

CPS Displaced Worker Supplement

Ebba Mark

2025-07-30

As part of the Current Population Survey, the US Census Bureau conducts an annual Displaced Worker Supplement in which workers who have lost their job in the last three years are asked additional questions about their unemployment experiences and (if re-employed) their re-employment conditions.

"The universe for the Displaced Workers Supplement is civilians 20 or older. Respondents are further categorized as a "displaced worker" if they meet additional characteristics (see DWSTAT). After 1998, displaced workers are those who lost or left a job due to layoffs or shutdowns within the past 3 years... were not self-employed, and did not expect to be recalled to work within the next six months.

The data used below is from annual survey responses between 2000-2025. I use the supplement sample weights in all results below. I note where I have clipped the sample for outliers (wage ratio between [0.25, 2] and unemployment duration less than 96 weeks (~24 months).

Below I:

1. **Data Cleaning Procedure:** Show data cleaning just for reference (feel free to ignore)!.
2. **Descriptives:** Show some descriptives about the data itself.
3. **Regression Results on Non-Uniform Sample:** Regression results with ratio of new wage to wage at the lost job (W_h and W_w) regressed (cross-sectionally) on unemployment duration with and without various combinations of control variables (whether or not an individual received unemployment compensation, age, race, sex, marital status, education, previous wage level.) Note that the wages are reported in hourly and weekly values but this reporting is inconsistent across observations. In other words, though most individuals (4600/6198) report their wage in both units, 270 report only hourly and 1328 report only weekly. I have not reconciled the inconsistency so I use hourly wage ratios in majority of the below document. I could try to reconcile this.
4. Outline some considerations for further improvement of the analysis:
 1. **Reweighted Samples:** The sample is non-uniform in unemployment duration (less observations as unemployment duration increases). Try two methods of reweighting to address selection issues (Heckman Selection correction - though I think this is inappropriate for this particular selection issue) and non-uniform (entropy-balancing to deal with representativeness of population over unemployment durations) sample confirm regression results in non-uniform sample.
 2. **Representativeness of the Sample (Education, Age, Gender, and Wage):** Representativeness of the data to motivate data limitations and inform the ultimate reweighting scheme.

Overall result (at the moment): Individuals accept a ~1-percentage point change in the wage ratio per additional month of unemployment. Variations using model reweighting, different samples, combinations of control variables, reported hourly and weekly wage ratios do not seem to affect the result. However, the data seems to follow a non-linear relationship (we see little satisfying until around ~12 months of unemployment) after which the wage ratio begins to decrease. Individuals seem to accept a below-1 relative wage ratio (current wage:wage at lost job) following a year of unemployment. If we fit this model with a quadratic fit this could inform our reservation wage adjustment parameter in the model.

Important Considerations/Limitations:

- 1. Displaced worker classification as outlined above.** Can we generalise from this definition to all unemployed workers?
- 2. The reported ‘current wage’ is not necessarily the the realised wage post-re-employment.** Individuals report the wage at the lost job, the amount of time unemployed until they were re-employed, and the wage they hold at their current job. However, it is not indicated whether the current job is the same job as the first they were re-employed at. Given various comments in the literature about finding “stop-gap” employment, this might not be a problem in the sense that the “current wage” would more accurately indicate the wage an individual has “landed” at post-unemployment spell. But curious what you think about the defensibility of this.
- 3. Outcome variable:** How do we feel about the outcome variable as the ratio of current to latest held job? Might we want to take the log or consider simply the (log) level regressed on the previous wage. Wondering if a ratio-based outcome variable might muddle interpretation. Curious for your reactions.

Data cleaning

Feel free to ignore this code chunk immediately below - included for now for transparency in case you spot issues. I include it for your info on binning and outlier trimming.

```
# From the original dataset, I include only those that reported having lost a
# FT job in the last three years
df <- readRDS(here("data/behav_params/cps_displaced_worker_supplement/cps_disp_filtered.RDS")) %>%
  select(hwtfnl, cpsid, wtfnl, age, sex, race, marst, educ,
         # age, sex, race, marital status, educational attainment
         dwsuppwt, # Survey weight
         dwyears, # Years worked at lost job
         dwben, # Received unemployment benefits
         dwexben, # Exhausted unemployment benefits
         dwlastwrk, # Time since worked at last job
         dwweekc, # Weekly earnings at current job
         dwweekl, # Weekly earnings at lost job
         dwwagel, # Hourly earnings at lost job
         dwwagec, # Hourly wage at current job
         dwhrswkc, # Hours worked each week at current job
         dwresp, # Eligibility and interview status for Displaced Wrkr Supplement
         # Interestingly the unemployment duration is not directly linked to
         # CURRENT job and we cannot see the wage of the start of the next job...
         # thought this feels problematic, it does indicate more accurately the
         # ultimate "recovered" wage...will need to declare as a limitation
         # but also not completely indefensible
         dwwksun) %>% # Number of weeks not working between between end of lost
         # or left job and start of next job
# I remove anyone who is Not in Universe (99) and declaring greater than 160
# weeks unemployed between jobs
filter(dwhrswkc != 99 & dwwksun <= 160) %>%
  # Replacing NIU values with NA values
  mutate(dwagel = ifelse(round(dwagel) == 100, NA, dwagel),
         dwagec = ifelse(round(dwagec) == 100, NA, dwagec),
         dwweekl = ifelse(round(dwweekl) == 10000, NA, dwweekl),
         dwweekc = ifelse(round(dwweekc) == 10000, NA, dwweekc),
         # duwage_rec_l = ifelse(is.na(dwagel) & !is.na(dwweekl) ~ dwweekl),
         # duweekc = ifelse(round(dwweekc) == 10000, NA, dwweekc),
         # Binning educational categories
         educ_cat = factor(case_when(educ %in% c(1) ~ NA, # (NIU)
```

```

educ > 1 & educ <= 71 ~ "Less than HS", # Includes
# "None" - Grade 12 no diploma
# (8 subcategories (grade 1-11 etc))
educ %in% c(73, 81) ~ "HS Diploma", # Includes
# "High school Diploma or equivalent" and
# "some college, but no degree"
educ %in% c(91, 92) ~ "Associate's", # Include
# "[Associate's degree, occupational/vocational
# program]" and "Associate's
# [Associate's degree, academic program]"
educ %in% c(111) ~ "Bachelor's", # Bachelor's degree
educ > 111 ~ "Postgraduate Degree" # Includes
# Master's, Professional School & Doctorate
), , levels = c("Less than HS", "HS Diploma",
"Associate's", "Bachelor's",
"Postgraduate Degree")),
# Marital status to binary indicator
marst = case_when(marst == 1 ~ 1, # Married with a present spouse
# Might consider dividing this differently
TRUE ~ 0), # Married with absent spouse, separated,
# divorced, widowed, never married/single

female = sex == 2, # gender to 0,1 values
white = race == 100, # race to higher-level categories w binary values
black = race == 200,
mixed = race %in% c(801, 802, 803, 804, 805,
806, 810, 812, 813, 820, 830),
aapi = race %in% c(650, 651, 652, 808, 809),
native = race == 300
# age is continuous which seems fine...binning likely unnecessary
) %>%

mutate(ratio_wage = dwwagec/dwwagel, # Ratio of hourly wage current:lost job
ratio_weekly = dwweekc/dwweekl, # Ratio of wkly wage of current:lost job
# Reconciling missing reporting between weekly and hourly wage.
# Take either the min, max or mean value.
ratio_reconciled_min = case_when(is.na(ratio_wage) ~ ratio_weekly,
is.na(ratio_weekly) ~ ratio_wage,
TRUE ~ pmin(ratio_weekly, ratio_wage)),
ratio_reconciled_max = case_when(is.na(ratio_wage) ~ ratio_weekly,
is.na(ratio_weekly) ~ ratio_wage,
TRUE ~ pmax(ratio_weekly, ratio_wage)),
ratio_reconciled_mean = case_when(is.na(ratio_wage) ~ ratio_weekly,
is.na(ratio_weekly) ~ ratio_wage,
TRUE ~ rowMeans(across(c(ratio_wage,
ratio_weekly)),
na.rm = TRUE)),
# Create monthly unemployment duration for continuous
dwmosun = floor(dwwksun/4),
dwmosun2 = dwmosun^2,
dwmosun3 = dwmosun^3,
# Unemployment duration:
# (reported as time between lost job and start of next job)

```

```

# I bin in...
# monthly intervals (4 weeks) from 1-6 months
# quarterly intervals (12 weeks) from 7 mos-1 year
# half-year interval from 1-2.5 years
# single bin for anyone about 120 weeks
dwwksun_bin = case_when(
  # Monthly intervals (4 weeks) from 1-6 months
  dwwksun <= 4 ~ 1, #"Less than 4 weeks",
    dwwksun > 4 & dwwksun <= 8 ~ 2,
    dwwksun > 8 & dwwksun <= 12 ~ 3,
    dwwksun > 12 & dwwksun <= 16 ~ 4,
    dwwksun > 16 & dwwksun <= 20 ~ 5,
    dwwksun > 20 & dwwksun <= 24 ~ 6,
    # Quarterly Intervals (12 wks) from 6+ mos-1 yr
    dwwksun > 24 & dwwksun <= 36 ~ 7,
    dwwksun > 36 & dwwksun <= 48 ~ 8,
    # Half-year Intervals (24 weeks) from 1-2.5 yrs
    dwwksun > 48 & dwwksun <= 72 ~ 9,
    dwwksun > 72 & dwwksun <= 96 ~ 10,
    dwwksun > 96 & dwwksun <= 120 ~ 11,
    # Anyone above-recall this is capped at 160 wks
    # as per filter above
    dwwksun > 120 ~ 12),
  # Bin labels
dwwksun_bin_labs = case_when(dwwksun_bin == 1 ~ "<= 1 mo.", #< 4 wks",
  dwwksun_bin == 2 ~ "1-2 mos.",
  dwwksun_bin == 3 ~ "2-3 mos.",
  dwwksun_bin == 4 ~ "3-4 mos.",
  dwwksun_bin == 5 ~ "4-5 mos.",
  dwwksun_bin == 6 ~ "5-6 mos.",
  # Quarterly Intervals (12 wks) from 6+ mos-1 yr
  dwwksun_bin == 7 ~ "6-9 mos.",
  dwwksun_bin == 8 ~ "9-12 mos.",
  # Half-year Intervals (24 wks) from 1-2.5 years
  dwwksun_bin == 9 ~ "12-18 mos.",
  dwwksun_bin == 10 ~ "18-24 mos.",
  dwwksun_bin == 11 ~ "24-30 mos.",
  # Anyone above-recall this is capped at 160 wks
  # as per filter above
  dwwksun_bin == 12 ~ "30+ mos."),
log_ratio_wage = log(ratio_wage),
log_ratio_weekly = log(ratio_weekly),
# I clip the sample to an accepted wage ratio between [0.5, 2]
# and less than 96 weeks of unemployment
clipped_sample_hwage =
  ratio_wage >= 0.5 & ratio_wage <= 2 & dwwksun_bin < 11,
clipped_sample_wwage =
  ratio_weekly >= 0.5 & ratio_weekly <= 2 & dwwksun_bin < 11,
clipped_sample_rec_min =
  ratio_reconciled_min>=0.5 & ratio_reconciled_min<=2 & dwwksun_bin<11,
clipped_sample_rec_max =
  ratio_reconciled_max>=0.5 & ratio_reconciled_max<=2 & dwwksun_bin<11,
clipped_sample_rec_mean =

```

```
ratio_reconciled_mean>=0.5 & ratio_reconciled_mean<=2 & dwwksun_bin<11)
```

Descriptives

All descriptives below use the Displaced Worker Sample Weights.

