#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

# PROJECT: KILLEWALD- GENDER WAGE DISTRIBUTION

# ANALYSIS FILE: MAIN DECOMPOSITION ANALYSES

# AUTHOR: NINO CRICCO

# LAST UPDATED: 09/06/2020 (dmy)

# RUNTIME: 1131.871 sec (~19 min)

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

library(tidyverse)

library(survey)

library(mitools)

library(fastDummies)

library(Hmisc)

library(knitr)

library(kableExtra)

library(weights)

library(gridExtra)

library(mice)

library(janitor)

library(ggrepel)

source("Jobs/helperfunctions.R")

options(survey.lonely.psu = "certainty")

# Loading the imputed dataset

psid\_imp <- read\_csv("clean\_data/psid\_final\_2019.csv")

# Creating the analytic sample

psid\_analytic <- psid\_imp %>%

filter(.imp > 0) %>% # Removing non-imputed data %>%

filter(samp.inc.final == 1) # Selecting sample members

# Creating an object to use in the decompositions using the sample means

means.year <- psid\_analytic %>%

group\_by(year, female, .imp) %>% # For each year, sex, and imputation,

# Compute the weighted means for the variables below, using the weights specified in the function

dplyr::select(lnhrlywage, age, Northeast, Northcentral, South, West, Black, Hispanic, Other, White,

married, prev.married, unmarried, housework,

numkids.0, numkids.1to2, numkids.3plus,

afb.cat\_21minus, afb.cat\_22to27, afb.cat\_27plus,

LessthanHS, HighSchool, SomeCollege, ba.advdeg,

union, govt.job, expf, log.expf, hrswrk.dummy\_50plus, emp.tenure,

occ.pct.female, occ.managers, manuf, perwt) %>%

summarize\_all(list(wmean = ~weighted.mean(., w = perwt))) %>%

mutate("(Intercept)" = 1) %>% # Creating an intercept column

filter(.imp != 0) %>% # Filtering out the non-imputed data

dplyr::select(year, female, "(Intercept)", everything()) %>% # Ordering the columns

dplyr::select(-c(.imp, perwt\_wmean)) %>% # Removing weight and imputation column

ungroup() %>%

group\_by(year, female) %>% # For each year and female, compute means mean across imputations

summarise\_all(mean)

# Setting names to the weighted means table

names(means.year) <- gsub("\_wmean", "", names(means.year))

# We use this table of year-and-sex specific means for the decomposition analyses

# Each row is a year X sex combination, columns are year-sex specific means of selected variables

means.year\_table <- means.year %>%

filter(year %in% c(1981, 1991, 2001, 2011, 2019))

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

# TABLE 1: DESCRIPTIVE STATISTICS

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

# First summarize the sample size in our analytic sample within each year

sampsize <- psid\_analytic %>%

filter(.imp == 1) %>% # Selecting data from a single imputation

group\_by(year, female) %>% # Getting the n's by sex and year

filter(year %in% c(1981, 1991, 2001, 2011, 2019)) %>%

summarise(n = n())

# Creating table of descriptive statistics from our table of year-and-sex specific means

t1\_means <- means.year\_table %>%

# Joining table of sample sizes with table of year-and-sex specific means

left\_join(., sampsize) %>%

# Creating column of actual hourly wages as opposed to logged hourly wages for desc stats table

mutate(outcome\_exp = exp(lnhrlywage)) %>%

# For categorical variables, convert decimals to percentage terms

mutate\_at(vars(Northeast, Northcentral,South, West, Black, Hispanic, Other, White,

married, prev.married, unmarried,

numkids.0, numkids.1to2, numkids.3plus,

afb.cat\_21minus, afb.cat\_22to27, afb.cat\_27plus,

hrswrk.dummy\_50plus, LessthanHS, HighSchool, SomeCollege, ba.advdeg,

union, govt.job, occ.pct.female, occ.managers, manuf),

function (x) x\*100) %>%

# Rounding all values to the first decimal point

round(digits = 1) %>%

# Setting the order of the variables, removing intercept, logged hourly wage, and log of experience column

dplyr::select(year, female, outcome\_exp, everything(), -c("(Intercept)", "lnhrlywage", "log.expf")) %>%

# Reshaping output so that each row is a variable and each column is a year-sex specific mean

gather(vars, means, -c(year, female)) %>%

pivot\_wider(id\_cols = c(vars),

names\_from = c(female, year), values\_from = means) %>%

# Re-ordering columns to have var names first, year-sex specific means for men first, then women

dplyr::select(vars, starts\_with("0"), everything()) %>%

as.data.frame() # Turning object to dataframe to change column and row names

# Setting column names for the descriptive stats table

colnames(t1\_means) <- (c("vars", "1981", "1991", "2001", "2011", "2019",

"1981 ", "1991 ", "2001 ", "2011 ", "2019 "))

# Setting row names for the descriptive stats table

rownames(t1\_means) <- c("Hourly Earnings", "Age", "Northeast", "Northcentral",

"South", "West", "Black", "Latinx", "Other", "White", "Married",

"Previously Married", "Unmarried", "Housework", "0", "1-2", "3+",

"<= 21", "22 to 27", "28 +",

"Less than High School", "High School", "Some College", "BA/Advanced Degree",

"Unionized Job", "Government Job", "Full-Time", "Overwork",

"Employer Tenure", "Percent Female in Occupation", "Management Occupation", "Manufacturing",

"Sample Size")

# Generating the descriptive statistics table

knitr::kable(t1\_means %>% dplyr::select(-vars), booktabs = T, format = "latex",

caption = "Sample Descriptive Statistics") %>%

add\_header\_above(c(" ", "Male" = 5, "Female" = 5)) %>%

footnote("Descriptive statistics show weighted averages for the analytic sample by year and sex. The sample consists of PSID heads and wives aged 30 to 60 who are full-time workers, defined as working 35 hours a week or more for at least 26 weeks a year and reporting non-zero wages. It excludes individuals who report being self-employed and individuals employed in agriculture or the military",

threeparttable = T) %>%

pack\_rows("Region", 3, 6, bold = F) %>% pack\_rows("Race", 7, 10, bold = F) %>%

pack\_rows("Marital Status", 11, 13, bold = F) %>% pack\_rows("Number of Kids", 15, 17, bold = F) %>%

pack\_rows("Age at First Birth", 18, 20, bold = F) %>% pack\_rows("Education", 21, 24, bold = F)

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

# FIGURE 1: Changes in Gender Pay Gap Among FT Workers

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

fig1\_values <- means.year\_table %>%

mutate(outcome\_exp = exp(lnhrlywage)) %>%

dplyr::select(year, female, outcome\_exp) %>%

pivot\_wider(names\_from = female, values\_from = outcome\_exp) %>%

rename(Women = "1", Men = "0") %>%

mutate(ratio = (Women/Men) \* 100,

year = case\_when(year == 1981 ~ 1980,

year == 1991 ~ 1990,

year == 2001 ~ 2000,

year == 2011 ~ 2010,

year == 2019 ~ 2018))

fig1 <- fig1\_values %>%

ggplot(aes(x = year, y = ratio)) +

geom\_point() +

theme\_bw() +

labs(title = "Changes in Gender Pay Gap Among Full-time Workers, 1980-2018",

y = "Mean Hrly Wage Women / Men",

x = "") +

geom\_line() +

theme(plot.title = element\_text(hjust = 0.5), legend.position = "bottom") +

scale\_x\_continuous(breaks=c(1980, 1990, 2000, 2010, 2018)) +

geom\_hline(yintercept = 100, linetype = "dashed") +

ylim(50, 100)

ggsave(plot = fig1, "/Users/ninocricco/Desktop/PSID\_CodeReview/output/fig1.jpg",

width = 7, height = 6, units = "in", device='jpeg', dpi=700)

