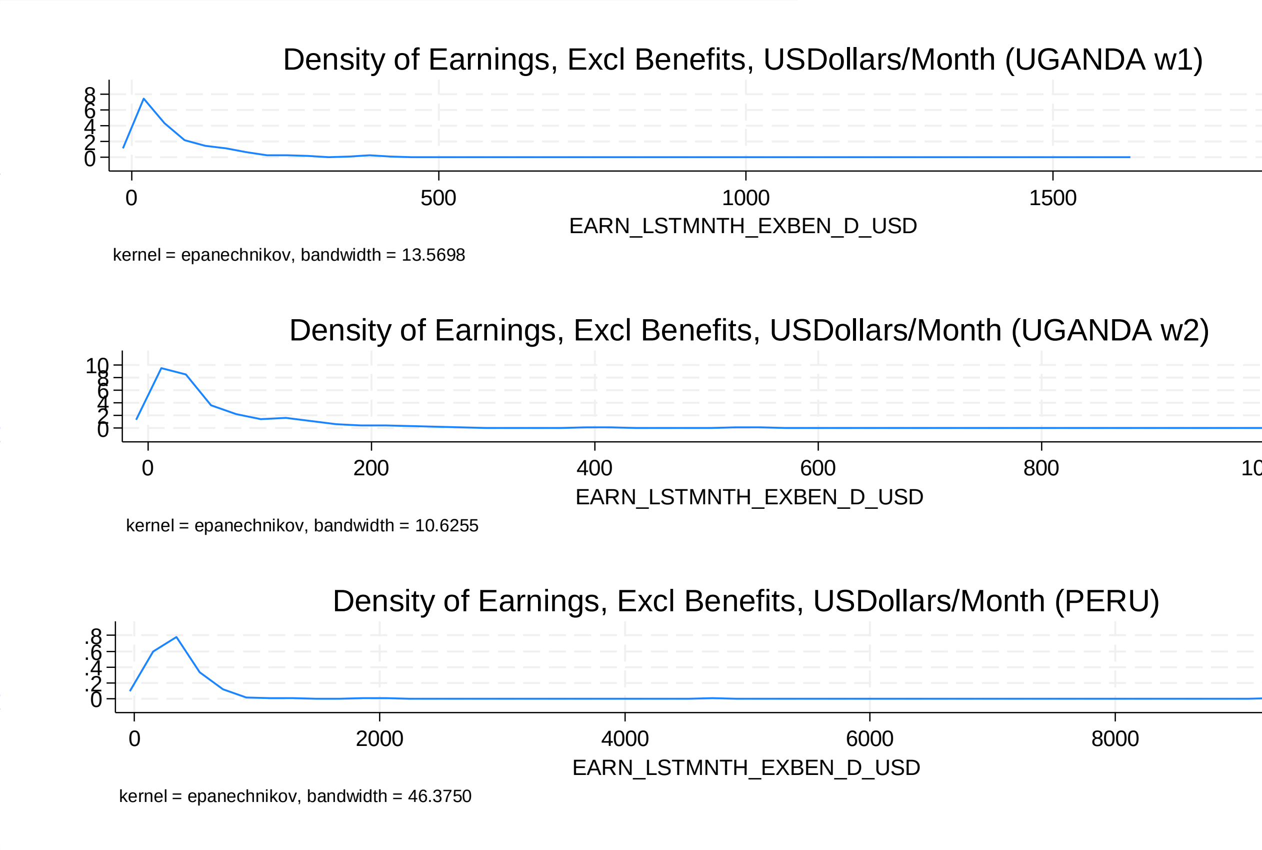
kernel = epanechnikov, bandwidth = 15.7776

Density of Earnings, Incl Benefits, USDollars/Month (UGANDA w2)

kernel = epanechnikov, bandwidth = 11.9895

Density of Earnings, Incl Benefits, USDollars/Month (PERU)

kernel = epanechnikov, bandwidth = 62.4079



Density of Earnings, Excl Benefits, USDollars/Month (UGANDA w1)

kernel = epanechnikov, bandwidth = 13.5698

Density of Earnings, Excl Benefits, USDollars/Month (UGANDA w2)

kernel = epanechnikov, bandwidth = 10.6255

Density of Earnings, Excl Benefits, USDollars/Month (PERU)

kernel = epanechnikov, bandwidth = 46.3750

‘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

The 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 Spearmans 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 explosive salary levels among a small group. Taking logarithm does not fully solve this problem. In regression, you can offer a Bonus Binary variable. This takes a high positive coefficient. It then has a role as a % of the explained variation of the logged pay. This is useful for for Y as pay, and for decomposition of 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, it can be useful to 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.

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

c) Add a hurdle to explain the zeroes part of the non-normality of the wage distribution, if necessary. A Tobit model, a negative binomial model, or a zero-inlated binomial model will work, too But these are harder to decompose.

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

But if you prefer, standardise the Wage variable and all other variables. You then have also the same units in all variables. But standardising the binary variables is very much argued about. Gelman

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)