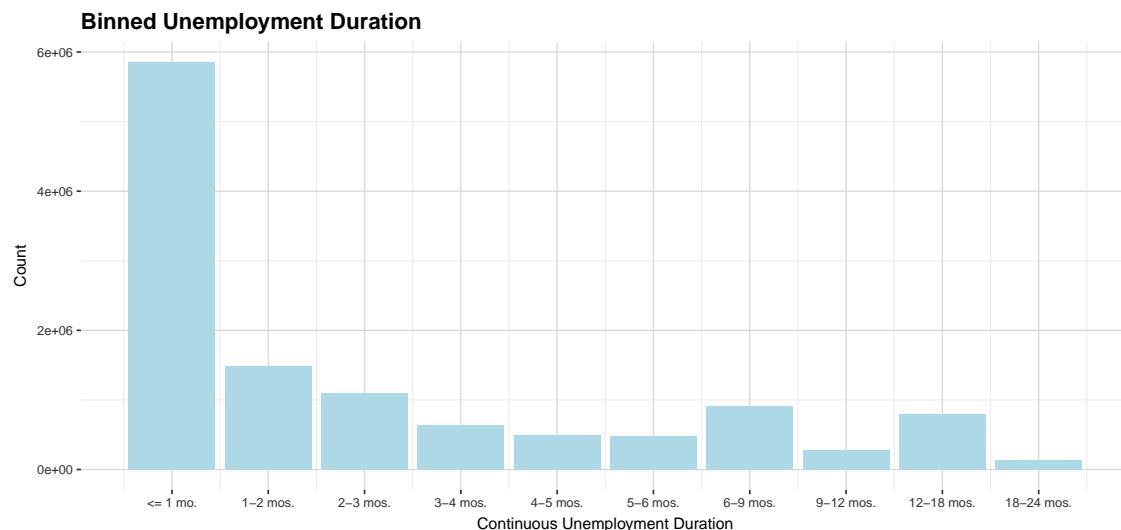
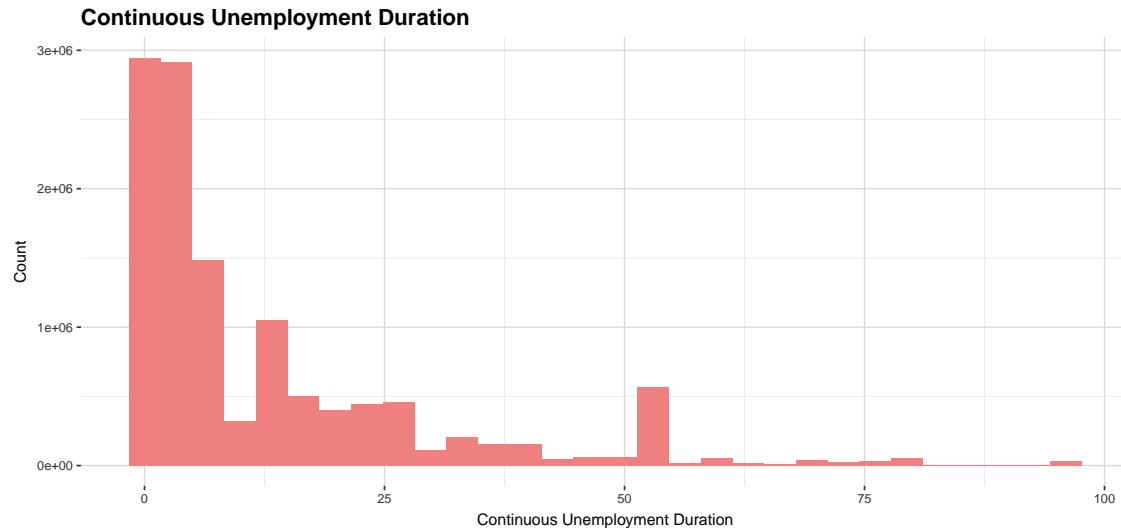
Histogram: sample is skewed (see reweighting alternatives at end of document).

Box plots: Looking at the reported wage ratios in weekly and hourly values, the mean is fixed near 1 until >12 mos of unemployment in hourly wage reporting. In weekly wage reporting, the “satisficing” seems to start earlier in unemployment duration (sample size is larger for weekly reporting - might be worth focusing on those wages).

Scatter plot: I fit a linear and spline fit to the scattered plot of the wage ratio to unemployment duration before using the regression. Indicates decline in the wage ratio with unemployment duration that has a potentially non-linear fit.

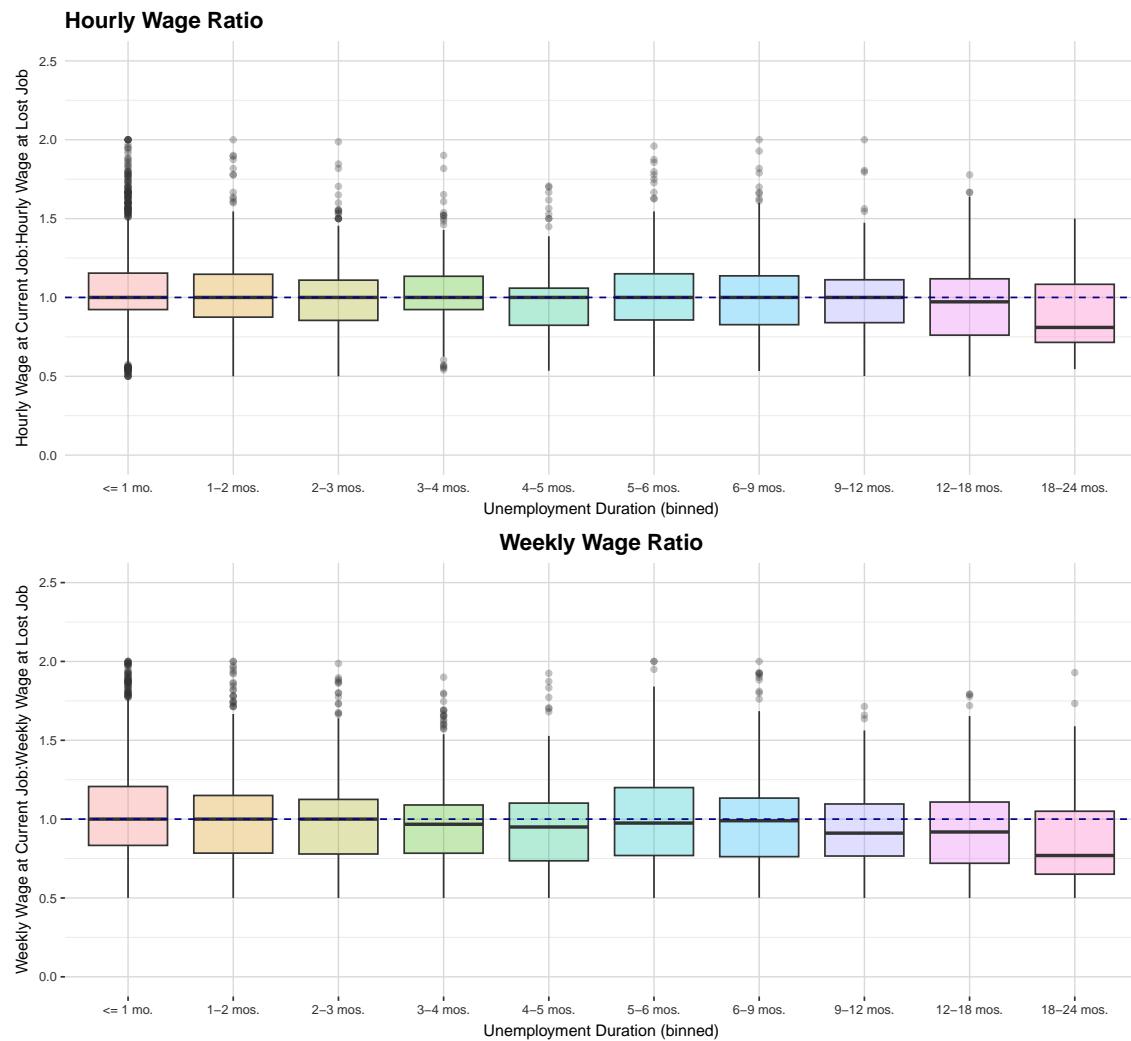
Reported unemployment duration between lost job and next job

Observations weighted by Displaced Worker Supplement Weights. Annual data from 2000–2025. Exclude observations reporting > 96 weeks of unemployment.



Reported ratio of current wage to lost wage by unemployment duration

Observations weighted by Displaced Worker Supplement Weights.
 Annual data from 2000–2025.
 Exclude observations reporting > 96 weeks of unemployment.



Reported ratio of current wage to lost wage by unemployment duration

Observations weighted by Displaced Worker Supplement Weights.

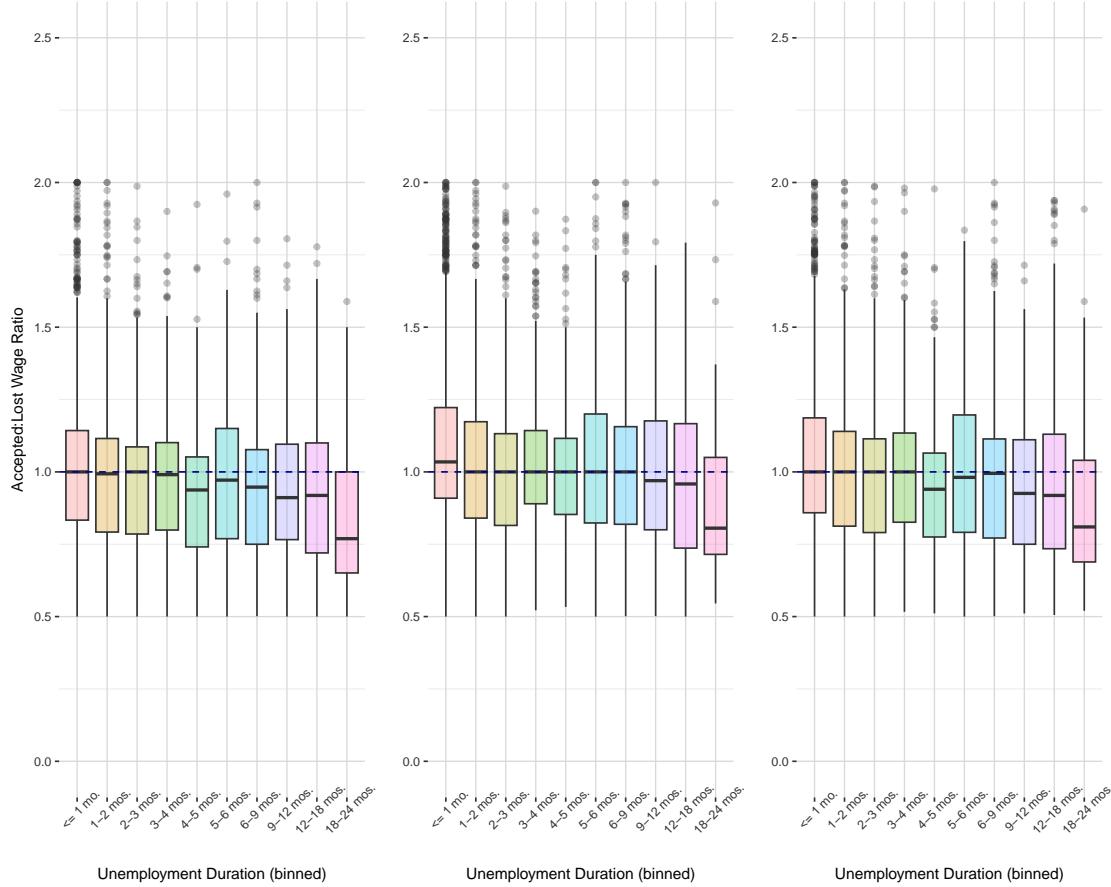
Annual data from 2000–2025.

