```
clear matrix
clear
set more off

set scheme s1color

cd "C:\Users\mario\Documents\Undocu_Mismatch_Wage_Research_2024 Data"
use "(ML)EO_Final_Sample.dta", clear




***elig_year variable creation***
gen eventyear = year
label define eventyr 2009 "2009" 2010 "2010" 2011 "2011" 2012 "2012"
2013 "2013" ///
          2014 "2014" 2015 "2015" 2016 "2016" 2017 "2017" 2018 "2018"
2019 "2019"
label values eventyear eventyr

forvalues y=2013(1)2019 {
        gen elig_year`y' = elig*(eventyear==`y')
}
drop elig_year2016

forvalues y=2013(1)2019 {
        gen undocu_year`y' = undocu_logit*(eventyear==`y')
}
drop undocu_year2016

***Mismatch and other regression covariate modifications/labeling***
gen hmatch = 1 if hundermatched==1
replace hmatch=2 if hundermatched==0 & hovermatched==0
replace hmatch=3 if  hovermatched==1

gen elig_stem=elig*stem_deg
gen post_stem=post*stem_deg
gen elig_post_stem=elig*post*stem_deg


label define hmatch_label 1 "Hundermatched" 2 "Hmatched" 3
"Hovermatched"
label values hmatch hmatch_label

replace post=0 if year==2012
replace immig_by_ten=1 if bpl_foreign==0


gen annual_total_dummy = 0 if annual_total<0
replace annual_total_dummy = 1 if annual_total==0
```

```
replace annual_total_dummy = 2 if annual_total>0

label define annual_total_label 0 "Exclusive" 1 "Neutral" 2
"Inclusive"
label values annual_total_dummy annual_total_label

gen exclusive = 1 if annual_total<0
replace exclusive = 0 if annual_total>=0

gen inclusive = 1 if annual_total>0
replace inclusive = 0 if annual_total<=0

gen undocu_inclusive = undocu*inclusive
gen undocu_knn_inclusive = undocu_knn*inclusive
gen undocu_rf_inclusive = undocu_rf*inclusive

gen undocu_exclusive = undocu*exclusive


gen undocu_annual_total = undocu_logit*annual_total


gen everify_inclusive=(e_verify==2)
gen undocu_everify=undocu*everify_inclusive
gen undocu_knn_everify=undocu_knn*everify_inclusive
gen undocu_rf_everify=undocu_rf*everify_inclusive

gen license_inclusive=(professional_licensure==2)
gen undocu_license=undocu*license_inclusive
gen undocu_knn_license=undocu_knn*license_inclusive
gen undocu_rf_license=undocu_rf*license_inclusive

gen drive_inclusive=(drivers_license==2)
gen undocu_drive=undocu*drive_inclusive
gen undocu_knn_drive=undocu_knn*drive_inclusive
gen undocu_rf_drive=undocu_rf*drive_inclusive



clear matrix
set more off

xtset statefip
/*global individual_ipc b1.pub_insurance_immigrant_kids
b1.prenatal_care_pregnant_immigrant
b1.pub_insurance_pregnant_immigrant
b1.pub_insurance_immigrant_older_ad
b1.food_assistance_for_lpr_adults         b1.tuition_equity
b1.financial_aid          b1.blocks_enrollment
b1.professional_licensure         b1.drivers_license       b1.omnibus
```

```
        b1.cooperation_federal_immigration        b1.e_verify
b1.secure_communities_participated

save "(ML) Pre Regression sample", replace
*/