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

# FIGURE 2: Changes in Levels, Family Variables

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

# For the means for figure 2, we compute new means: while the "married" mean is the same, we

# use the continuous measure for number of children and we compute only age at first birth

# among those who have children. We compute these sets of means both for our analytic sample

# and for the full PSID sample

fig2\_values <- psid\_analytic %>%

filter(year %in% c(1981, 1991, 2001, 2011, 2019)) %>%

group\_by(year, female, .imp) %>% # For each year, sex, and imputation,

# Compute the weighted means for the variables below, using the weights specified in the function

dplyr::select(married, num.children.synth, perwt) %>%

summarize\_all(list(wmean = ~weighted.mean(., w = perwt))) %>%

dplyr::select(year, female, everything()) %>% # Ordering the columns

dplyr::select(-c(.imp, perwt\_wmean)) %>% # Removing weight and imputation column

ungroup() %>%

group\_by(year, female) %>% # For each year and female, compute means mean across imputations

summarise\_all(mean) %>%

dplyr::select(year, female, married\_wmean, num.children.synth\_wmean) %>%

ungroup() %>%

# Here, we compute the mean age at first birth only among those who have children

left\_join(psid\_analytic %>%

group\_by(year, female, .imp) %>% # For each year, sex, and imputation,

# Compute the weighted means for the variables below, using the weights specified in the function

dplyr::select(afb.final, perwt) %>%

filter(afb.final != 9999) %>% # filtering out those with no children

summarize\_all(list(wmean = ~weighted.mean(., w = perwt))) %>%

dplyr::select(year, female, everything()) %>% # Ordering the columns

dplyr::select(-c(.imp, perwt\_wmean)) %>% # Removing weight and imputation column

ungroup() %>%

group\_by(year, female) %>% # For each year and female, compute means mean across imputations

summarise\_all(mean) %>%

ungroup(), by = c("year", "female")) %>%

mutate(Sex = case\_when(female == 0 ~ "Men",

female == 1 ~ "Women"),

year = factor(year)) %>%

dplyr::select(-female) %>%

rename("Age at First Birth" = "afb.final\_wmean",

"Percent Married" = "married\_wmean",

"Number of Children" = "num.children.synth\_wmean") %>%

gather(key, value, -c(year, Sex)) %>%

mutate(sample = "Full-Time Workers") %>%

# Here, we compute these same sets of means but in the full PSID sample

rbind(psid\_imp %>%

filter(.imp > 0) %>%

filter(year %in% c(1981, 1991, 2001, 2011, 2019)) %>%

group\_by(year, female, .imp) %>% # For each year, sex, and imputation,

# Compute the weighted means for the variables below, using the weights specified in the function

dplyr::select(married, num.children.synth, perwt) %>%

summarize\_all(list(wmean = ~weighted.mean(., w = perwt))) %>%

ungroup() %>%

dplyr::select(year, female, everything()) %>% # Ordering the columns

dplyr::select(-c(.imp, perwt\_wmean)) %>% # Removing weight and imputation column

group\_by(year, female) %>% # For each year and female, compute means mean across imputations

summarise\_all(mean) %>%

ungroup() %>%

dplyr::select(year, female, married\_wmean, num.children.synth\_wmean) %>%

left\_join(psid\_imp %>%

filter(.imp > 0) %>%

group\_by(year, female, .imp) %>% # For each year, sex, and imputation,

# Compute the weighted means for the variables below, using the weights specified in the function

dplyr::select(afb.final, perwt) %>%

filter(afb.final != 9999) %>%

summarize\_all(list(wmean = ~weighted.mean(., w = perwt))) %>%

ungroup() %>%

dplyr::select(year, female, everything()) %>% # Ordering the columns

dplyr::select(-c(.imp, perwt\_wmean)) %>% # Removing weight and imputation column

group\_by(year, female) %>% # For each year and female, compute means mean across imputations

summarise\_all(mean) %>%

ungroup(), by = c("year", "female")) %>%

mutate(Sex = case\_when(female == 0 ~ "Men",

female == 1 ~ "Women"),

year = factor(year)) %>%

dplyr::select(-female) %>%

rename("Age at First Birth" = "afb.final\_wmean",

"Percent Married" = "married\_wmean",

"Number of Children" = "num.children.synth\_wmean") %>%

gather(key, value, -c(year, Sex)) %>%

mutate(sample = "Full PSID Sample")) %>%

mutate(year = case\_when(year == 1981 ~ 1980,

year == 1991 ~ 1990,

year == 2001 ~ 2000,

year == 2011 ~ 2010,

year == 2019 ~ 2018))

fig2 <- fig2\_values %>%

ggplot(aes(y = value, x = year, color = Sex, linetype = sample, alpha = sample, shape = sample)) +

geom\_point() +

geom\_line() +

theme\_bw() +

labs(title = "Changes in Family Formation, 1980-2018",

y = "Mean Number of Children") +

labs(title = "Changes in Mean Levels, Family Variables, 1980-2018") +

facet\_wrap(~key, scales = "free") +

theme(plot.title = element\_text(hjust = 0.5, face = "bold", size = 12), legend.position = "bottom") +

scale\_linetype\_manual(values = c(3, 1)) +

guides(linetype = FALSE) +

scale\_x\_continuous(breaks=c(1980, 1990, 2000, 2010, 2018)) +

scale\_alpha\_manual(values = c(0.8, 1))

# This line outputs the values for figure 2 in a more readable wide format

# fig2\_values %>% pivot\_wider(names\_from = key, values\_from = value)

ggsave(plot = fig2, "/Users/ninocricco/Desktop/PSID\_CodeReview/output/fig2.jpg",

width = 6, height = 6, units = "in", device='jpeg', dpi=700)

# Next, we run the regressions

cov.years <- c(1981, 2019) # sets the reference years for the decomposition

psid\_t1 <- psid\_analytic %>%

filter(year %in% cov.years) # Selecting focal years

# For each year by sex group, we create list objects where each element contains an imputed dataset

# We then create a survey design object that takes into account the uncertainty in the PSID's sampling design

# For Men in t1

mice.out.imp.mt1 <- imputationList(psid\_t1 %>%

filter(.imp > 0) %>%

filter(year == cov.years[1], female == 0) %>%

group\_split(.imp))

design.mt1 <-svydesign(ids = ~samp\_error\_cluster, strata = ~samp\_error\_stratum,

data = mice.out.imp.mt1, weights = ~perwt, nest = T)

# For Men in t2

mice.out.imp.mt2 <- imputationList(psid\_t1 %>%

filter(.imp > 0) %>%

filter(year == cov.years[2], female == 0) %>%

group\_split(.imp))

design.mt2 <-svydesign(ids = ~samp\_error\_cluster, strata = ~samp\_error\_stratum,

data = mice.out.imp.mt2, weights = ~perwt, nest = T)

# For Women in t1

mice.out.imp.ft1 <- imputationList(psid\_t1 %>%

filter(.imp > 0) %>%

filter(year == cov.years[1], female == 1) %>%

group\_split(.imp))

design.ft1 <-svydesign(ids = ~samp\_error\_cluster, strata = ~samp\_error\_stratum,

data = mice.out.imp.ft1, weights = ~perwt, nest = T)

# For Women in t2

mice.out.imp.ft2 <- imputationList(psid\_t1 %>%

filter(.imp > 0) %>%

filter(year == cov.years[2], female == 1) %>%

group\_split(.imp))

design.ft2 <-svydesign(ids = ~samp\_error\_cluster, strata = ~samp\_error\_stratum,

data = mice.out.imp.ft2, weights = ~perwt, nest = T)

# We then run the regressions corresponding to each of our models on each year by sex group

# Using the survey design object on each imputed dataset, then combine the regression results

# across imputations

# Family model for Men in t1

fam.mt1 <- with(design.mt1, svyglm(

lnhrlywage ~ married + prev.married + numkids.3plus +

afb.cat\_21minus + afb.cat\_22to27 + afb.cat\_27plus

))