Exclude observations reporting > 96 weeks of unemployment.

In many cases, only hourly OR weekly wages are reported.

To be able to combine information on all workers to one value, we select the present statistic for those missing one and retain either the minimum, maximum, or mean of the hourly versus weekly wage for those reporting both.

Wage Ratio – Reconciled with min **Wage Ratio – Reconciled with max** **Wage Ratio – Reconciled with r**

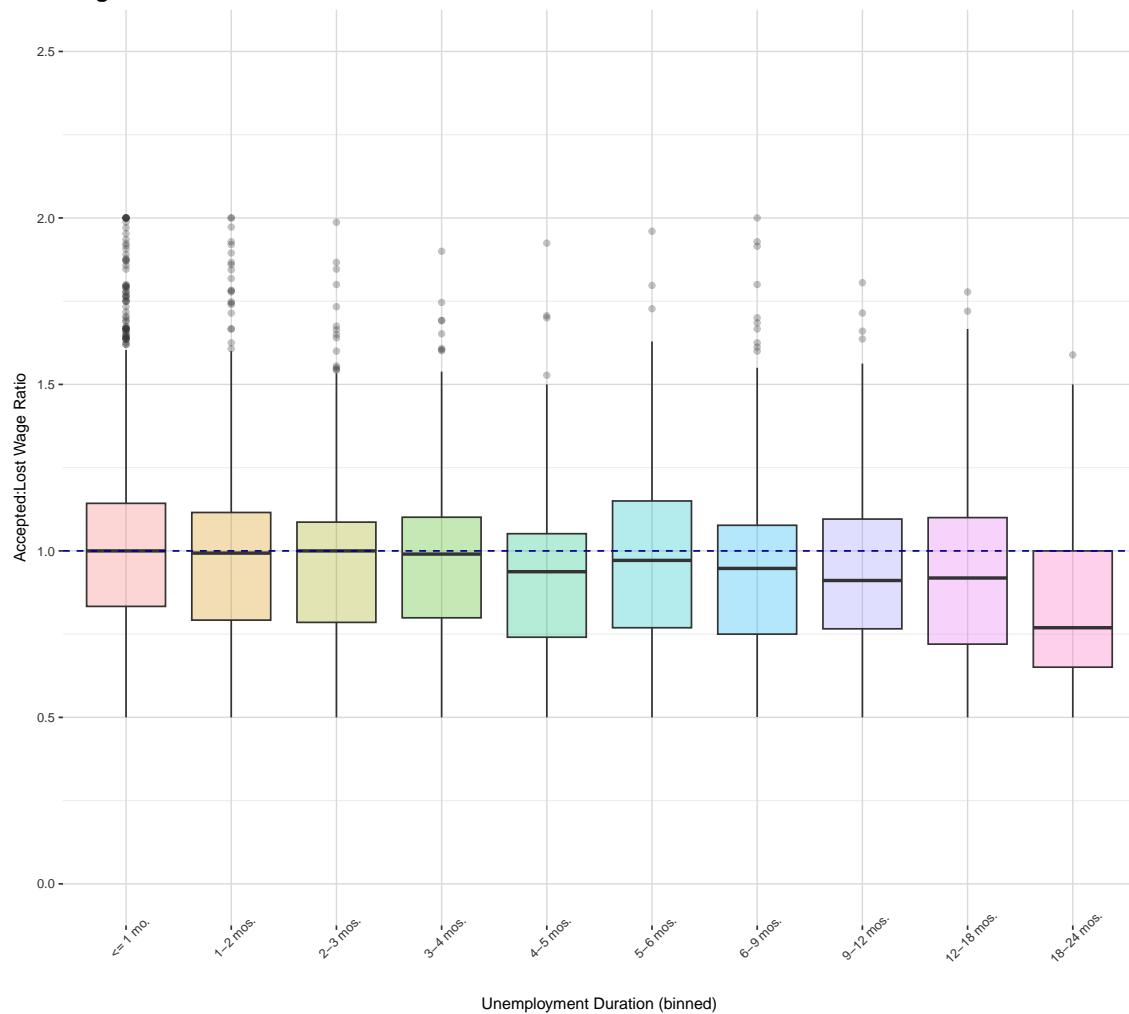


Reported ratio of current wage to lost wage by unemployment duration

Observations weighted by Displaced Worker Supplement Weights. Annual data from 2000–2025.

Exclude observations reporting > 96 weeks of unemployment.

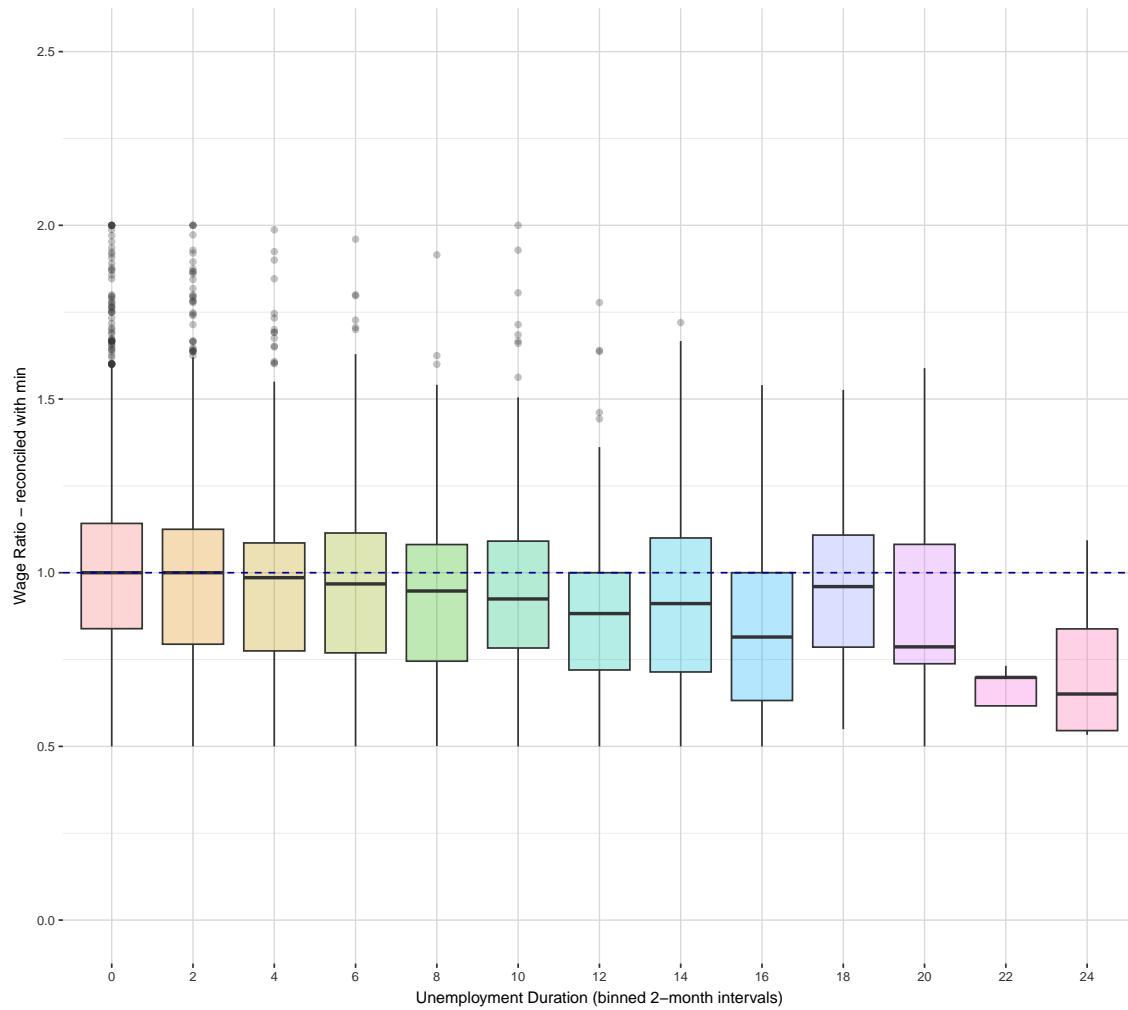
Wage Ratio – Reconciled with min



Reported ratio of current wage to lost wage by unemployment duration

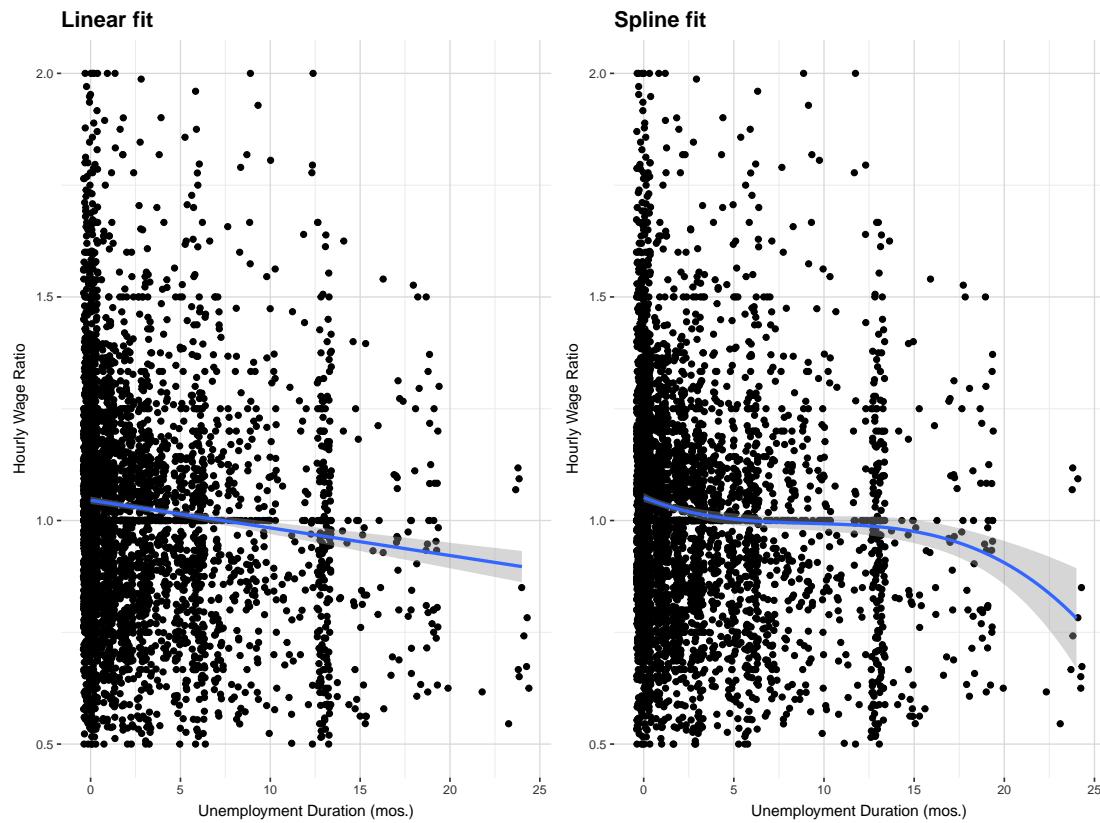
Observations weighted by Displaced Worker Supplement Weights.
Annual data from 2000–2025.
Exclude observations reporting > 96 weeks of unemployment.