global covars_redhdfe  hisp male gov_worker bpl_foreign immig_by_ten
nonfluent yrsed metropolitan


**********************************************************************
******************
******Individual Mismatch regressions with Degree Interactions
************************
**********************************************************************
******************
clear matrix
set more off
eststo clear

*Vertical Mismatch
reghdfe vmismatched hundermatched hovermatched
undocu##ib5.degfield_broader $covars_reghdfe    [pweight=perwt],
absorb(statefip##year age ) vce(cluster statefip)
estadd ysumm
eststo logical_vmismatch

reghdfe vmismatched hundermatched hovermatched
undocu_knn##ib5.degfield_broader $covars_reghdfe   [pweight=perwt],
absorb(statefip##year age ) vce(cluster statefip)
estadd ysumm
eststo knn_vmismatch

reghdfe vmismatched hundermatched hovermatched
undocu_rf##ib5.degfield_broader $covars_reghdfe    [pweight=perwt],
absorb(statefip##year age ) vce(cluster statefip)
estadd ysumm
eststo rf_vmismatch

cd "C:
\Users\mario\Documents\GitHub\Undocu_Mismatch_Wage_Research_2024\(ML)
Appendix"
esttab logical_vmismatch knn_vmismatch rf_vmismatch using
vmismatch_regressions_degree.tex, replace label booktabs
drop($covars_reghdfe) ///
rename(1.degfield_broader "STEM" 2.degfield_broader "STEM Related"
3.degfield_broader "Business" 4.degfield_broader "Education" ///
1.undocu_rf  "Undocumented" 1.undocu_rf#2.degfield_broader
```

```
"Undocumented x STEM Related" 1.undocu_rf#3.degfield_broader
"Undocumented x Business" 1.undocu_rf#4.degfield_broader "Undocumented
x Education" 1.undocu_rf#1.degfield_broader "Undocumented x STEM" ///
1.undocu_knn  "Undocumented" 1.undocu_knn#2.degfield_broader
"Undocumented x STEM Related"  1.undocu_knn#3.degfield_broader
"Undocumented x Business" 1.undocu_knn#4.degfield_broader
"Undocumented x Education" 1.undocu_knn#1.degfield_broader
"Undocumented x STEM" ///
1.undocu "Undocumented" 1.undocu#2.degfield_broader "Undocumented x
STEM Related"  1.undocu#3.degfield_broader "Undocumented x Business"
1.undocu#4.degfield_broader "Undocumented x Education"
1.undocu#1.degfield_broader "Undocumented x STEM") ///
stats( ymean r2 N  , labels(  "Mean of Dep. Var." "R-squared" N )
fmt(    %9.2f %9.2f %9.0fc ) ) ///
title("Regressions of Undocumented Status on Vmismatch (Degree
Interaction Terms)") ///
mlabel("Logical edits"  "KNN" "RF") ///
r2(4) b(4) se(4) brackets star(* .1 ** 0.05 *** 0.01) ///
note("Additional controls include:") ///
addn("dummy age indicators, gender, race/ethnicity, metropolitan
residence, statefip##year age" ///
        "government occupation, English-speaking fluency, foreign
born, immigration by age 10," ///
        "STEM degree indicators, years of schooling, state and year
interaction fixed effects." ///
        "Robust standard errors are all clustered by state.")