# Combining the regression results across imputations

fam.mt1 <- MIcombine((fam.mt1))

# From the combined results object, create a dataframe that contains

# the coefficients and standard errors

fam.mt1 <- bind\_rows(coef(fam.mt1), vcov::se(fam.mt1)) %>%

# We add a row to the dataframe that labels the year and sex groups

mutate(year = cov.years[1], female = 0, estimate = c("coef", "se"))

# Repeating the same procedure for each year by sex group

# Family model for Men in t2

fam.mt2 <- with(design.mt2, svyglm(

lnhrlywage ~ married + prev.married + numkids.3plus +

afb.cat\_21minus + afb.cat\_22to27 + afb.cat\_27plus

))

fam.mt2 <- MIcombine((fam.mt2))

fam.mt2 <- bind\_rows(coef(fam.mt2), vcov::se(fam.mt2)) %>%

mutate(year = cov.years[2], female = 0, estimate = c("coef", "se"))

# Family model for women in t1

fam.ft1 <- with(design.ft1, svyglm(

lnhrlywage ~ married + prev.married + numkids.3plus +

afb.cat\_21minus + afb.cat\_22to27 + afb.cat\_27plus

))

fam.ft1 <- MIcombine((fam.ft1))

fam.ft1 <- bind\_rows(coef(fam.ft1), vcov::se(fam.ft1)) %>%

mutate(year = cov.years[1], female = 1, estimate = c("coef", "se"))

# Family model for women in t2

fam.ft2 <- with(design.ft2, svyglm(

lnhrlywage ~ married + prev.married + numkids.3plus +

afb.cat\_21minus + afb.cat\_22to27 + afb.cat\_27plus

))

fam.ft2 <- MIcombine((fam.ft2))

fam.ft2 <- bind\_rows(coef(fam.ft2), vcov::se(fam.ft2)) %>%

mutate(year = cov.years[2], female = 1, estimate = c("coef", "se"))

# We then create a dataframe that combines the year by sex

# coefficients and standard errors for the family model

fam.coefs <- bind\_rows(fam.mt1, fam.ft1, fam.mt2, fam.ft2)

# We repeat the same procedure with our full model

# Full model for men in t1

full.mt1 <- with(design.mt1, svyglm(

lnhrlywage ~ Northeast + Northcentral + South + Black + Hispanic + Other +

married + prev.married +

housework +

numkids.3plus + afb.cat\_21minus + afb.cat\_22to27 + afb.cat\_27plus +

HighSchool + SomeCollege + ba.advdeg + union + govt.job +

log.expf + hrswrk.dummy\_50plus +

emp.tenure +

occ.pct.female + occ.managers + manuf))

full.mt1 <- MIcombine((full.mt1))

full.mt1 <- bind\_rows(coef(full.mt1), vcov::se(full.mt1)) %>%

mutate(year = cov.years[1], female = 0, estimate = c("coef", "se"))

# Full model for men in t2

full.mt2 <- with(design.mt2, svyglm(

lnhrlywage ~ Northeast + Northcentral + South + Black + Hispanic + Other +

married + prev.married +

housework +

numkids.3plus + afb.cat\_21minus + afb.cat\_22to27 + afb.cat\_27plus +

HighSchool + SomeCollege + ba.advdeg + union + govt.job +

log.expf + hrswrk.dummy\_50plus +

emp.tenure +

occ.pct.female + occ.managers + manuf))

full.mt2 <- MIcombine((full.mt2))

full.mt2 <- bind\_rows(coef(full.mt2), vcov::se(full.mt2)) %>%

mutate(year = cov.years[2], female = 0, estimate = c("coef", "se"))

# Full model for women in t1

full.ft1 <- with(design.ft1, svyglm(

lnhrlywage ~ Northeast + Northcentral + South + Black + Hispanic + Other +

married + prev.married +

housework +

numkids.3plus + afb.cat\_21minus + afb.cat\_22to27 + afb.cat\_27plus +

HighSchool + SomeCollege + ba.advdeg + union + govt.job +

log.expf + hrswrk.dummy\_50plus +

emp.tenure +

occ.pct.female + occ.managers + manuf))

full.ft1 <- MIcombine((full.ft1))

full.ft1 <- bind\_rows(coef(full.ft1), vcov::se(full.ft1)) %>%

mutate(year = cov.years[1], female = 1, estimate = c("coef", "se"))

# Full model for women in t2

full.ft2 <- with(design.ft2, svyglm(

lnhrlywage ~ Northeast + Northcentral + South + Black + Hispanic + Other +

married + prev.married +

housework +

numkids.3plus + afb.cat\_21minus + afb.cat\_22to27 + afb.cat\_27plus +

HighSchool + SomeCollege + ba.advdeg + union + govt.job +

log.expf + hrswrk.dummy\_50plus +

emp.tenure +

occ.pct.female + occ.managers + manuf))

full.ft2 <- MIcombine((full.ft2))

full.ft2 <- bind\_rows(coef(full.ft2), vcov::se(full.ft2)) %>%

mutate(year = cov.years[2], female = 1, estimate = c("coef", "se"))

# Creatinig a dataframe that combines the year by sex coefs and se's for the full model

full.coefs <- bind\_rows(full.mt1, full.ft1, full.mt2, full.ft2)

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

# TABLE 2: FAMILY COEFFICIENTS FOR THE FAMILY AND FULL MODELS

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

# We create a table that combines the family coefficients for each model

# Each row is a quantity for each model-estimate-sex-year, and each row is a family variable

t2\_coefs <- fam.coefs %>% mutate(model = "Family") %>%

dplyr::select(model, estimate, female, year, married, prev.married,

numkids.3plus, starts\_with("afb.cat\_")) %>%

mutate(housework = NA) %>%

# Combining the family coefficients from the family model

# with the family coefficients from the full model

rbind(full.coefs %>% mutate(model = "Full") %>%

dplyr::select(model, estimate, female, year, married, prev.married,

numkids.3plus, starts\_with("afb.cat\_"), housework)) %>%

# The following two lines reshape the data such that we have separate coefficient and se

# columns for each model-sex-year-variable combination

gather(variable, value, -c(model, female, year, estimate)) %>%

pivot\_wider(names\_from = estimate, values\_from = value) %>%

# Rounding the estimates of the coefficients and standard errors

mutate(coef = round(coef, digits = 3),

se = round(se, digits = 3),

se = paste("(", se, ")", sep = "")) %>%

# The following two lines reshape the data such that each row contains the estimates

# (coef or se) for each model and variable for each year-sex group

gather(estimate, value, -c(model, female, year, variable)) %>%

pivot\_wider(names\_from = c(female, year), values\_from = value) %>%

# This line rearranges the rows to group the coefficient and standard errors for each model-variable combination

arrange(model, variable) %>%

# Creating a column that labels each characteristic

mutate(Characteristic = c("Age First Birth: <= 21", "", "Age First Birth: 22 to 27", "",

"Age First Birth: 28 +", "", "Housework", "", "Married", "",

"Number of Kids: 3+", "", "Previously Married", "",

"Age First Birth: <= 21", "", "Age First Birth: 22 to 27", "",

"Age First Birth: 28 +", "", "Housework", "", "Married", "",

"Number of Kids: 3+", "", "Previously Married", "")) %>%

dplyr::select(model, Characteristic, starts\_with("0"), everything(), -c(estimate, variable))

# This sets the column names for the table

colnames(t2\_coefs) <- (c("model", "Characteristic", "1980", "2018", "1980 ", "2018 "))