Wage Ratio – Reconciled with min



Linear and spline fit to scatter plot of wage ratio vs. unemployment duration in months.

Weighted by Displaced Worker Supplement Weights. Annual data from 2000–2025. Exclude observations reporting > 96 weeks of unemployment and wage ratios below



Regressions (non-uniform sample)

Next, (ignoring for now the non-uniformity of the sample ie. that there are less observations present for higher unemployment durations) I run the following regression (with various modifications to sample and control variables). $W_i = \alpha_i + \beta_1 d_i + \beta_2 UI_i + \beta_3 X_i + \epsilon_i$

where W_i : Ratio of accepted wage to wage at lost job (hourly values).

d_i : Unemployment duration (continuous or binned).

UI_i : Control variable for having used or exhausted unemployment benefits.

X_i : Vector of control variables (sex, age, race (white, black, mixed), marital status (married or not), whether individual used UI benefits, whether individual exhausted UI benefits, education level, and previous wage level).

There are 48 models present with all combinations of the following:

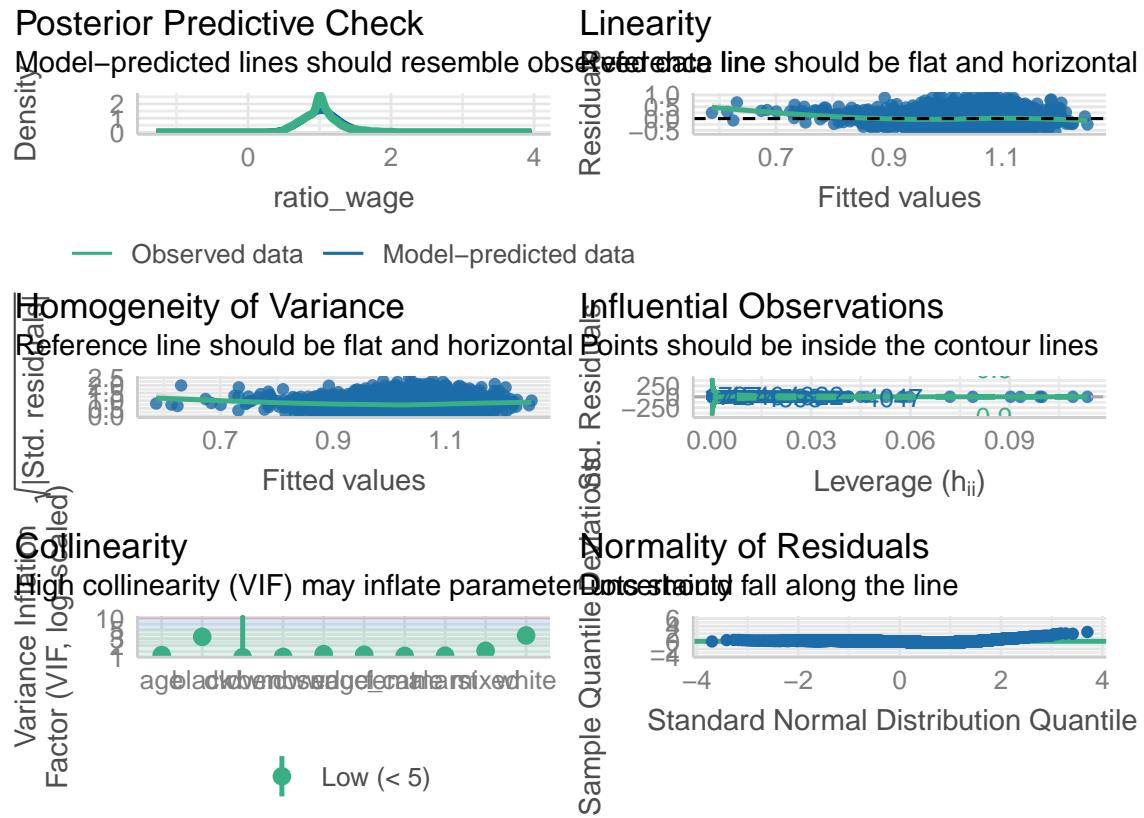
- **Continuous vs. Discrete Treatment Variable (2 alternatives):** Continuous (monthly) versus binned unemployment duration.
- **w. UI vs w. Exhausted UI (3 alternatives):** The data includes a variable for whether individuals USE and/or EXHAUST unemployment benefits. I run the regressions without these UI controls, with control for having used UI, with control for having exhausted UI.
- **w. Controls (2 alternatives):** With or without additional demographic controls (sex, age, race, married, education)

- **w. Wage Level (2 alternatives):** With or without wage level of lost job to control for income. The level of the previous wage likely affects the wage ratio.
- **Outlier clipped sample (2 alternatives):** (As described in the intro section) Remove outliers where the wage ratio is within [0.25, 2.5] and reported unemployment duration is below 96 weeks (~ 2 years).

I include the full set of coefficients (again, apologies for verbose output) in case you find the coefficients on the controls interesting (I think the coefficient on age and holding a Bachelor's degree particularly interesting). But I highlight in blue our main interest in β_1 .

Across all models in the tabs below we see a consistently negative coefficient on unemployment duration (~0.7-1 percentage point increase in the wage ratio for each additional month spent in unemployment). If we look more closely at the performance of our model with continuous unemployment duration, UI use (not exhaustion), all controls, wage levels, and outlier correction we see that the model performs fairly well across various diagnostic tests.

```
## [1] "Continuous U Duration. w. UI Control w. demographic controls (clipped sample)"
```



Continuous UE Duration

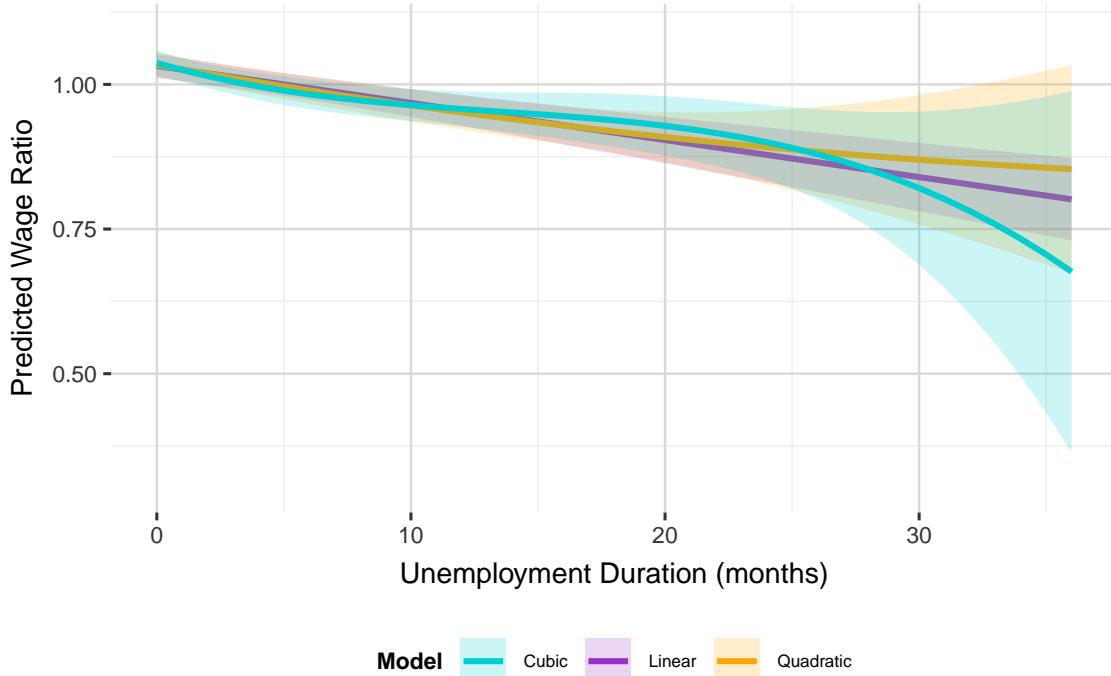
Continuous UE duration treatment is reported in monthly values. A one-unit increase in the treatment variable = 1 additional month of unemployment.

Table 1: Continuous UE Duration w.o Wage Level Control

Table 2: Continuous UE Duration w. Wage Level Control

Predicted Wage Ratios by Unemployment Duration

From non-reweighted regressions: linear, quadratic, and cubic spec



Binned UE Duration

Binned UE duration treatment is reported in bins as indicated in the box plots and code cleaning above.

Additional Considerations

Below I:

Table 3: Binned UE Duration w/o Wage Level Control

	Disc.	Disc. (clipped)	Disc. w. UI	Disc. w. UI (clipped)	Disc. w. exhausted UI	Disc. w. exhausted UI (clipped)	Disc. w. controls	Disc. w. controls (clipped)	Disc. w. UI w. controls	Disc. w. UI w. controls (clipped)	Disc. w. exhausted UI w. controls	Disc. w. exhausted UI w. controls (clipped)
Intercept	1.069*** (0.008)	1.057*** (0.005)	1.069*** (0.008)	1.055*** (0.005)	1.016*** (0.012)	1.016*** (0.008)	1.190*** (0.011)	1.170*** (0.021)	1.190*** (0.011)	1.170*** (0.021)	1.127*** (0.034)	1.176*** (0.023)
Unemployment Duration (Binned)	-0.013*** (0.002)	-0.009*** (0.001)	-0.009*** (0.001)	-0.009*** (0.001)	-0.008*** (0.002)	-0.008*** (0.001)	-0.011*** (0.002)	-0.006*** (0.001)	-0.011*** (0.002)	-0.006*** (0.001)	-0.007*** (0.002)	-0.005*** (0.001)
Received Unemployment Compensation			0.000 (0.001)	0.000 (0.001)				0.000 (0.001)		0.000 (0.001)		0.000 (0.001)
Exhausted Unemployment Compensation					0.001*** (0.000)	0.001*** (0.000)					0.001*** (0.000)	0.001*** (0.000)
Female					0.003 (0.011)	-0.003 (0.007)	0.003 (0.011)	-0.003 (0.007)	0.003 (0.011)	-0.003 (0.007)	0.003 (0.011)	0.003 (0.007)
Age					-0.005*** (0.006)	-0.005*** (0.006)	-0.005*** (0.006)	-0.005*** (0.006)	-0.005*** (0.006)	-0.005*** (0.006)	-0.005*** (0.006)	-0.005*** (0.006)
White					-0.035 (0.023)	-0.052** (0.016)	-0.035 (0.023)	-0.035 (0.023)	-0.035 (0.023)	-0.035 (0.023)	-0.033 (0.023)	-0.033 (0.023)
Black					-0.047 (0.026)	-0.047 (0.018)	-0.047 (0.026)	-0.047 (0.018)	-0.047 (0.026)	-0.047 (0.018)	-0.046 (0.018)	-0.046 (0.018)
Mixed					0.014 (0.010)	-0.070*** (0.007)	0.014 (0.011)	-0.070*** (0.007)	0.014 (0.011)	-0.070*** (0.007)	0.017 (0.011)	-0.068* (0.017)
Married					0.000 (0.001)	0.011 (0.007)	0.004 (0.011)	0.011 (0.007)	0.001 (0.011)	0.005 (0.007)	0.012 (0.011)	0.012 (0.007)
High School					0.006 (0.016)	0.001 (0.016)	0.000 (0.016)	0.000 (0.016)	0.000 (0.016)	0.000 (0.016)	0.005 (0.011)	0.005 (0.011)
educ_catAssociate's					0.003 (0.021)	-0.009 (0.024)	0.003 (0.021)	-0.009 (0.024)	0.000 (0.021)	-0.009 (0.021)	-0.006 (0.021)	-0.006 (0.021)
Bachelor's Degree					0.082*** (0.011)	0.067*** (0.011)	0.082*** (0.011)	0.067*** (0.011)	0.067*** (0.011)	0.067*** (0.011)	0.087*** (0.011)	0.071*** (0.011)
Postgraduate Degree					0.116** (0.045)	0.009 (0.031)	0.116** (0.045)	0.009 (0.031)	0.000 (0.031)	0.000 (0.031)	0.028 (0.031)	0.028 (0.031)
Num.Obs.	4870	4644	8780	4644	4870	4644	4870	4644	4870	4644	4870	4644
R2	0.040	0.041	0.040	0.041	0.040	0.042	0.040	0.041	0.040	0.041	0.040	0.040
R2 Adj.	0.009	0.011	0.009	0.010	0.016	0.022	0.020	0.022	0.020	0.022	0.020	0.020
F	47.638		23.816		40.199		11.165		10.232		12.314	
RMSE	0.37	0.24	0.37	0.24	0.37	0.24	0.37	0.24	0.37	0.24	0.37	0.24