*Horizontal Undermatch
reghdfe  hundermatched vmismatched undocu##ib5.degfield_broader
$covars_reghdfe   [pweight=perwt], absorb(statefip##year age )
vce(cluster statefip)
estadd ysumm
eststo logical_hunder

reghdfe  hundermatched vmismatched undocu_knn##ib5.degfield_broader
$covars_reghdfe    [pweight=perwt], absorb(statefip##year age )
vce(cluster statefip)
estadd ysumm
eststo knn_hunder

reghdfe  hundermatched vmismatched undocu_rf##ib5.degfield_broader
$covars_reghdfe   [pweight=perwt], absorb(statefip##year age )
vce(cluster statefip)
estadd ysumm
eststo rf_hunder

*cd "C:
\Users\mario\Documents\GitHub\Undocu_Mismatch_Wage_Research_2024\Undoc
u Research Figures ML"
```

```
esttab logical_hunder knn_hunder rf_hunder using
hunder_regressions_degree.tex, replace label booktabs
drop($covars_reghdfe) ///
rename(1.degfield_broader "STEM" 2.degfield_broader "STEM Related"
3.degfield_broader "Business" 4.degfield_broader "Education" ///
1.undocu_rf  "Undocumented" 1.undocu_rf#2.degfield_broader
"Undocumented x STEM Related" 1.undocu_rf#3.degfield_broader
"Undocumented x Business" 1.undocu_rf#4.degfield_broader "Undocumented
x Education" 1.undocu_rf#1.degfield_broader "Undocumented x STEM" ///
1.undocu_knn  "Undocumented" 1.undocu_knn#2.degfield_broader
"Undocumented x STEM Related"  1.undocu_knn#3.degfield_broader
"Undocumented x Business" 1.undocu_knn#4.degfield_broader
"Undocumented x Education" 1.undocu_knn#1.degfield_broader
"Undocumented x STEM" ///
1.undocu "Undocumented" 1.undocu#2.degfield_broader "Undocumented x
STEM Related"  1.undocu#3.degfield_broader "Undocumented x Business"
1.undocu#4.degfield_broader "Undocumented x Education"
1.undocu#1.degfield_broader "Undocumented x STEM") ///
stats( ymean r2 N  , labels(  "Mean of Dep. Var." "R-squared" N )
fmt(    %9.2f %9.2f %9.0fc ) ) ///
title("Regressions of Undocumented Status on Horizontal Undermatch
(Degree Interaction Terms)") ///
mlabel("Logical edits"  "KNN" "RF") ///
r2(4) b(4) se(4) brackets star(* .1 ** 0.05 *** 0.01) ///
note("Additional controls include:") ///
addn("dummy age indicators, gender, race/ethnicity, metropolitan
residence, statefip##year age" ///
        "government occupation, English-speaking fluency, foreign
born, immigration by age 10," ///
        "STEM degree indicators, years of schooling, state and year
interaction fixed effects." ///
        "Robust standard errors are all clustered by state.")


*Log Wages
reghdfe ln_adj vmismatched hundermatched hovermatched
undocu##ib5.degfield_broader $covars_reghdfe   [pweight=perwt],
absorb(statefip##year age ) vce(cluster statefip)
estadd ysumm
eststo logical_wage

reghdfe ln_adj vmismatched hundermatched hovermatched
undocu_rf##ib5.degfield_broader $covars_reghdfe    [pweight=perwt],
absorb(statefip##year age ) vce(cluster statefip)
estadd ysumm
eststo rf_wage

reghdfe ln_adj vmismatched hundermatched hovermatched
undocu_knn##ib5.degfield_broader $covars_reghdfe   [pweight=perwt],
absorb(statefip##year age ) vce(cluster statefip)
```

```
estadd ysumm
eststo knn_wage

*cd "C:
\Users\mario\Documents\GitHub\Undocu_Mismatch_Wage_Research_2024\Undoc
u Research Figures ML"
esttab logical_wage knn_wage rf_wage using
wage_regressions_degree.tex, replace label booktabs
drop($covars_reghdfe) ///
rename(1.degfield_broader "STEM" 2.degfield_broader "STEM Related"
3.degfield_broader "Business" 4.degfield_broader "Education" ///
1.undocu_rf  "Undocumented" 1.undocu_rf#2.degfield_broader
"Undocumented x STEM Related" 1.undocu_rf#3.degfield_broader
"Undocumented x Business" 1.undocu_rf#4.degfield_broader "Undocumented
x Education" 1.undocu_rf#1.degfield_broader "Undocumented x STEM" ///
1.undocu_knn  "Undocumented" 1.undocu_knn#2.degfield_broader
"Undocumented x STEM Related"  1.undocu_knn#3.degfield_broader
"Undocumented x Business" 1.undocu_knn#4.degfield_broader
"Undocumented x Education" 1.undocu_knn#1.degfield_broader
"Undocumented x STEM" ///
1.undocu "Undocumented" 1.undocu#2.degfield_broader "Undocumented x
STEM Related"  1.undocu#3.degfield_broader "Undocumented x Business"
1.undocu#4.degfield_broader "Undocumented x Education"
1.undocu#1.degfield_broader "Undocumented x STEM") ///
stats( ymean r2 N  , labels(  "Mean of Dep. Var." "R-squared" N )
fmt(    %9.2f %9.2f %9.0fc ) ) ///
title("Regressions of Undocumented Status on Log Wages (Degree
Interaction Terms)") ///
mlabel("Logical edits"  "KNN" "RF") ///
r2(4) b(4) se(4) brackets star(* .1 ** 0.05 *** 0.01) ///
note("Additional controls include:") ///
addn("dummy age indicators, gender, race/ethnicity, metropolitan
residence, statefip##year age" ///
        "government occupation, English-speaking fluency, foreign
born, immigration by age 10," ///
        "STEM degree indicators, years of schooling, state and year
interaction fixed effects." ///
        "Robust standard errors are all clustered by state.")


