



```

cap log close
log using "$log_save\create_ind_panel_adj", replace

use "$saveaddress_data\merged_ind_panel", clear
    keep amt3601 amt3805 busdistmuni_geo buslocmuni_geo busmainplc_geo busprov_ge
> o ///
    province_geo districtmunicip_geo localmunicip_geo mainplace_geo dateofbirth ///
    gender idno mainincomesourcecode natureofperson payereferenceno ///
    periodemployedfrom periodemployedto taxrefno taxyear totalperiodsinyearofassessment /
> //
    totalperiodsworked passportno certificateno revisionnumber kerr_income kerr_emp kerr_
> emp_inc FYE

* create individual identifier: code from Marlies Piek: "Set_up_4_march"
gen id_new=idno
replace id_new=passportno if id_new==" "
label var id_new "Unique identifier, from SA ID or passport no."
egen n_id=group(id_new)

sort idno payeref taxyear
br idno payeref taxyear

    deststring totalperiodsinyearofassessment, gen(totpy) force
    replace totpy = abs(totpy)
    deststring totalperiodsworked, gen(totwork) force
    replace totwork=abs(totwork)
    gen weight = totwork/totpy

    gen paye_next = 0
>
    gen paye_prev = 0

    replace paye_next = 1 if idno == idno[_n+1] & taxyear+1 == taxyear[_n+1] & tax
> year !=2018 /*see in the next tax year*/

    replace paye_prev = 1 if idno == idno[_n-1] & taxyear-1 == taxyear[_n-1] & tax
> year !=2010 /*seen in the previous taxyear*/

    cap drop taxyear_start_date
    gen taxyear_start_date = (monthly(strofreal(taxyear) + "/" + "01", "YM") -12)
> + 2

    cap drop taxyear_end_date
    gen taxyear_end_date = taxyear_start_date +12

    cap drop taxyear_start_date_tmp
    cap drop taxyear_end_date_tmp
    gen taxyear_start_date_tmp =dofm(taxyear_start_date)
    gen taxyear_end_date_tmp =dofm((taxyear_end_date))-1

    cap drop taxyear_start_date
    cap drop taxyear_end_date

    ren taxyear_start_date_tmp taxyear_start_date
    format taxyear_start_date %tdDD/NN/CCYY
    cap ren taxyear_end_date_tmp taxyear_end_date
    format taxyear_end_date %tdDD/NN/CCYY

    gen c_taxyear_start_date = taxyear_start_date
    gen c_taxyear_end_date = taxyear_end_date
    format c_taxyear_start_date %tdDD/NN/CCYY
    format c_taxyear_end_date %tdDD/NN/CCYY

    generate days_in_year = c_taxyear_end_date - c_taxyear_start_date
>

    gen start_imp = taxyear_start_date if weight>=1
> *Start date for workers working full taxyear or more* /
    gen end_imp = taxyear_end_date if weight>=1
> *end date for worker working full taxyear or more */ /
    format start_imp %tdDD/NN/CCYY

```

```

format end_imp %tdDD/NN/CCYY

/*****
X.1.1. IMPUTE
*****/

/*****
Impute only for id's with no-unity weight not seen both previously and after
Using next period data
*****/
    gen c_start_imp = start_imp
    gen c_end_imp = end_imp
    format c_start_imp %tdDD/NN/CCYY
    format c_end_imp %tdDD/NN/CCYY

    sort idno payeref taxyear

    replace c_end_imp = c_taxyear_start_date[_n+1] -1 if paye_next ==1 & paye_prev
> == 0 & weight<1 & c_end_imp ==.
    replace c_start_imp = c_end_imp - weight*days_in_year if c_start_imp == . & c_
> end_imp!=.

/*
> Using previous period data*/
    replace c_start_imp = c_taxyear_end_date[_n-1] +1 if paye_prev ==1 & paye_next
> == 0 & weight<1 & c_start_imp ==.
    replace c_end_imp = c_start_imp + weight*days_in_year if c_end_imp == . & c_st
> art_imp!=.

    replace c_start_imp = round(c_start_imp)
    replace c_end_imp = round(c_end_imp)

>

/*****
Same firm dates work done with old approach to calculate actual days worked.
*****/

    gen tag_unimp = 1 if weight !=. & c_start_imp == . & c_end_imp == .
    replace tag_unimp = 0 if tag_unimp == .

    replace c_start_imp = c_taxyear_start_date if tag_unimp == 1
    replace c_end_imp = c_taxyear_end_date if tag_unimp == 1

    save "$saveaddress_data\IRP5_imputingweights_fulldata_DELETE1.dta", replace
    forvalues year= 2010/2018 {
    use if taxyear==`year' using "$saveaddress_data\IRP5_imputingweights_fulldata_
> DELETE1.dta", clear
    rename amt3601 a3601_income
    rename c_end_imp ind_end
    rename c_start_imp ind_start
    rename c_taxyear_start_date taxyear_start
    rename c_taxyear_end_date taxyear_end

    capture drop tag_endmissing
    gen tag_endmissing=1*(ind_end==.)
    capture drop tag_startmissing
    gen tag_startmissing=1*(ind_start==.)
    capture drop tag_startsafter
    gen tag_startsafter = 1*(taxyear_end<ind_start)
    capture drop tag_endsbefore
    gen tag_endsbefore = 1*(taxyear_start>ind_end)