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Table 4: Binned UE Duration w. Wage Level Control

	Disc.	Disc. (clipped)	Disc. w. UI	Disc. w. UI (clipped)	Disc. w. exhausted UI	Disc. w. exhausted UI (clipped)	Disc. w. controls	Disc. w. controls (clipped)	Disc. w. UI w. controls	Disc. w. UI w. controls (clipped)	Disc. w. exhausted UI w. controls	Disc. w. exhausted UI w. controls (clipped)
Intercept	1.069*** (0.012)	1.057*** (0.008)	1.069*** (0.012)	1.055*** (0.008)	1.016*** (0.016)	1.016*** (0.011)	1.272*** (0.011)	1.272*** (0.011)	1.226*** (0.011)	1.226*** (0.011)	1.229*** (0.011)	1.176*** (0.013)
Hourly Wage of Lost Job	-0.009*** (0.006)	-0.006*** (0.006)	-0.009*** (0.006)	-0.006*** (0.006)	-0.009*** (0.006)	-0.006*** (0.006)	-0.007*** (0.006)	-0.007*** (0.006)	-0.011*** (0.006)	-0.011*** (0.006)	-0.011*** (0.006)	-0.007*** (0.006)
Unemployment Duration (Binned)	-0.011*** (0.002)	-0.009*** (0.001)	-0.011*** (0.002)	-0.009*** (0.001)	-0.008*** (0.002)	-0.008*** (0.001)	-0.011*** (0.002)	-0.011*** (0.002)	-0.011*** (0.002)	-0.011*** (0.002)	-0.011*** (0.002)	-0.005*** (0.001)
Received Unemployment Compensation			0.000 (0.001)	0.000 (0.001)			0.000 (0.001)		0.000 (0.001)		0.000 (0.001)	
Exhausted Unemployment Compensation					0.000*** (0.000)	0.000*** (0.000)					0.000*** (0.000)	0.000*** (0.000)
Female					-0.028* (0.011)	-0.023** (0.007)	-0.028* (0.011)	-0.023** (0.007)	-0.023** (0.011)	-0.023** (0.007)	-0.023** (0.011)	-0.023** (0.007)
Age					-0.049*** (0.006)	-0.049*** (0.006)	-0.049*** (0.006)	-0.049*** (0.006)	-0.049*** (0.006)	-0.049*** (0.006)	-0.049*** (0.006)	-0.049*** (0.006)
White					-0.034 (0.023)	-0.052** (0.016)	-0.034 (0.023)	-0.052** (0.016)	-0.034 (0.023)	-0.052** (0.016)	-0.032 (0.023)	-0.049** (0.016)
Black					-0.057* (0.026)	-0.063*** (0.015)	-0.057* (0.026)	-0.063*** (0.015)	-0.057* (0.026)	-0.063*** (0.015)	-0.054** (0.015)	-0.059*** (0.015)
Mixed					0.017 (0.017)	-0.037* (0.027)	0.017 (0.017)	-0.037* (0.027)	0.017 (0.017)	-0.037* (0.027)	0.013 (0.017)	-0.035* (0.017)
Married					0.034* (0.015)	0.019+ (0.015)	0.034* (0.015)	0.019+ (0.015)	0.034* (0.015)	0.019+ (0.015)	0.038* (0.015)	0.022* (0.015)
High School					0.002*** (0.006)	0.025** (0.021)	0.002*** (0.006)	0.025** (0.021)	0.002*** (0.006)	0.025** (0.021)	0.000 (0.006)	0.000 (0.006)
educ_catAssociate's					0.019 (0.021)	0.014 (0.014)	0.019 (0.021)	0.014 (0.014)	0.019 (0.021)	0.014 (0.014)	0.013 (0.014)	0.013* (0.014)
Bachelor's Degree					0.163*** (0.012)	0.122*** (0.012)	0.163*** (0.012)	0.122*** (0.012)	0.163*** (0.012)	0.122*** (0.012)	0.166*** (0.012)	0.124*** (0.012)
Postgraduate Degree					0.246*** (0.045)	0.120*** (0.031)	0.246*** (0.045)	0.120*** (0.031)	0.246*** (0.045)	0.120*** (0.031)	0.250*** (0.045)	0.124*** (0.031)
Num.Obs.	4870	4644	8780	4644	4870	4644	4870	4644	4870	4644	4870	4644
R2	0.047	0.045	0.047	0.045	0.051	0.052	0.049	0.052	0.049	0.052	0.052	0.048
R2 Adj.	0.047	0.045	0.047	0.045	0.050	0.051	0.047	0.052	0.047	0.050	0.050	0.047
F	130.632	86.22	86.995	86.995	30.690	27.774	30.690	27.774	30.690	27.774	30.690	27.774
RMSE	0.37	0.24	0.37	0.24	0.37	0.24	0.37	0.24	0.37	0.24	0.37	0.24

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

1. Show results from some sample reweighting to address the non-uniformity of our cross-sectional data.
2. Histogram of an additional explanatory variable that might be interesting - the tenure of the lost job. How long (in years) did the individual hold the job they lost.
3. Additionally, I show some rough graphs/figures about the sample population (age, education, gender, wage distributions). I am still working on an occupational distribution graph to understand the “skills”/occupational distribution of the sample.

Selection Issues with Non-Random Sample

NOTE: Skip ahead to “Regression Results with Sample Reweighting for Regression Results if you don’t wish to look at the reweighting details below.

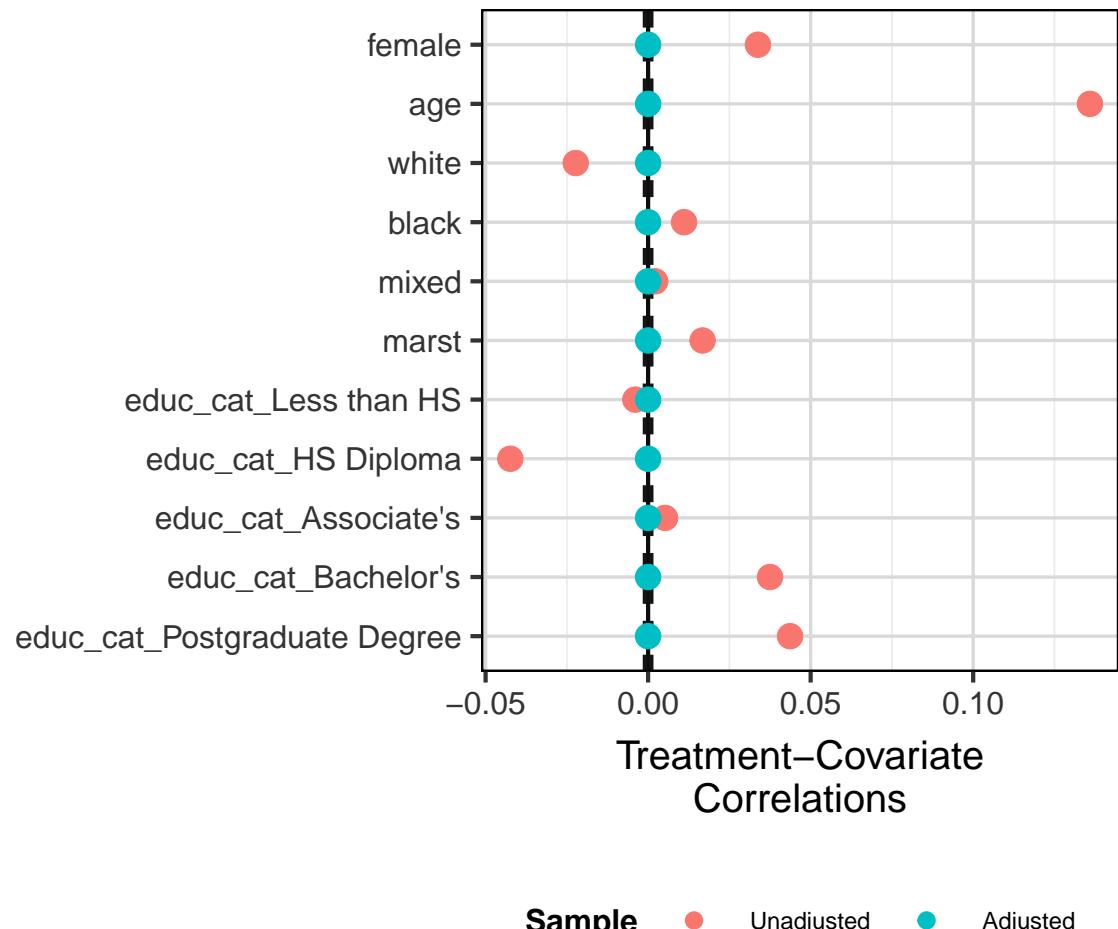
One of the challenges with this data is that the sample grows significantly smaller for higher reported of unemployment duration (see scatter plots in Descriptives section). One option is a sample reweighting (beyond the census weights) to ensure population similarity across bins (below I choose GLM propensity score matching & entropy-balancing) or a Heckman Selection. Again, I include the code below (apologies for verbose output), mainly because I am not yet 100% sure of the implementation as I have never implemented such sample correction in a cross-sectional study). Open to suggestions and corrections :)

Conclusion: With this implementation (which may very well be wrong for now!), the coefficients on unemployment duration remain stable.

Entropy Balancing Entropy balancing simply reweights observations to ensure population matching across the key dependent variable.