**Coefficient Plots****

*vertical mismatch
coefplot (logical_vmismatch, label(Logical Edits) ) (knn_vmismatch,
label(KNN) ) (rf_vmismatch, label(Random Forest) ) ///
 ||, keep(1.undocu* *.undocu*#*.degfield_broader)  xline(0)
byopts( cols(1)) ///
rename(1.undocu_rf = "Undocumented" 1.undocu_rf#2.degfield_broader=
"Undocumented x STEM Related" 1.undocu_rf#3.degfield_broader=
"Undocumented x Business" 1.undocu_rf#4.degfield_broader=
```

```
"Undocumented x Education" 1.undocu_rf#1.degfield_broader=
"Undocumented x STEM" ///
1.undocu_knn = "Undocumented" 1.undocu_knn#2.degfield_broader=
"Undocumented x STEM Related"  1.undocu_knn#3.degfield_broader=
"Undocumented x Business" 1.undocu_knn#4.degfield_broader=
"Undocumented x Education" 1.undocu_knn#1.degfield_broader=
"Undocumented x STEM" ///
1.undocu= "Undocumented" 1.undocu#2.degfield_broader= "Undocumented x
STEM Related"  1.undocu#3.degfield_broader= "Undocumented x Business"
1.undocu#4.degfield_broader= "Undocumented x Education"
1.undocu#1.degfield_broader= "Undocumented x STEM") ///
 xline(0) title("Vertical Mismatch")
  graph export degree_vmismatch.png, replace


 *horizontal undermatch
coefplot (logical_hunder, label(Logical Edits) ) (knn_hunder,
label(KNN) ) (rf_hunder, label(Random Forest) ) ///
 ||, keep(1.undocu* *.undocu*#*.degfield_broader)  xline(0)
byopts( cols(1)) ///
rename(1.undocu_rf = "Undocumented" 1.undocu_rf#2.degfield_broader=
"Undocumented x STEM Related" 1.undocu_rf#3.degfield_broader=
"Undocumented x Business" 1.undocu_rf#4.degfield_broader=
"Undocumented x Education" 1.undocu_rf#1.degfield_broader=
"Undocumented x STEM" ///
1.undocu_knn = "Undocumented" 1.undocu_knn#2.degfield_broader=
"Undocumented x STEM Related"  1.undocu_knn#3.degfield_broader=
"Undocumented x Business" 1.undocu_knn#4.degfield_broader=
"Undocumented x Education" 1.undocu_knn#1.degfield_broader=
"Undocumented x STEM" ///
1.undocu= "Undocumented" 1.undocu#2.degfield_broader= "Undocumented x
STEM Related"  1.undocu#3.degfield_broader= "Undocumented x Business"
1.undocu#4.degfield_broader= "Undocumented x Education"
1.undocu#1.degfield_broader= "Undocumented x STEM") ///
 xline(0) title("Horizontal Undermatch")
   graph export degree_hunder.png, replace

  *wages
coefplot (logical_wage, label(Logical Edits) ) (knn_wage, label(KNN) )
(rf_wage, label(Random Forest)) ///
 ||, keep(1.undocu* *.undocu*#*.degfield_broader)  xline(0)
byopts( cols(1)) ///
rename(1.undocu_rf = "Undocumented" 1.undocu_rf#2.degfield_broader=
"Undocumented x STEM Related" 1.undocu_rf#3.degfield_broader=
"Undocumented x Business" 1.undocu_rf#4.degfield_broader=
"Undocumented x Education" 1.undocu_rf#1.degfield_broader=
"Undocumented x STEM" ///
1.undocu_knn = "Undocumented" 1.undocu_knn#2.degfield_broader=
"Undocumented x STEM Related"  1.undocu_knn#3.degfield_broader=