# Generating the table output

knitr::kable(t2\_coefs %>% dplyr::select(-model), booktabs = T, format = "latex",

caption = "Family Coefficients in Year-and-Sex specific regressions") %>%

add\_header\_above(c(" ", "Male" = 2, "Female" = 2)) %>%

footnote("Table shows coefficients for family characteristics from sex-and-year-specific regressions on hourly wages. The Full model includes controls for demographic characteristics (region, race), education, and job traits (years of experience, employer tenure, overwork, percent female in occupation, whether job is government job or covered by union, and whether employee works more than 50 hours a week)",

threeparttable = T) %>%

pack\_rows("Family Model", 1, 14, bold = T) %>% pack\_rows("Full Model", 15, 28, bold = T)

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

# TABLE A1: REGRESSION COEFFICIENTS FOR ALL VARIABLES IN EACH MODEL

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

# Appendix Table: Regression coefficients, full table

coeftable <- fam.coefs %>% #filter(estimate == "coef") %>%

mutate(model = "Family") %>%

gather(key, value, -c(year, female, model ,estimate)) %>%

rbind(full.coefs %>% #filter(estimate == "coef") %>%

mutate(model = "Full") %>%

gather(key, value, -c(year, female, model, estimate))) %>%

pivot\_wider(names\_from = estimate, values\_from = value) %>%

mutate(coef = round(coef, digits = 3),

se = round(se, digits = 3),

se = paste("(", se, ")", sep = "")) %>%

gather(estimate, value, -c(model, female, year, key)) %>%

pivot\_wider(names\_from = c(female, year), values\_from = value) %>%

arrange(model, key) %>%

mutate(Characteristic = c("Intercept", "", "Age First Birth: <= 21", "", "Age First Birth: 22 to 27", "",

"Age First Birth: 28 +", "", "Married", "",

"Number of Kids: 3+", "", "Previously Married", "",

"Intercept", "", "Age First Birth: <= 21", "", "Age First Birth: 22 to 27", "",

"Age First Birth: 28 +", "", "BA/Advanced Degree", "", "Black", "",

"Employer Tenure", "", "Government Job" , "", "High School", "",

"Hispanic", "", "Housework", "", "Overwork", "", "Log, Full-time Experience", "",

"Manufacturing Occupation", "", "Married", "", "Northcentral", "", "Northeast", "",

"Number of Kids: 3+", "", "Management Occupation", "", "Percent Female in Occupation", "",

"Other", "", "Previously Married", "", "Some College", "", "South", "", "Union Job", "")) %>%

dplyr::select(model, Characteristic, starts\_with("0"), everything(), -c(estimate))

knitr::kable(coeftable %>% dplyr::select(-c(key,model)), booktabs = T, format = "latex",

caption = "Coefficients in Year-and-Sex specific regressions", digits = 2) %>%

add\_header\_above(c(" ", "Male" = 2, "Female" = 2)) %>%

footnote("Table shows coefficients for all characteristics from sex-and-year-specific regressions on hourly wages",

threeparttable = T) %>%

pack\_rows("Family Model", 1, 14, bold = T) %>% pack\_rows("Full Model", 15, 64, bold = T)

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

# THIS SECTION USES THE MEANS AND REGRESSION COEFFICIENTS GENERATED

# ABOVE FOR OUR DECOMPOSITION ANALYSES OF CHANGES IN THE GAP

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

# We create these vectors that group coefficients, used later for analyses

int <- "(Intercept)"

age <- c("age", "agesq")

ed <- c("HighSchool", "SomeCollege", "ba.advdeg")

exp <- c("log.expf", "log.expparttime")

exp.ft <- c("log.expf")

exp.pt <- c("log.expparttime")

region <- c("Northeast", "Northcentral", "South")

race <- c("Black", "Hispanic", "Other")

union <- "union"

wrk.hrs <- "hrswrk.dummy\_50plus"

manuf <- "manuf"

govt.job <- "govt.job"

marital.status <- c("married", "prev.married")

fertility <- c("afb.cat\_21minus", "afb.cat\_22to27", "afb.cat\_27plus", "numkids.3plus")

housework <- "housework"

tenure <- "emp.tenure"

occ.char <- "occ.pct.female"

# First we set the reference years. For our main specification:

cov.years <- c(1981, 2019)

# For alternative specifications, comment out line above and instead run one of the following lines

# cov.years <- c(2019, 1981) # When swapping order of reference years

# cov.years <- c(1980, 2017) # When using prior survey wave covariates

# Our first analyses decompose the changes in the gender wage gap across years

# into changes in characteristics, changes in returns, and changes in the interaction

# between these characteristics and returns

# We scale all three components by changes in the gender wage gap:

denom <- (means.year$lnhrlywage[means.year$female == 0 & means.year$year == cov.years[2]] -

means.year$lnhrlywage[means.year$female == 0 & means.year$year == cov.years[1]]) -

(means.year$lnhrlywage[means.year$female == 1 & means.year$year == cov.years[2]] -

means.year$lnhrlywage[means.year$female == 1 & means.year$year == cov.years[1]])

# For the characteristics component, the numerator is:

# Men's coef, t1 \* (Men's char, t2 - Men's char, t1) -

# Women's coef, t1 \* (Women's char, t2 - Women's char, t1)

# For the returns component, the numerator is:

# Men's char, t1 \* (Men's coef, t2 - Men's coef, t1) -

# Women's char, t1 \* (Women's coef, t2 - Women's coef, t1)

# For the interactions component, the numerator is:

# (Men's coef t2 - Men's coef t1) \* (Men's char t2 - Men's char t1) -

# (Women's coef t2 - Women's coef t1) \* (Women's char t2 - Women's char t1)

# First, we create objects that compute sex-specific characteristic changes

# and sex-specific characteristic levels at t1

mchargap <- means.year %>% filter(female == 0, year == cov.years[2]) -

means.year %>% filter(female == 0, year == cov.years[1])

wchargap <- means.year %>% filter(female == 1, year == cov.years[2]) -

means.year %>% filter(female == 1, year == cov.years[1])

m\_chart1 <- means.year %>% filter(female == 0, year == cov.years[1])

w\_chart1 <- means.year %>% filter(female == 1, year == cov.years[1])

# Using these characteristics gaps and the regression coefficients for each

# model, we compute the components of the decomposition. For expositional clarity,

# we create objects that separate out each sub-component that go into the decomposition

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

# FAMILY MODEL

# CHARACTERISTICS CHANGE COMPONENT, FAMILY MODEL

# Men's family coefficients at t1

m\_famcoeft1 <- fam.coefs %>% filter(female == 0, year == cov.years[1], estimate == "coef") %>%

dplyr::select(-c(estimate, female, year)) %>%

t() %>% as.matrix()

# Women's family coefficients at t1

w\_famcoeft1 <- fam.coefs %>% filter(female == 1, year == cov.years[1], estimate == "coef") %>%

dplyr::select(-c(estimate, female, year)) %>%

t() %>% as.matrix()

# Characteristic change among men, family characteristics

mchargap\_fam <- mchargap[names(mchargap) %in% c(int, marital.status, fertility)] %>%

dplyr::select("(Intercept)", everything()) %>%

as.matrix()

# Characteristic change among women, family characteristics

wchargap\_fam <- wchargap[names(wchargap) %in% c(int, marital.status, fertility)] %>%

dplyr::select("(Intercept)", everything()) %>%

as.matrix()

# Numerator for the detailed characteristics component, Family model

char\_num\_fam <- m\_famcoeft1 \* t(mchargap\_fam) - (w\_famcoeft1 \* t(wchargap\_fam))

# Scaled by the denominator (change in the gender wage gap):

char\_fam <- char\_num\_fam/denom

# RETURNS COMPONENT, FAMILY MODEL

# Changes in men's regression coefficients from the family model

mcoefgap\_fam <- fam.coefs %>% filter(female == 0, estimate == "coef", year == cov.years[2]) %>%

dplyr::select(-c(estimate, female, year)) -

fam.coefs %>% filter(female == 0, estimate == "coef", year == cov.years[1]) %>%

dplyr::select(-c(estimate, female, year))