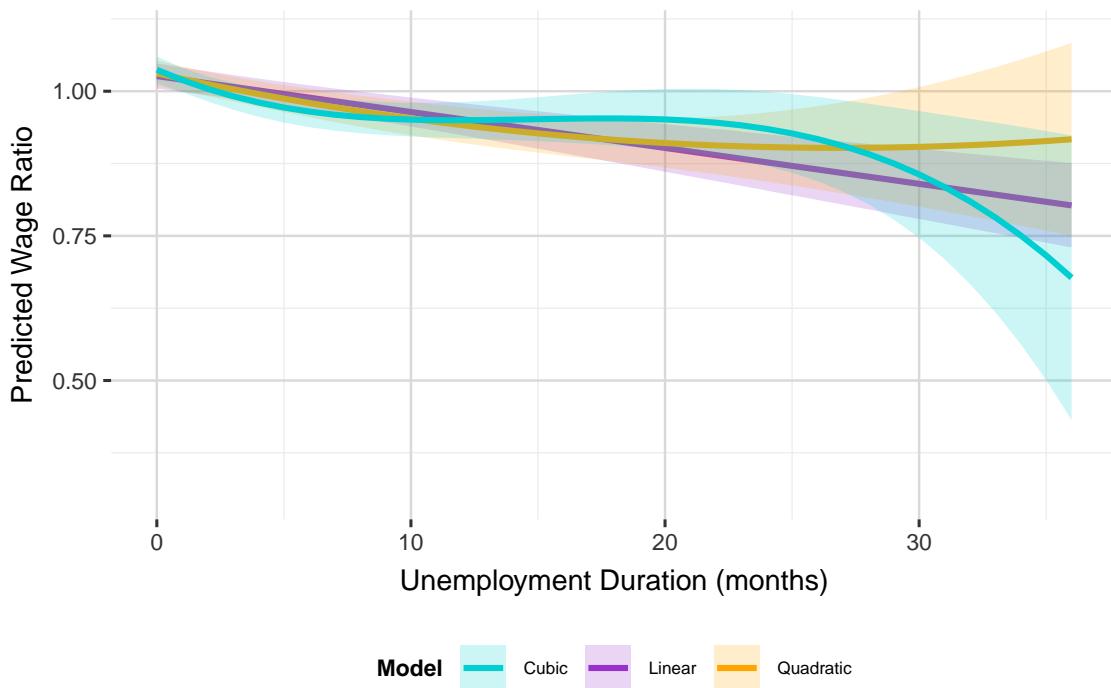
Covariate Balance

Love plot of adjusted and unadj

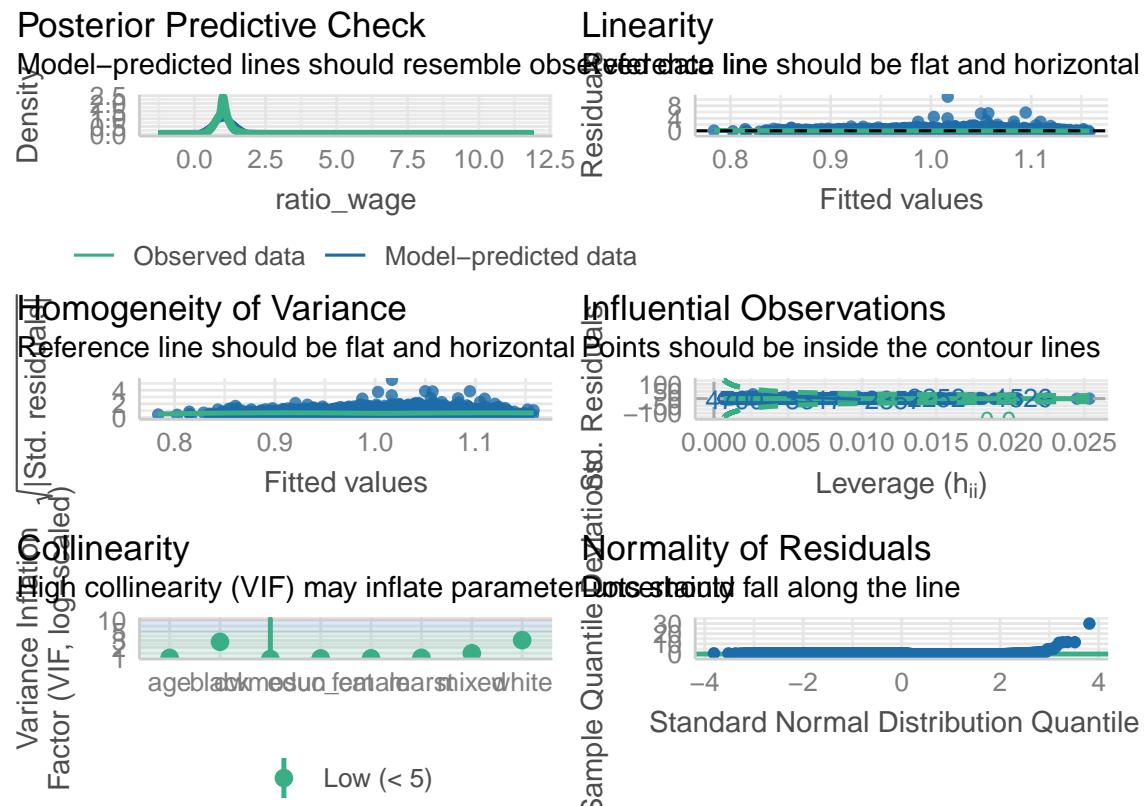


Predicted Wage Ratios by Unemployment Duration (E)