"Undocumented x Business" 1.undocu_knn#4.degfield_broader=
```

```
"Undocumented x Education" 1.undocu_knn#1.degfield_broader=
"Undocumented x STEM" ///
1.undocu= "Undocumented" 1.undocu#2.degfield_broader= "Undocumented x
STEM Related"  1.undocu#3.degfield_broader= "Undocumented x Business"
1.undocu#4.degfield_broader= "Undocumented x Education"
1.undocu#1.degfield_broader= "Undocumented x STEM") ///
 xline(0) title("Log Wages")
   graph export degree_wage.png, replace


 *all together
coefplot (logical_vmismatch, label(Logical Edits) ) (knn_vmismatch,
label(KNN) ) (rf_vmismatch, label(Random Forest) ), bylabel(Vertical
Mismatch)  ///
||  (logical_hunder, label(Logical Edits) ) (knn_hunder, label(KNN) )
(rf_hunder, label(Random Forest) ), bylabel(Horizontal Undermatch) ///
||  (logical_wage, label(Logical Edits) ) (knn_wage, label(KNN) )
(rf_wage, label(Random Forest)), bylabel(Log Wage) ///
||, keep(1.undocu* *.undocu*#*.degfield_broader) xline(0)  ///
rename(1.undocu_rf = "Undocumented" 1.undocu_rf#2.degfield_broader=
"Undocumented x STEM Related" 1.undocu_rf#3.degfield_broader=
"Undocumented x Business" 1.undocu_rf#4.degfield_broader=
"Undocumented x Education" 1.undocu_rf#1.degfield_broader=
"Undocumented x STEM" ///
1.undocu_knn = "Undocumented" 1.undocu_knn#2.degfield_broader=
"Undocumented x STEM Related"  1.undocu_knn#3.degfield_broader=
"Undocumented x Business" 1.undocu_knn#4.degfield_broader=
"Undocumented x Education" 1.undocu_knn#1.degfield_broader=
"Undocumented x STEM" ///
1.undocu= "Undocumented" 1.undocu#2.degfield_broader= "Undocumented x
STEM Related"  1.undocu#3.degfield_broader= "Undocumented x Business"
1.undocu#4.degfield_broader= "Undocumented x Education"
1.undocu#1.degfield_broader= "Undocumented x STEM")
graph export deg_coeff.png, replace

*************************************************************************
******************
******Individual Mismatch regressions with IPC Interactions
*************************
*************************************************************************
******************
clear matrix
set more off
eststo clear

*Vertical Mismatch
reghdfe vmismatched hundermatched hovermatched undocu undocu_inclusive
$covars_reghdfe    [pweight=perwt], absorb(statefip##year age
degfield_broader) vce(cluster statefip)
estadd ysumm
```

```
eststo logical_vmismatch

reghdfe vmismatched hundermatched hovermatched undocu_knn
undocu_knn_inclusive $covars_reghdfe   [pweight=perwt],
absorb(statefip##year age degfield_broader ) vce(cluster statefip)
estadd ysumm
eststo knn_vmismatch

reghdfe vmismatched hundermatched hovermatched undocu_rf
undocu_rf_inclusive $covars_reghdfe    [pweight=perwt],
absorb(statefip##year age degfield_broader) vce(cluster statefip)
estadd ysumm
eststo rf_vmismatch

*cd "C:
\Users\mario\Documents\GitHub\Undocu_Mismatch_Wage_Research_2024\Undoc