# Changes in women's regression coefficients from the family model

wcoefgap\_fam <- fam.coefs %>% filter(female == 1, estimate == "coef", year == cov.years[2]) %>%

dplyr::select(-c(estimate, female, year)) -

fam.coefs %>% filter(female == 1, estimate == "coef", year == cov.years[1]) %>%

dplyr::select(-c(estimate, female, year))

# Men's characteristics levels at t1 \* changes in men's regression coefficients from the family model

returns\_men\_fam\_num <- m\_chart1[names(m\_chart1) %in% c(int, marital.status, fertility)] %>%

dplyr::select("(Intercept)", everything()) %>%

as.matrix() \* mcoefgap\_fam

# Women's characteristics levels at t1 \* changes in women's regression coefficients from the family model

returns\_women\_fam\_num <- w\_chart1[names(w\_chart1) %in% c(int, marital.status, fertility)] %>%

dplyr::select("(Intercept)", everything()) %>%

as.matrix() \* wcoefgap\_fam

# Numerator for the detailed returns component, Family model

returns\_num\_fam <- (returns\_men\_fam\_num-returns\_women\_fam\_num)

# Scaled by the denominator (change in the gender wage gap):

returns\_fam <- returns\_num\_fam/denom

# INTERACTIONS COMPONENT, FAMILY MODEL

interaction\_num\_fam <- (mcoefgap\_fam \* mchargap\_fam) - (wcoefgap\_fam \* wchargap\_fam)

# Scaled by the denominator (change in the gender wage gap):

interaction\_fam <- interaction\_num\_fam/denom

# Groups the component from the family model into a single object

t3\_fam\_detailed <- bind\_cols(char\_fam %>% as.data.frame() %>% rename("Characteristics Gap" = V1),

t(returns\_fam) %>% as.data.frame() %>% rename("Returns" = V1),

t(interaction\_fam) %>% as.data.frame() %>% rename("Interaction" = V1))

# Summarizes the detailed results for each variable group by

# summing across the relevant category

t3\_fam\_sum <- t3\_fam\_detailed %>%

mutate(group = case\_when(rownames(.) %in% marital.status ~ "Marriage",

rownames(.) %in% fertility ~ "Fertility",

TRUE ~ "Intercept")) %>%

group\_by(group) %>%

summarise\_each(funs(sum(., na.rm = TRUE))) %>%

adorn\_totals("row") %>%

filter(group != "Intercept") %>%

mutate\_if(is.numeric, ~.\*100) %>%

mutate\_if(is.numeric, round, digits = 2) %>%

rename("Variables" = group) %>%

mutate(model = "Family")

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

# FULL MODEL

# CHARACTERISTICS CHANGE COMPONENT, FULL MODEL

# Men's full coefficients at t1

m\_fullcoeft1 <- full.coefs %>% filter(female == 0, year == cov.years[1], estimate == "coef") %>%

dplyr::select(-c(estimate, female, year)) %>%

t() %>% as.matrix()

# Women's full coefficients at t1

w\_fullcoeft1 <- full.coefs %>% filter(female == 1, year == cov.years[1], estimate == "coef") %>%

dplyr::select(-c(estimate, female, year)) %>%

t() %>% as.matrix()

# Characteristic change among men, full characteristics

mchargap\_full <- mchargap[names(mchargap) %in%

c(int, region, race, marital.status, housework, fertility,

ed, union, govt.job, exp.ft, wrk.hrs,

tenure, occ.char, "occ.managers", manuf)] %>%

dplyr::select("(Intercept)", everything()) %>%

as.matrix()

# Characteristic change among women, full characteristics

wchargap\_full <- wchargap[names(wchargap) %in%

c(int, region, race, marital.status, housework, fertility,

ed, union, govt.job, exp.ft, wrk.hrs,

tenure,occ.char, "occ.managers", manuf)] %>%

dplyr::select("(Intercept)", everything()) %>%

as.matrix()

# Numerator for the detailed characteristics component, Full model

char\_num\_full <- m\_fullcoeft1 \* t(mchargap\_full) - (w\_fullcoeft1 \* t(wchargap\_full))

# Scaled by the denominator (change in the gender wage gap):

char\_full <- char\_num\_full/denom

# RETURNS COMPONENT, FULL MODEL

# Changes in men's regression coefficients from the full model

mcoefgap\_full <- full.coefs %>% filter(female == 0, estimate == "coef", year == cov.years[2]) %>%

dplyr::select(-c(estimate, female, year)) -

full.coefs %>% filter(female == 0, estimate == "coef", year == cov.years[1]) %>%

dplyr::select(-c(estimate, female, year))

# Changes in women's regression coefficients from the full model

wcoefgap\_full <- full.coefs %>% filter(female == 1, estimate == "coef", year == cov.years[2]) %>%

dplyr::select(-c(estimate, female, year)) -

full.coefs %>% filter(female == 1, estimate == "coef", year == cov.years[1]) %>%

dplyr::select(-c(estimate, female, year))

# Men's characteristics levels at t1 \* changes in men's regression coefficients from the full model

returns\_men\_full\_num <- m\_chart1[names(m\_chart1) %in%

c(int, region, race, marital.status, housework, fertility,

ed, union, govt.job, exp.ft, wrk.hrs,

tenure, occ.char, "occ.managers", manuf)] %>%

dplyr::select("(Intercept)", everything()) %>%

as.matrix() \* mcoefgap\_full

# Women's characteristics levels at t1 \* changes in women's regression coefficients from the full model

returns\_women\_full\_num <- w\_chart1[names(w\_chart1) %in%

c(int, region, race, marital.status, housework, fertility,

ed, union, govt.job, exp.ft, wrk.hrs,

tenure, occ.char, "occ.managers", manuf)] %>%

dplyr::select("(Intercept)", everything()) %>%

as.matrix() \* wcoefgap\_full

# Numerator for the detailed returns component, Full model

returns\_num\_full <- (returns\_men\_full\_num-returns\_women\_full\_num)

# Scaled by the denominator (change in the gender wage gap):

returns\_full <- returns\_num\_full/denom

# INTERACTIONS COMPONENT, FULL MODEL

interaction\_num\_full <- (mcoefgap\_full \* mchargap\_full) - (wcoefgap\_full \* wchargap\_full)

# Scaled by the denominator (change in the gender wage gap):

interaction\_full <- interaction\_num\_full/denom

# Groups the component from the full model into a single object

t3\_full\_detailed <- bind\_cols(char\_full %>% as.data.frame() %>% rename("Characteristics Gap" = V1),

t(returns\_full) %>% as.data.frame() %>% rename("Returns" = V1),

t(interaction\_full) %>% as.data.frame() %>% rename("Interaction" = V1))

# Summarizes the detailed results for each variable group by

# summing across the relevant category

t3\_full <- t3\_full\_detailed %>%

mutate(group = case\_when(rownames(.) %in% marital.status ~ "Marital Status",

rownames(.) %in% fertility ~ "Fertility",

rownames(.) %in% c(race, region) ~ "Demographic",

rownames(.) %in% housework ~ "Housework",

rownames(.) %in% ed ~ "Education",

rownames(.) %in% int ~ "Intercept",

TRUE ~ "Job Traits")) %>%

group\_by(group) %>%

summarise\_each(funs(sum(., na.rm = TRUE))) %>%

adorn\_totals("row") %>%

filter(group != "Intercept") %>%

mutate\_if(is.numeric, ~.\*100) %>%

mutate\_if(is.numeric, round, digits = 2) %>%

rename("Variables" = group)

# Summarizes the detailed results distinguishing role of family variables from other variables

t3\_full\_sum <- t3\_full %>%

mutate(vargroup = case\_when(Variables %in% c("Marital Status", "Fertility", "Housework") ~ "Sum, Family",