From EB-weighted regressions: linear, quadratic, and cubic specifica

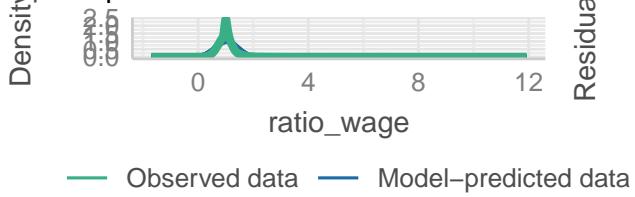


Diagnostic Tests for Entropy-balanced Reweighted Sample



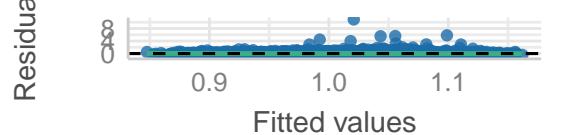
Posterior Predictive Check

Model-predicted lines should resemble observed data



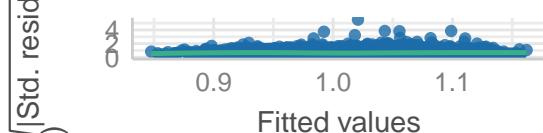
Linearity

Residuals vs fitted values line should be flat and horizontal



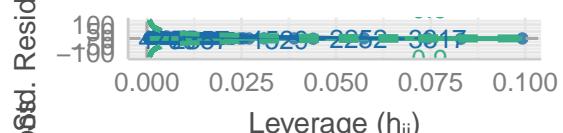
Homogeneity of Variance

Reference line should be flat and horizontal



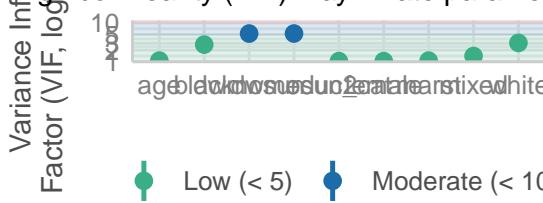
Influential Observations

Points should be inside the contour lines



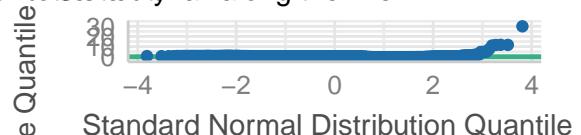
Collinearity

High collinearity (VIF) may inflate parameter estimates and standard errors



Normality of Residuals

Dots should fall along the line



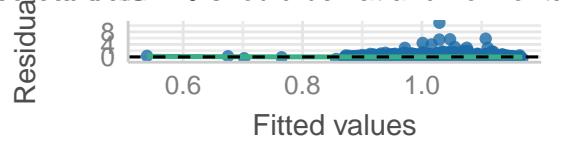
Posterior Predictive Check

Model-predicted lines should resemble observed data



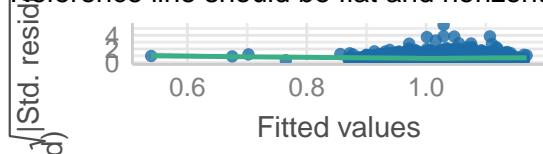
Linearity

Residuals vs fitted values line should be flat and horizontal



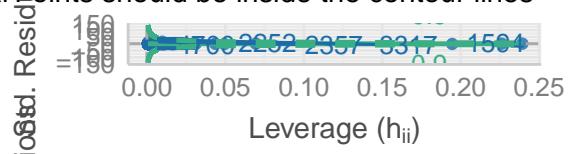
Homogeneity of Variance

Reference line should be flat and horizontal



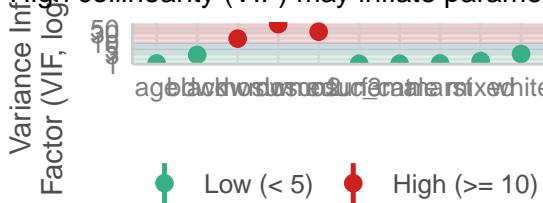
Influential Observations

Points should be inside the contour lines



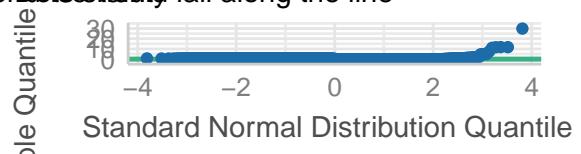
Collinearity

High collinearity (VIF) may inflate parameter estimates and standard errors



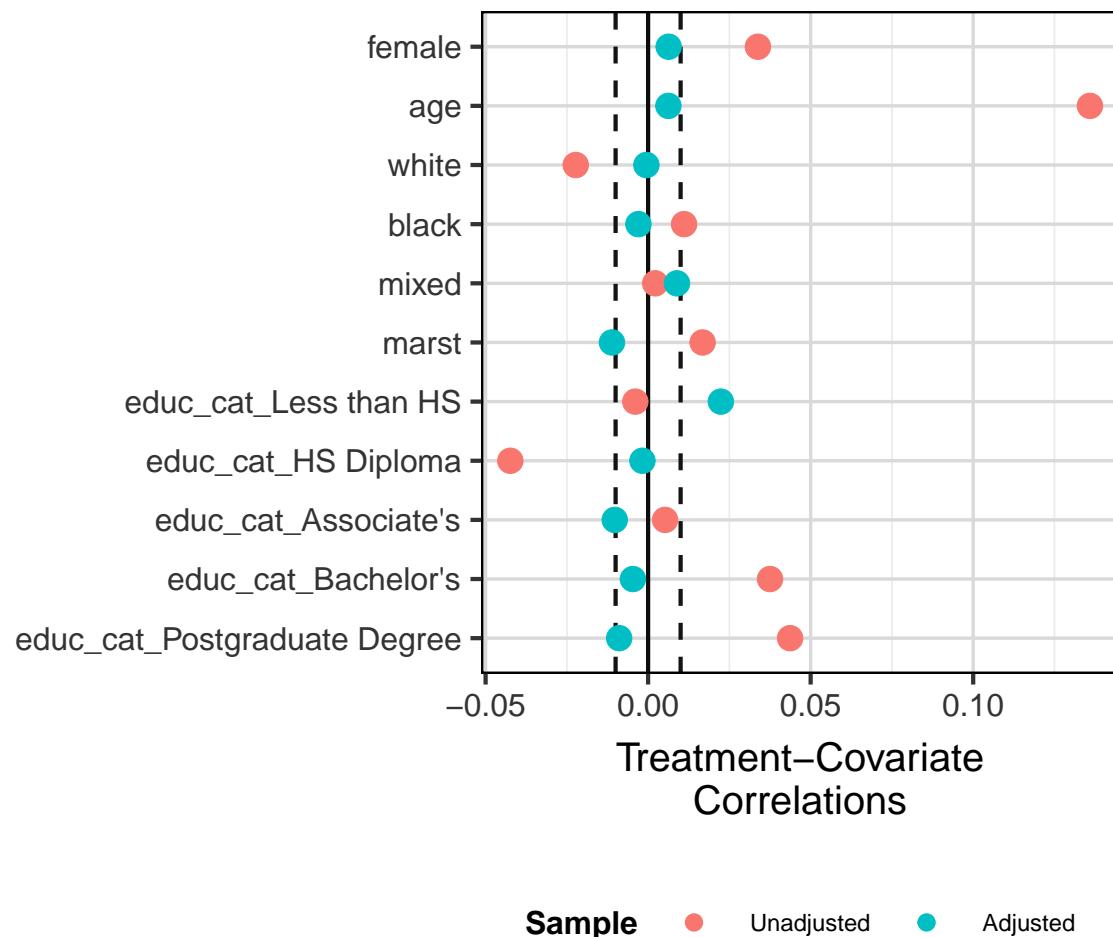
Normality of Residuals

Dots should fall along the line



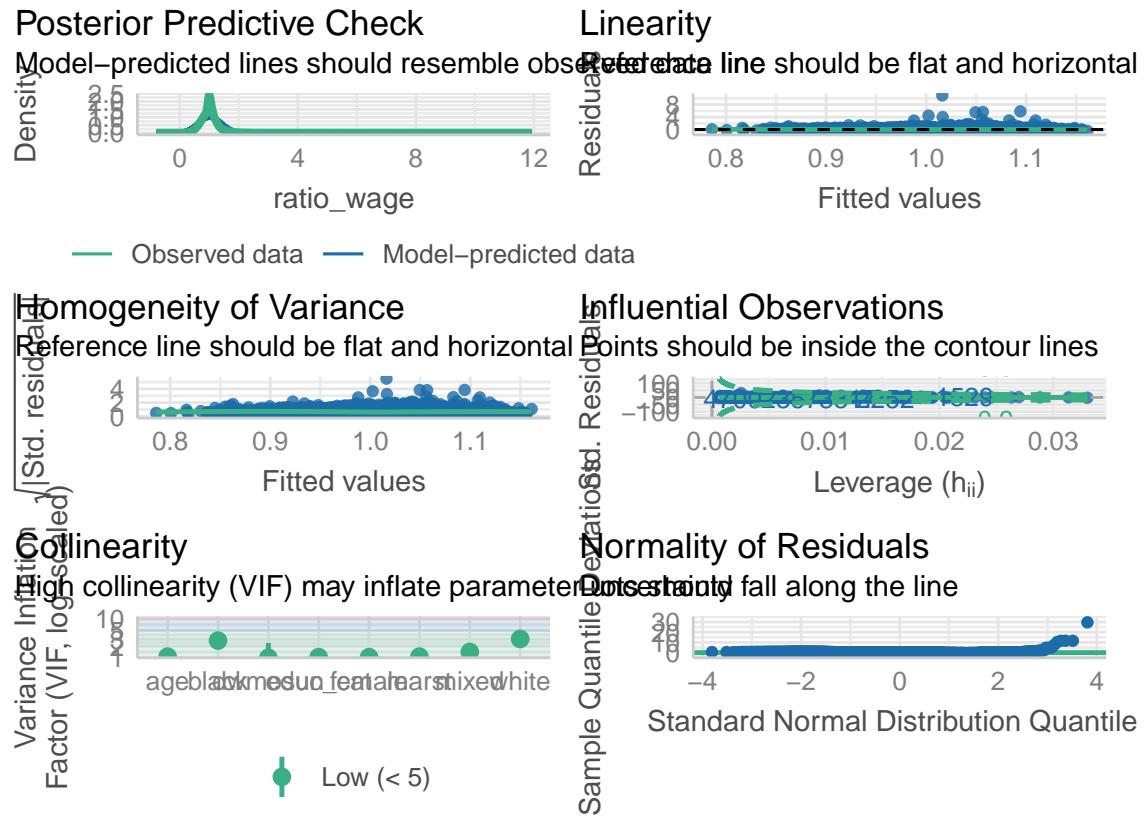
Covariate Balance

Love plot of adjusted and unadj



Sample ● Unadjusted ● Adjusted

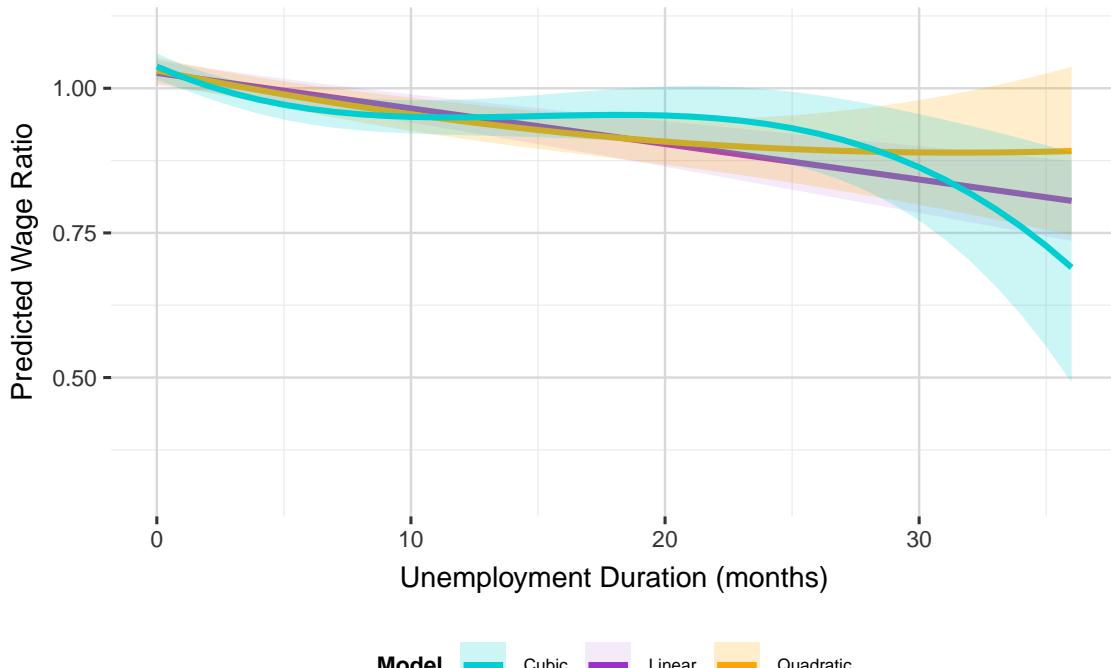
Diagnostic Tests for Propensity Score Matching (GLM) Reweighted Sample



Predicted Reservation Wage using GLM Reweighted Sample

Predicted Wage Ratios by Unemployment Duration (G)

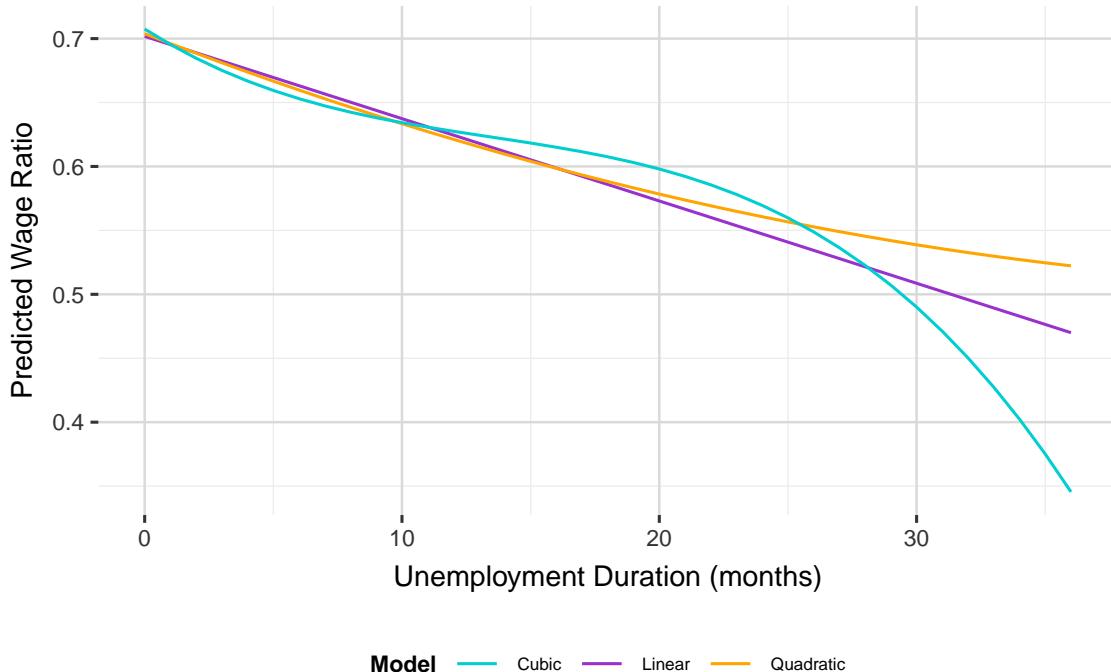
From GLM-weighted regressions: linear, quadratic, and cubic spec



Heckman Selection Another option is a Heckman Selection correction though I do not think this addresses the particular selection concern we have where there are simply less observations in longer unemployment durations.

Predicted Wage Ratios by Unemployment Duration (Heckman Selection)