u Research Figures ML"
esttab logical_vmismatch knn_vmismatch rf_vmismatch using
vmismatch_regressions_ipc.tex, replace label booktabs
drop($covars_reghdfe) ///
rename(undocu  "Undocumented" undocu_knn  "Undocumented" undocu_rf
"Undocumented" undocu_inclusive "Undocumented x Inclusive"
undocu_knn_inclusive "Undocumented x Inclusive" undocu_rf_inclusive
"Undocumented x Inclusive") ///
stats( ymean r2 N  , labels(  "Mean of Dep. Var." "R-squared" N )
fmt(    %9.2f %9.2f %9.0fc ) ) ///
title("Regressions of Undocumented Status on Vmismatch (IPC
Interaction Terms)") ///
mlabel("Logical edits"  "KNN" "RF") ///
r2(4) b(4) se(4) brackets star(* .1 ** 0.05 *** 0.01) ///
note("Additional controls include:") ///
addn("dummy age indicators, gender, race/ethnicity, metropolitan
residence, statefip##year age" ///
        "government occupation, English-speaking fluency, foreign
born, immigration by age 10," ///
        "STEM degree indicators, years of schooling, state and year
interaction fixed effects." ///
        "Robust standard errors are all clustered by state.")


*Horizontal Undermatch
reghdfe  hundermatched vmismatched undocu undocu_inclusive
$covars_reghdfe   [pweight=perwt], absorb(statefip##year age
degfield_broader) vce(cluster statefip)
estadd ysumm
eststo logical_hunder

reghdfe  hundermatched vmismatched undocu_knn undocu_knn_inclusive
$covars_reghdfe    [pweight=perwt], absorb(statefip##year age
degfield_broader) vce(cluster statefip)
```

```
estadd ysumm
eststo knn_hunder

reghdfe  hundermatched vmismatched undocu_rf undocu_rf_inclusive
$covars_reghdfe   [pweight=perwt], absorb(statefip##year age
degfield_broader) vce(cluster statefip)
estadd ysumm
eststo rf_hunder

*cd "C:
\Users\mario\Documents\GitHub\Undocu_Mismatch_Wage_Research_2024\Undoc
u Research Figures ML"
esttab logical_hunder knn_hunder rf_hunder using
hunder_regressions_ipc.tex, replace label booktabs
drop($covars_reghdfe) ///
rename(undocu  "Undocumented" undocu_knn  "Undocumented" undocu_rf
"Undocumented" undocu_inclusive "Undocumented x Inclusive"
undocu_knn_inclusive "Undocumented x Inclusive" undocu_rf_inclusive
"Undocumented x Inclusive") ///
stats( ymean r2 N  , labels(  "Mean of Dep. Var." "R-squared" N )
fmt(    %9.2f %9.2f %9.0fc ) ) ///
title("Regressions of Undocumented Status on Horizontal Undermatch
(IPC Interaction Terms)") ///
mlabel("Logical edits"  "KNN" "RF") ///
r2(4) b(4) se(4) brackets star(* .1 ** 0.05 *** 0.01) ///
note("Additional controls include:") ///
addn("dummy age indicators, gender, race/ethnicity, metropolitan
residence, statefip##year age" ///
        "government occupation, English-speaking fluency, foreign
born, immigration by age 10," ///
        "STEM degree indicators, years of schooling, state and year
interaction fixed effects." ///
        "Robust standard errors are all clustered by state.")