TRUE ~ "Other")) %>%

group\_by(vargroup) %>%

summarise\_if(is.numeric, sum) %>%

filter(vargroup != "Other") %>%

rename("Variables" = vargroup) %>%

bind\_rows(t3\_full, .) %>%

mutate(model = "Full")

# Combining the summarised results from all components across both models

t3\_allcomponents <- bind\_rows(t3\_fam\_sum, t3\_full\_sum)

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

# TABLE A2: DETAILED DECOMPOSITION SHOWING CONTRIBUTIONS OF CHANGING

# CHARACTERISTICS, RETURNS, AND INTERACTIONS TO THE CHANGING WAGE GAP

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

knitr::kable(t3\_allcomponents %>% dplyr::select(-model), booktabs = T, format = "latex",

caption = "Decomposition, Change in Gender Pay Gap") %>%

pack\_rows("Family Model", 1, 3, bold = T) %>% pack\_rows("Full Model", 4, 11, bold = T)

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

# THIS SECTION DECOMPOSES CHANGES OVER TIME AMONG MEN AND WOMEN

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

# We compute each decomposition component (changing characteristics,

# returns, and interactions) for each group, for each model

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

# FAMILY MODEL

# CHANGING CHARACTERISTICS BY SEX

# Changing coefficients for men, family model

m\_famcoeft1 <- fam.coefs %>% filter(female == 0, year == cov.years[1], estimate == "coef") %>%

dplyr::select(-c(estimate, female, year)) %>%

t() %>% as.matrix()

# Changing family characteristics for men

mchargap\_fam <- mchargap[names(mchargap) %in% c(int, marital.status, fertility)] %>%

dplyr::select("(Intercept)", everything()) %>%

as.matrix()

# Numerator for the detailed characteristics change component among men, Family model

mchar\_num\_fam <- m\_famcoeft1 \* t(mchargap\_fam)

# Scaled by changes in the gender wage gap

char\_men\_fam <- mchar\_num\_fam/denom

# Changing coefficients for women, family model

w\_famcoeft1 <- fam.coefs %>% filter(female == 1, year == cov.years[1], estimate == "coef") %>%

dplyr::select(-c(estimate, female, year)) %>%

t() %>% as.matrix()

# Changing family characteristics for women

wchargap\_fam <- wchargap[names(wchargap) %in% c(int, marital.status, fertility)] %>%

dplyr::select("(Intercept)", everything()) %>%

as.matrix()

# Numerator for the detailed characteristics change component among women, Family model

wchar\_num\_fam <- w\_famcoeft1 \* t(wchargap\_fam)

# Scaled by changes in the gender wage gap

char\_women\_fam <- wchar\_num\_fam/denom

# CHANGING RETURNS BY SEX

# (we already computed these when decomposing the contribution of changes in the different returns in lines above)

returns\_men\_fam <- returns\_men\_fam\_num/denom

returns\_women\_fam <- returns\_women\_fam\_num/denom

# CHANGING INTERACTIONS BY SEX

interaction\_men\_num\_fam <- (mcoefgap\_fam \* mchargap\_fam)

interaction\_men\_fam <- (mcoefgap\_fam \* mchargap\_fam)/denom

interaction\_women\_num\_fam <- (wcoefgap\_fam \* wchargap\_fam)

interaction\_women\_fam <- (wcoefgap\_fam \* wchargap\_fam)/denom

# Groups the component from the family model for each sex into a single object

t3\_fam\_sex\_detailed <- bind\_cols(char\_men\_fam %>% as.data.frame() %>% rename("Characteristics Gap Men" = V1),

t(returns\_men\_fam) %>% as.data.frame() %>% rename("Returns Men" = V1),

t(interaction\_men\_fam) %>% as.data.frame() %>% rename("Interaction Men" = V1),

char\_women\_fam %>% as.data.frame() %>% rename("Characteristics Gap Women" = V1),

t(returns\_women\_fam) %>% as.data.frame() %>% rename("Returns Women" = V1),

t(interaction\_women\_fam) %>% as.data.frame() %>% rename("Interaction Women" = V1))

# Summarizes the detailed results for each variable group by

# summing across the relevant category

t3\_fam\_sex\_sum <- t3\_fam\_sex\_detailed %>%

mutate(group = case\_when(rownames(.) %in% marital.status ~ "Marriage",

rownames(.) %in% fertility ~ "Fertility",

TRUE ~ "Intercept")) %>%

group\_by(group) %>%

summarise\_each(funs(sum(., na.rm = TRUE))) %>%

adorn\_totals("row") %>%

filter(group != "Intercept") %>%

mutate\_if(is.numeric, ~.\*100) %>%

mutate\_if(is.numeric, round, digits = 2) %>%

rename("Variables" = group) %>%

mutate(model = "Family")

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

# FULL MODEL

# CHANGING CHARACTERISTICS BY SEX

# Coefficients at t1 for men, full model

m\_fullcoeft1 <- full.coefs %>% filter(female == 0, year == cov.years[1], estimate == "coef") %>%

dplyr::select(-c(estimate, female, year)) %>%

t() %>% as.matrix()

# Changing characteristics for men, full characteristics

mchargap\_full <- mchargap[names(mchargap) %in%

c(int,region, race, marital.status, housework, fertility,

ed, union, govt.job, exp.ft, wrk.hrs,

tenure, occ.char, "occ.managers", manuf)] %>%

dplyr::select("(Intercept)", everything()) %>%

as.matrix()

# Numerator for the detailed characteristics change component among men, Full model

mchar\_num\_full <- m\_fullcoeft1 \* t(mchargap\_full)

# Scaled by changes in the gender wage gap

char\_men\_full <- mchar\_num\_full/denom

# Coefficients at t1 for women, full model

w\_fullcoeft1 <- full.coefs %>% filter(female == 1, year == cov.years[1], estimate == "coef") %>%

dplyr::select(-c(estimate, female, year)) %>%

t() %>% as.matrix()

# Changing characteristics for women, full characteristics

wchargap\_full <- wchargap[names(wchargap) %in%

c(int, region, race, marital.status, housework, fertility,

ed, union, govt.job, exp.ft, wrk.hrs,

tenure, occ.char, "occ.managers", manuf)] %>%

dplyr::select("(Intercept)", everything()) %>%

as.matrix()

# Numerator for the detailed characteristics change component among women, Full model

wchar\_num\_full <- w\_fullcoeft1 \* t(wchargap\_full)

# Scaled by changes in the gender wage gap

char\_women\_full <- wchar\_num\_full/denom

# CHANGING RETURNS BY SEX

# (we already computed these when decomposing the contribution of changes in the different returns in lines above)

returns\_men\_full <- returns\_men\_full\_num/denom

returns\_women\_full <- returns\_women\_full\_num/denom

# CHANGING INTERACTIONS BY SEX

interaction\_men\_num\_full <- (mcoefgap\_full \* mchargap\_full)

interaction\_men\_full <- (mcoefgap\_full \* mchargap\_full)/denom

interaction\_women\_num\_full <- (wcoefgap\_full \* wchargap\_full)

interaction\_women\_full <- (wcoefgap\_full \* wchargap\_full)/denom

# Groups the component from the full model for each sex into a single object

t3\_full\_sex\_detailed <- bind\_cols(char\_men\_full %>% as.data.frame() %>% rename("Characteristics Gap Men" = V1),

t(returns\_men\_full) %>% as.data.frame() %>% rename("Returns Men" = V1),

t(interaction\_men\_full) %>% as.data.frame() %>% rename("Interaction Men" = V1),

char\_women\_full %>% as.data.frame() %>% rename("Characteristics Gap Women" = V1),

t(returns\_women\_full) %>% as.data.frame() %>% rename("Returns Women" = V1),

t(interaction\_women\_full) %>% as.data.frame() %>% rename("Interaction Women" = V1))