From Heckman–corrected regressions: linear, quadratic, and cubic models

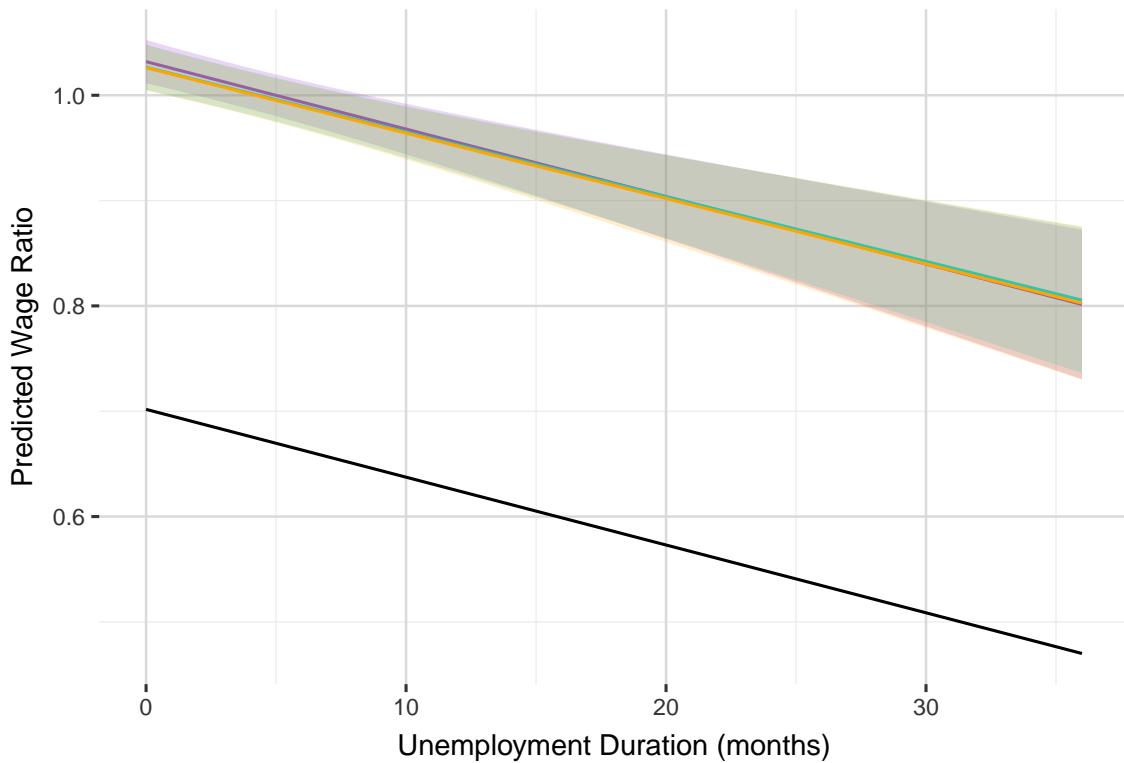


Regression Results with Sample Reweighting

	Heckman Correction	Entropy Balanced Reweight	GLM Reweight
Intercept	1.131*** (0.041)	1.147*** (0.033)	1.143*** (0.033)
Unemployment Duration (Months)	-0.006*** (0.001)	-0.006*** (0.001)	-0.006*** (0.001)
Female	0.018 (0.014)	0.001 (0.011)	0.001 (0.011)
Age	-0.007*** (0.002)	-0.002*** (0.000)	-0.002*** (0.000)
White	-0.162* (0.074)	-0.027 (0.025)	-0.023 (0.025)
Black	-0.125* (0.050)	-0.040 (0.030)	-0.036 (0.030)
Mixed	-0.054 (0.055)	0.003 (0.044)	0.007 (0.044)
Married	0.003 (0.011)	0.005 (0.011)	0.004 (0.011)
High School	-0.014 (0.019)	-0.014 (0.017)	-0.014 (0.017)
XOeduc_catAssociate's	-0.078 (0.064)		
Bachelor's Degree	-0.217 (0.165)	0.054* (0.023)	0.054* (0.023)
Postgraduate Degree	-0.479 (0.330)	0.083+ (0.048)	0.086+ (0.047)
Inverse Mills Ratio	0.870+ (0.479)		
educ_catAssociate's		0.007 (0.022)	0.006 (0.022)
Num.Obs.	4870	4870	4870
R2	0.893	0.014	0.015
R2 Adj.	0.893	0.012	0.013
F		6.487	6.798
RMSE	0.37	0.37	0.37

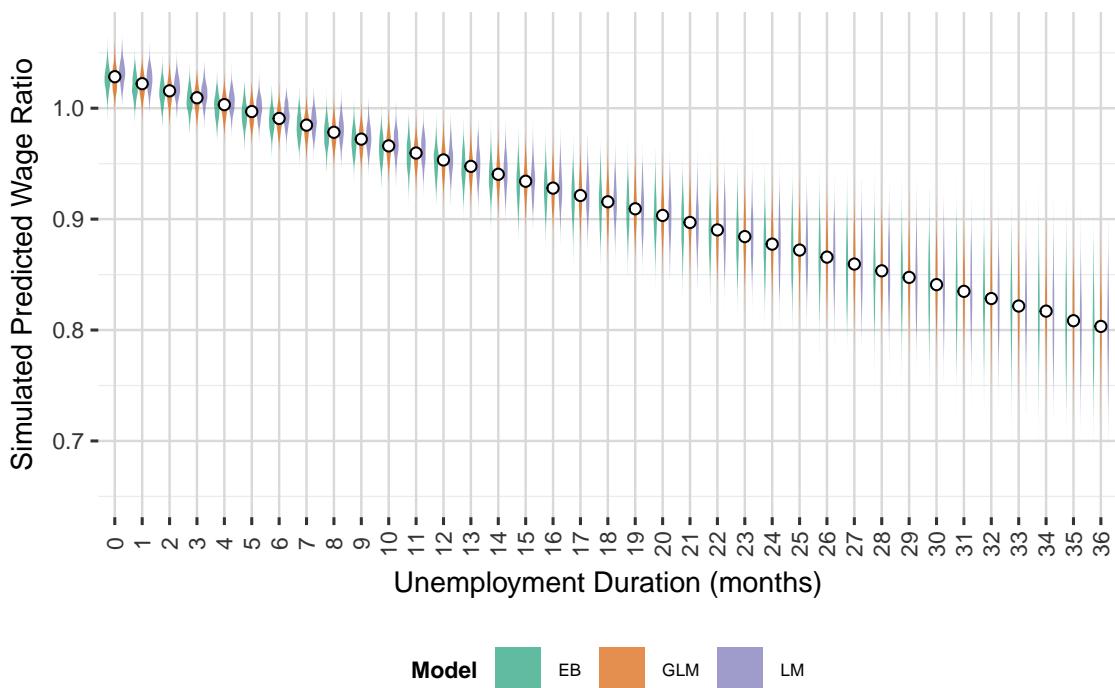
+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Predicted Wage Ratio vs. Unemployment Duration



Simulated Predicted Wage Ratio Distributions by Unen

Violin plots from LM, GLM, and EB model predictions



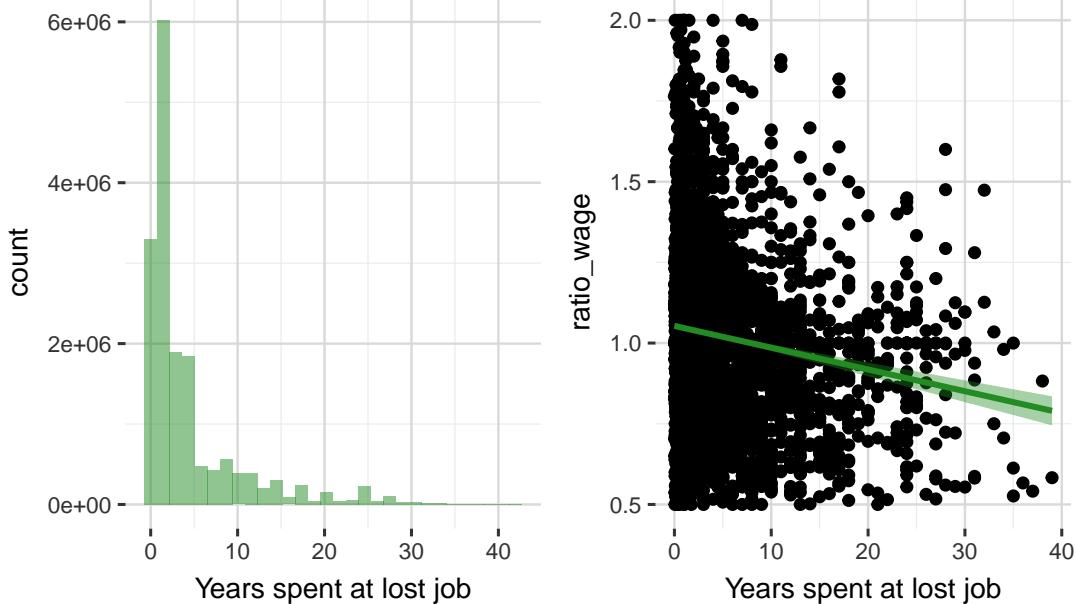
Job Tenure

We have information on the tenure spent at the last job which could impact the result. This could speak to the “adaptability” of individuals. Wage ratio seems to decrease (although not sure if meaningfully) with tenure at previous job.

Tenure at Lost Job (years)

→d Worker Supplement Weights. Annual data from 2000–2025. Exclude observations repor

Histogram of Reported Tenure at Lost Job versus Report



Representation

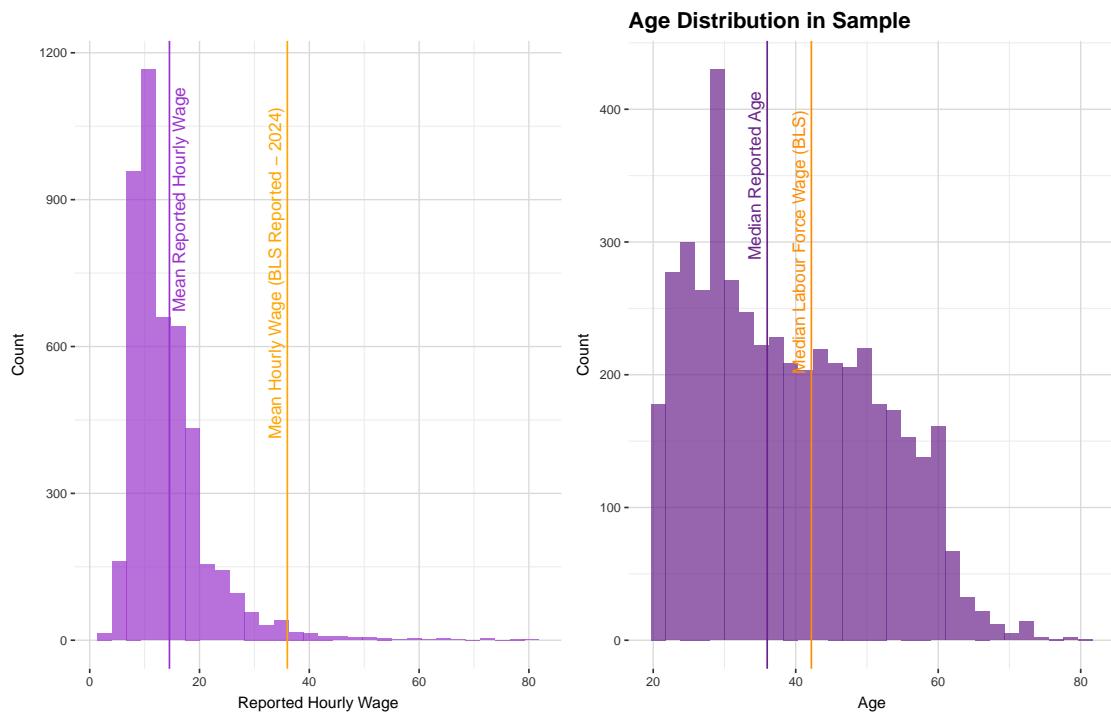
Although the survey does provide sample weights which we use above, it's still likely that those who are laid off might be systematically more susceptible to layoffs (lower-wage, low-skill occupation, male, etc). Below, some (very rough) graphs to indicate what the sample looks like.

Headline result: it seems the sample over-represents below-mean wage earners and women. Age looks reasonably accurate (in relation to a simple median though....have not checked spread). Have not yet checked match to educational attainment. Individuals with only a HS diploma is strong majority in sample - not sure how accurate this is (likely correlated with wage however... so this might be cause for concern and confirm a skewed sample in that sense).

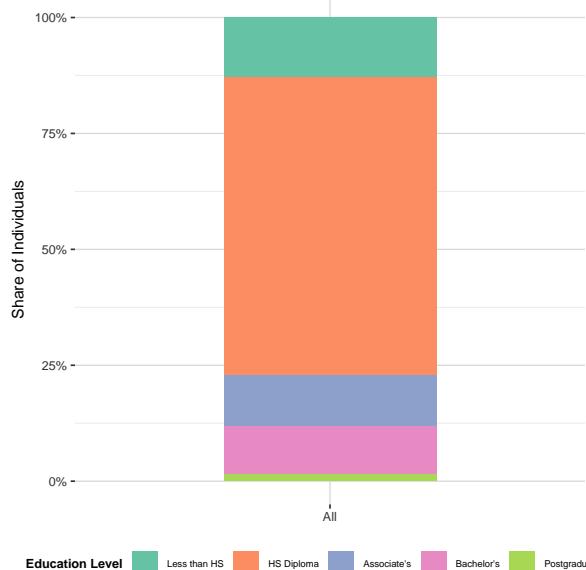
If we wish to pursue this data, I could improve on the below but it will have to do for now.

Preliminary Look at Sample Composition by Age, Wage, Education, Gender, Occupation

Observations weighted by Displaced Worker Supplement Weights. Annual data from 2000–2025. Exclude observations reporting > 96 weeks of unemployment.



Education Attainment in Sample



Gender Share of Sample

Female representation (ie. Female LFPR) as calculated in the sample is ~41% whereas female LFPR in the US is ~57% and their unemployment rate is ~10%.