```

```

di in red "Clean up ending and starting dates"
capture drop ind_end_orig
gen ind_end_orig = ind_end
capture drop ind_start_orig
gen ind_start_orig = ind_start
di in red "Replace start and end if start>end"
replace ind_end = ind_start_orig if ind_start_orig>ind_end_ori
> g
replace ind_start = ind_end_orig if ind_end_orig<ind_start_ori
> g

di in red "Make dates missing if end and start are either afte
> r or before start of tax year"
replace ind_end = . if tag_endsbefore==1 | tag_startsafter==1
>
replace ind_start = . if tag_endsbefore==1 | tag_startsafter==
> 1
di in red "Replace dates with date of firms tax year end if en
> ds after"
replace ind_end = taxyear_end if ind_end>taxyear_end & ind_end
> !=.
di in red "Replace start date with date of firm start if pers
> on started employment before firm fin statements"
replace ind_start = taxyear_start if ind_start<taxyear_start &
> ind_start!=.

capture format %tdDD/NN/CCYY ind_start
capture format %tdDD/NN/CCYY ind_end
capture format %tdDD/NN/CCYY ind_start_orig
capture format %tdDD/NN/CCYY ind_end_orig

>
> hat is going on here? */
> rename FYE firm_end
> //for now
format %tdDD/NN/CCYY firm_end

/*****
THIS IS A PROBLEM AREA
*****/
di in red "Generate firm start and end dates"

> /*these dates need to be updated with each inclusion of a leap year*/

/* ****
I1. ISSUE - TO DO
*****
Write program that auto-generates firm start date from firm end
date.

Author: FK
***** */
cap drop firm_end_tmp
gen firm_end_tmp = mofd(firm_end)
format firm_end_tmp %tmDD/NN/CCYY
cap drop firm_start_tmp
gen firm_start_tmp = firm_end_tmp - 11
format firm_start_tmp %tmDD/NN/CCYY
gen firm_start = dofm(firm_start_tmp)
format firm_start %tdDD/NN/CCYY
drop firm_start_tmp firm_end_tmp