# Summarizes the detailed results for each variable group by

# summing across the relevant category

t3\_full\_sex <- t3\_full\_sex\_detailed %>%

mutate(group = case\_when(rownames(.) %in% marital.status ~ "Marital Status",

rownames(.) %in% fertility ~ "Fertility",

rownames(.) %in% c(race, region) ~ "Demographic",

rownames(.) %in% housework ~ "Housework",

rownames(.) %in% ed ~ "Education",

rownames(.) %in% int ~ "Intercept",

TRUE ~ "Job Traits")) %>%

group\_by(group) %>%

summarise\_each(funs(sum(., na.rm = TRUE))) %>%

adorn\_totals("row") %>%

filter(group != "Intercept") %>%

mutate\_if(is.numeric, ~.\*100) %>%

mutate\_if(is.numeric, round, digits = 2) %>%

rename("Variables" = group)

# Summarizes the detailed results distinguishing role of family variables from other variables

t3\_full\_sex\_sum <- t3\_full\_sex %>%

mutate(vargroup = case\_when(Variables %in% c("Marital Status", "Fertility", "Housework") ~ "Sum, Family",

TRUE ~ "Other")) %>%

group\_by(vargroup) %>%

summarise\_if(is.numeric, sum) %>%

filter(vargroup != "Other") %>%

rename("Variables" = vargroup) %>%

bind\_rows(t3\_full\_sex, .) %>%

mutate(model = "Full")

# Combining the summarised results from all components across both models

t3\_combined\_sex <- bind\_rows(t3\_fam\_sex\_sum, t3\_full\_sex\_sum)

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

# TABLE 3: SHOWING THE CONTRIBUTION OF CHANGES IN CHARACTERISTICS

# FOR MEN'S PAY, WOMEN'S PAY, AND THE PAY GAP

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

# Combining the summarised result from the sex-specific decompositions

# with the two-sexes decomposition

t3\_final <- bind\_cols(t3\_allcomponents %>% dplyr::select(Variables, "Characteristics Gap") %>%

rename("Pay Gap" = "Characteristics Gap"),

t3\_combined\_sex %>% dplyr::select("Characteristics Gap Men", "Characteristics Gap Women") %>%

rename("Men" = "Characteristics Gap Men",

"Women" = "Characteristics Gap Women")) %>%

rbind(c("Change in Pay", denom, denom, denom))

# Creating the table output

knitr::kable(t3\_final, booktabs = T, format = "latex", digits = 3,

caption = "Contribution of Changing Characteristics to Changes in Men’s Pay, Women’s Pay, and the Pay Gap") %>%

pack\_rows("Family Model", 1, 3, bold = F) %>% pack\_rows("Full Model", 4, 11, bold = F)

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

# IN THE FOLLOWING SECTION, WE REPEAT THE SAME ANALYSIS AS WE DID FOR

# THE TWO-SEXES DECOMPOSITION FOR EACH INTERDECADE PERIOD, FOCUSING

# ONLY ON THE CHANGING CHARACTERISTICS COMPONENT. WE SCALE THE

# CONTRIBUTION OF EACH COMPONENT BY THE CHANGE IN THE GENDER PAY GAP

# OVER THE WHOLE PERIOD

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

cov.years <- c(1981, 1991)

mchargap\_1991 <- means.year %>% filter(female == 0, year == cov.years[2]) -

means.year %>% filter(female == 0, year == cov.years[1])

wchargap\_1991 <- means.year %>% filter(female == 1, year == cov.years[2]) -

means.year %>% filter(female == 1, year == cov.years[1])

mchargap\_fam\_1991 <- mchargap\_1991[names(mchargap\_1991) %in% c(int, marital.status, fertility)] %>%

dplyr::select("(Intercept)", everything()) %>%

as.matrix()

wchargap\_fam\_1991 <- wchargap\_1991[names(wchargap\_1991) %in% c(int, marital.status, fertility)] %>%

dplyr::select("(Intercept)", everything()) %>%

as.matrix()

char\_num\_fam\_1991 <- m\_famcoeft1 \* t(mchargap\_fam\_1991) - (w\_famcoeft1 \* t(wchargap\_fam\_1991))

char\_fam\_1991 <- char\_num\_fam\_1991/denom

mchargap\_full\_1991 <- mchargap\_1991[names(mchargap\_1991) %in%

c(int, region, race, marital.status, housework, fertility,

ed, union, govt.job, exp.ft, wrk.hrs,

tenure, occ.char, "occ.managers", manuf)] %>%

dplyr::select("(Intercept)", everything()) %>%

as.matrix()

wchargap\_full\_1991 <- wchargap\_1991[names(wchargap\_1991) %in%

c(int, region, race, marital.status, housework, fertility,

ed, union, govt.job, exp.ft, wrk.hrs,

tenure,occ.char, "occ.managers", manuf)] %>%

dplyr::select("(Intercept)", everything()) %>%

as.matrix()

char\_num\_full\_1991 <- m\_fullcoeft1 \* t(mchargap\_full\_1991) - (w\_fullcoeft1 \* t(wchargap\_full\_1991))

char\_full\_1991 <- char\_num\_full\_1991/denom

cov.years <- c(1991, 2001)

mchargap\_2001 <- means.year %>% filter(female == 0, year == cov.years[2]) -

means.year %>% filter(female == 0, year == cov.years[1])

wchargap\_2001 <- means.year %>% filter(female == 1, year == cov.years[2]) -

means.year %>% filter(female == 1, year == cov.years[1])

mchargap\_fam\_2001 <- mchargap\_2001[names(mchargap\_2001) %in% c(int, marital.status, fertility)] %>%

dplyr::select("(Intercept)", everything()) %>%

as.matrix()

wchargap\_fam\_2001 <- wchargap\_2001[names(wchargap\_2001) %in% c(int, marital.status, fertility)] %>%

dplyr::select("(Intercept)", everything()) %>%

as.matrix()

char\_num\_fam\_2001 <- m\_famcoeft1 \* t(mchargap\_fam\_2001) - (w\_famcoeft1 \* t(wchargap\_fam\_2001))

char\_fam\_2001 <- char\_num\_fam\_2001/denom

mchargap\_full\_2001 <- mchargap\_2001[names(mchargap\_2001) %in%

c(int, region, race, marital.status, housework, fertility,

ed, union, govt.job, exp.ft, wrk.hrs,

tenure, occ.char, "occ.managers", manuf)] %>%

dplyr::select("(Intercept)", everything()) %>%

as.matrix()

wchargap\_full\_2001 <- wchargap\_2001[names(wchargap\_2001) %in%

c(int, region, race, marital.status, housework, fertility,

ed, union, govt.job, exp.ft, wrk.hrs,

tenure, occ.char, "occ.managers", manuf)] %>%

dplyr::select("(Intercept)", everything()) %>%

as.matrix()

char\_num\_full\_2001 <- m\_fullcoeft1 \* t(mchargap\_full\_2001) - (w\_fullcoeft1 \* t(wchargap\_full\_2001))

char\_full\_2001 <- char\_num\_full\_2001/denom

cov.years <- c(2001, 2011)

mchargap\_2011 <- means.year %>% filter(female == 0, year == cov.years[2]) -

means.year %>% filter(female == 0, year == cov.years[1])

wchargap\_2011 <- means.year %>% filter(female == 1, year == cov.years[2]) -

means.year %>% filter(female == 1, year == cov.years[1])

mchargap\_fam\_2011 <- mchargap\_2011[names(mchargap\_2011) %in% c(int, marital.status, fertility)] %>%

dplyr::select("(Intercept)", everything()) %>%

as.matrix()

wchargap\_fam\_2011 <- wchargap\_2011[names(wchargap\_2011) %in% c(int, marital.status, fertility)] %>%

dplyr::select("(Intercept)", everything()) %>%

as.matrix()

char\_num\_fam\_2011 <- m\_famcoeft1 \* t(mchargap\_fam\_2011) - (w\_famcoeft1 \* t(wchargap\_fam\_2011))

char\_fam\_2011 <- char\_num\_fam\_2011/denom

mchargap\_full\_2011 <- mchargap\_2011[names(mchargap\_2011) %in%

c(int, region, race, marital.status, housework, fertility,

ed, union, govt.job, exp.ft, wrk.hrs,

tenure, occ.char, "occ.managers", manuf)] %>%

dplyr::select("(Intercept)", everything()) %>%

as.matrix()

wchargap\_full\_2011 <- wchargap\_2011[names(wchargap\_2011) %in%

c(int, region, race, marital.status, housework, fertility,

ed, union, govt.job, exp.ft, wrk.hrs,

tenure, occ.char, "occ.managers", manuf)] %>%

dplyr::select("(Intercept)", everything()) %>%

as.matrix()

char\_num\_full\_2011 <- m\_fullcoeft1 \* t(mchargap\_full\_2011) - (w\_fullcoeft1 \* t(wchargap\_full\_2011))

char\_full\_2011 <- char\_num\_full\_2011/denom

cov.years <- c(2011, 2019)

mchargap\_2019 <- means.year %>% filter(female == 0, year == cov.years[2]) -

means.year %>% filter(female == 0, year == cov.years[1])

wchargap\_2019 <- means.year %>% filter(female == 1, year == cov.years[2]) -

means.year %>% filter(female == 1, year == cov.years[1])

mchargap\_fam\_2019 <- mchargap\_2019[names(mchargap\_2019) %in% c(int, marital.status, fertility)] %>%

dplyr::select("(Intercept)", everything()) %>%

as.matrix()

wchargap\_fam\_2019 <- wchargap\_2019[names(wchargap\_2019) %in% c(int, marital.status, fertility)] %>%

dplyr::select("(Intercept)", everything()) %>%

as.matrix()

char\_num\_fam\_2019 <- m\_famcoeft1 \* t(mchargap\_fam\_2019) - (w\_famcoeft1 \* t(wchargap\_fam\_2019))

char\_fam\_2019 <- char\_num\_fam\_2019/denom

mchargap\_full\_2019 <- mchargap\_2019[names(mchargap\_2019) %in%

c(int, region, race, marital.status, housework, fertility,

ed, union, govt.job, exp.ft, wrk.hrs,

tenure, occ.char, "occ.managers", manuf)] %>%

dplyr::select("(Intercept)", everything()) %>%

as.matrix()

wchargap\_full\_2019 <- wchargap\_2019[names(wchargap\_2019) %in%

c(int, region, race, marital.status, housework, fertility,

ed, union, govt.job, exp.ft, wrk.hrs,

tenure, occ.char, "occ.managers", manuf)] %>%

dplyr::select("(Intercept)", everything()) %>%

as.matrix()

char\_num\_full\_2019 <- m\_fullcoeft1 \* t(mchargap\_full\_2019) - (w\_fullcoeft1 \* t(wchargap\_full\_2019))

char\_full\_2019 <- char\_num\_full\_2019/denom

# Binds the characteristics change contributions by decade for the family model to one object

char\_fam\_decades <- bind\_cols(as.data.frame(char\_fam\_1991) %>% rename("1980-1990" = V1),

as.data.frame(char\_fam\_2001) %>% rename("1990-2000" = V1),

as.data.frame(char\_fam\_2011) %>% rename("2000-2010" = V1),

as.data.frame(char\_fam\_2019) %>% rename("2010-2018" = V1))

# Summarises the contributions of each variable group by decade for the family model

char\_fam\_decades\_summary <- char\_fam\_decades %>%

mutate(group = case\_when(rownames(.) %in% marital.status ~ "Marital Status",

rownames(.) %in% fertility ~ "Fertility")) %>%

group\_by(group) %>%

summarise\_each(funs(sum(., na.rm = TRUE))) %>%

filter(group != "Intercept") %>%

mutate\_if(is.numeric, ~.\*100) %>%

mutate\_if(is.numeric, round, digits = 2) %>%

rename("Variables" = group) %>%

mutate(model = "Family Model")

# Binds the characteristics change contributions by decade for the full model to one object

char\_full\_decades <- bind\_cols(as.data.frame(char\_full\_1991) %>% rename("1980-1990" = V1),

as.data.frame(char\_full\_2001) %>% rename("1990-2000" = V1),

as.data.frame(char\_full\_2011) %>% rename("2000-2010" = V1),

as.data.frame(char\_full\_2019) %>% rename("2010-2018" = V1))

# Summarises the contributions of each variable group by decade for the full model

char\_full\_decades\_summary <- char\_full\_decades %>%

mutate(group = case\_when(rownames(.) %in% marital.status ~ "Marital Status",

rownames(.) %in% fertility ~ "Fertility",

rownames(.) %in% c(race, region) ~ "Demographic",

rownames(.) %in% housework ~ "Housework",

rownames(.) %in% ed ~ "Education",

rownames(.) %in% int ~ "Intercept",

TRUE ~ "Job Traits")) %>%

group\_by(group) %>%

summarise\_each(funs(sum(., na.rm = TRUE))) %>%

filter(group != "Intercept") %>%

mutate\_if(is.numeric, ~.\*100) %>%

mutate\_if(is.numeric, round, digits = 2) %>%

rename("Variables" = group) %>%

mutate(model = "Full Model")

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

# FIGURE 3 SHOWS THE CONTRIBUTION OF CHANGING FAMILY CHARACTERISTICS

# BY DECADE FOR EACH MODEL

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

fig3 <- bind\_rows(char\_fam\_decades\_summary,

char\_full\_decades\_summary) %>%

filter(Variables %in% c('Marital Status', "Fertility", "Housework")) %>%

mutate(Variables = factor(Variables, levels = c('Marital Status', "Fertility", "Housework"))) %>%

gather(decade, value, - c("Variables", "model")) %>%

ggplot(aes(x = decade, y = value, fill = Variables)) +

geom\_bar(position = position\_stack(reverse = TRUE), stat="identity") +

facet\_wrap(~ model) +

labs(x = "", y = "% Change in Pay Gap Explained by Component",

title = "Contribution of Changing Family Life to the Gender Pay Gap, 1980-2018",

fill = "") +

theme\_bw() +

theme(legend.position = "right",

plot.title = element\_text(hjust = 0.5, face = "bold", size = 12),

axis.text.x = element\_text(angle = 12, size = 8, vjust = 0.5)) +

scale\_fill\_manual(values=c("#999999", "#56B4E9", "#E69F00"))

ggsave(plot = fig3, "/Users/ninocricco/Desktop/PSID\_CodeReview/output/fig3.jpg",

width = 10, height = 6.5, units = "in", device='jpeg', dpi=700)

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

# TABLE A4 SHOWS THE SUMMARIZED CONTRIBUTIONS OF CHANGES IN

# CHARACTERISTICS BY DECADE FOR EACH VARIABLE GROUP AND MODEL

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

knitr::kable(bind\_rows(char\_fam\_decades\_summary %>%

adorn\_totals("row"),

char\_full\_decades\_summary %>%

adorn\_totals("row"),

char\_full\_decades\_summary %>%

filter(Variables %in% c("Fertility", "Marital Status", "Housework")) %>%

adorn\_totals("row") %>%

filter(Variables == "Total")

) %>% dplyr::select(-model),

booktabs = T, format = "latex", digits = 3,

caption = "Contribution of Changing Characteristics to the Gender Pay Gap, 1980-2018") %>%

pack\_rows("Family Model", 1, 3, bold = T) %>%

pack\_rows("Full Model", 4, 10, bold = T)