```

```

di in red "Generate individual in firm data"
capture drop indfirm_start
gen indfirm_start = firm_start*(firm_start>=ind_start) + ind_s
> tart*(ind_start>firm_start)
replace indfirm_start = . if ind_end<firm_start
replace indfirm_start = . if ind_start>firm_end

capture drop indfirm_end
gen indfirm_end = firm_end*(firm_end<=ind_end) + ind_end*(ind_
> end<firm_end)

replace indfirm_end = . if ind_end<firm_start
replace indfirm_end = . if ind_start>firm_end
format %tdDD/NN/CCYY indfirm_end
format %tdDD/NN/CCYY indfirm_start

di in red "Generate Days Worked and days in firm, days after f
> irm fin year and days before firm fin year"
capture drop ind_days
gen ind_days = ind_end-ind_start+1
capture drop ind_firm
gen ind_firm = indfirm_end-indfirm_start+1
capture drop ind_firm_after
gen ind_firm_after = ind_end-indfirm_end
*replace ind_firm_after = ind_end-ind_start+1 if ind_end>firm_star
> t & ind_start>firm_start & ind_firm_after==.
capture drop ind_firm_before
gen ind_firm_before = indfirm_start - ind_start
*replace ind_firm_before = ind_end-ind_start+1 if ind_end<firm
> _start & ind_start<firm_start & ind_firm_before==.

di in red "Proportion of time in firm"
capture drop prop_in_firm
gen prop_in_firm = ind_firm/ind_days if tag_unimp == 0
capture drop prop_before_firm
gen prop_before_firm = ind_firm_before/ind_days if tag_unimp ==
> 0

capture drop prop_after_firm
gen prop_after_firm = ind_firm_after/ind_days if tag_unimp ==
> 0

replace prop_after_firm = ind_days/ind_days if ind_start>firm_
> end & tag_unimp == 0
replace prop_before_firm = ind_days/ind_days if ind_end<firm_s
> tart & tag_unimp == 0
replace prop_after_firm = 0 if prop_after_firm==. & tag_unimp
> == 0
replace prop_before_firm = 0 if prop_before_firm==. & tag_unim
> p == 0
/*capture rename grossnfundincomeamt grossnretfundincomeamt
> */
di in red "Generate prop_in_firm, prop_after_firm and prop_bef
> ore_firm for those with unimputed days but with weights"
replace prop_in_firm = (ind_firm*weight)/ind_days if t
> ag_unimp == 1
replace prop_before_firm = (ind_firm_before*weight)/ind
> _days if tag_unimp == 1
replace prop_after_firm = (ind_firm_after*weight)/ind_
> days if tag_unimp == 1

replace prop_after_firm = ind_days/ind_days if ind_sta
> rt>firm_end & tag_unimp == 1
replace prop_before_firm = ind_days/ind_days if ind_en
> d<firm_start & tag_unimp == 1
replace prop_after_firm = 0 if prop_after_firm==. & ta
> g_unimp == 1
replace prop_before_firm = 0 if prop_before_firm==. &
> tag_unimp == 1

save "$saveaddress_data\IRP5_`year'_temp_DELETE2.dta",
> replace
keep if prop_in_firm<. /*& prop_in_firm>0 */

```

```

>
> HY IS THIS COMMENTED OUT??? WHAT IS GOING ON?? */
/* FK to SR: W

>
> K to MK: I have attempted to correct this somewhat only using the pieterse kerr and
> 3601 measures. */
/* FK to SR, F

>
> ds kerr_income a3601_income
>
>
> foreach var in `r(varlist)' {
>     replace `var' = `var'*prop_in_firm
> }
>
> drop prop_before_firm prop_after_firm ind_firm_before ind_firm
> _after
> drop if firm_end==. & firm_start==.
> save "$saveaddress_data\IRP5_`year'_infirm.dta", replace
> local altbefore = "after"
> local altafter = "before"
> foreach reldate in before after {
>     use "$saveaddress_data\IRP5_`year'_temp_DELETE2.dta",
> clear
>     keep if prop_`reldate'_firm>0 & prop_`reldate'_firm<.
>
>     ds kerr_income a3601_income
>     foreach var in `r(varlist)' {
>         replace `var' = `var'*prop_`reldate'_firm
>     }
>     drop prop_in_firm prop_`alt`reldate'`_firm ind_firm in
> d_firm_`alt`reldate'`
> gen tag_`reldate'=1
> *drop if firm_end==. & firm_start==.
> save "$saveaddress_data\IRP5_`year'_`reldate'.dta", re
> place
>
> }

forvalues year= 2010/2018 {
>     use "$saveaddress_data\IRP5_`year'_infirm.dta", clear
>     local pre = `year'-1
>     local post = `year'+1
>     capture confirm file "$saveaddress_data\IRP5_`post'_'_
> before.dta"
>     if _rc==0 {
>         append using "$saveaddress_data\IRP5_`post'_'_be
> fore.dta", force
>     }
>     capture confirm file "$saveaddress_data\IRP5_`pre'_'_a
> fter.dta"
>     if _rc==0 {
>         append using "$saveaddress_data\IRP5_`pre'_'_aft
> er.dta", force
>     }
>     ren taxyear old_taxyear
>     gen taxyear = `year'
>
>     preserve
>
>     use "$citirp5_v4" , clear
>     keep if taxyear == `year'
>     tempfile cit
>     save `cit'
>     restore
>     merge m:1 taxrefno taxyear using `cit'

```

```
                keep if _merge==3
                save "$saveaddress_data\merged_`year'.dta", replace
            }

            use "$saveaddress_data\merged_2018", clear
forval i = 2017(-1)2010{
append using "$saveaddress_data\merged_`i'.dta"
}

save "$saveaddress_data\merged_ind_panel", replace
cap log close
```