*Log Wages
reghdfe ln_adj vmismatched hundermatched hovermatched undocu
undocu_inclusive $covars_reghdfe   [pweight=perwt],
absorb(statefip##year age degfield_broader) vce(cluster statefip)
estadd ysumm
eststo logical_wage

reghdfe ln_adj vmismatched hundermatched hovermatched undocu_knn
undocu_knn_inclusive $covars_reghdfe    [pweight=perwt],
absorb(statefip##year age degfield_broader) vce(cluster statefip)
estadd ysumm
eststo knn_wage

reghdfe ln_adj vmismatched hundermatched hovermatched undocu_rf
undocu_rf_inclusive $covars_reghdfe   [pweight=perwt],
```

```
absorb(statefip##year age degfield_broader) vce(cluster statefip)
estadd ysumm
eststo rf_wage

*cd "C:
\Users\mario\Documents\GitHub\Undocu_Mismatch_Wage_Research_2024\Undoc
u Research Figures ML"
esttab logical_wage knn_wage rf_wage using wage_regressions_ipc.tex,
replace label booktabs drop($covars_reghdfe) ///
rename(undocu  "Undocumented" undocu_knn  "Undocumented" undocu_rf
"Undocumented" undocu_inclusive "Undocumented x Inclusive"
undocu_knn_inclusive "Undocumented x Inclusive" undocu_rf_inclusive
"Undocumented x Inclusive") ///
stats( ymean r2 N  , labels(  "Mean of Dep. Var." "R-squared" N )
fmt(    %9.2f %9.2f %9.0fc ) ) ///
title("Regressions of Undocumented Status on Log Wages (IPC
Interaction Terms)") ///
mlabel("Logical edits"  "KNN" "RF") ///
r2(4) b(4) se(4) brackets star(* .1 ** 0.05 *** 0.01) ///
note("Additional controls include:") ///
addn("dummy age indicators, gender, race/ethnicity, metropolitan
residence, statefip##year age" ///
        "government occupation, English-speaking fluency, foreign
born, immigration by age 10," ///
        "STEM degree indicators, years of schooling, state and year
interaction fixed effects." ///
        "Robust standard errors are all clustered by state.")


**Coefficient Plots****
cd "C:
\Users\mario\Documents\GitHub\Undocu_Mismatch_Wage_Research_2024\(ML)
Figures"

*vertical mismatch
coefplot (logical_vmismatch, label(Logical Edits) ) (knn_vmismatch,
label(KNN) ) (rf_vmismatch, label(Random Forest) ) ///
 ||, drop($covars_reghdfe hundermatched hovermatched )  xline(0)   ///
rename(undocu_rf = "Undocumented" undocu= "Undocumented" undocu_knn=
"Undocumented" ///
undocu_inclusive= "Undocumented x Inclusive" undocu_knn_inclusive=
"Undocumented x Inclusive"  undocu_rf_inclusive= "Undocumented x
Inclusive" ) ///
 xline(0) title("Vertical Mismatch")
  graph save ipc_vmismatch, replace


 *horizontal undermatch
coefplot (logical_hunder, label(Logical Edits) ) (knn_hunder,
label(KNN) ) (rf_hunder, label(Random Forest) ) ///
```

```
 ||, drop($covars_reghdfe vmismatched)  xline(0)   ///
rename(undocu_rf = "Undocumented" undocu= "Undocumented" undocu_knn=
"Undocumented" ///
undocu_inclusive= "Undocumented x Inclusive" undocu_knn_inclusive=
"Undocumented x Inclusive"  undocu_rf_inclusive= "Undocumented x
Inclusive" ) ///
 xline(0) title("Horizontal Undermatch")
   graph save ipc_hunder, replace

 *wages
coefplot (logical_wage, label(Logical Edits) ) (knn_wage, label(KNN) )
(rf_wage, label(Random Forest)) ///
||, drop($covars_reghdfe vmismatched hundermatched hovermatched)
xline(0)   ///
rename(undocu_rf = "Undocumented" undocu= "Undocumented" undocu_knn=
"Undocumented" ///
undocu_inclusive= "Undocumented x Inclusive" undocu_knn_inclusive=
"Undocumented x Inclusive"  undocu_rf_inclusive= "Undocumented x
Inclusive" ) ///
 xline(0) title("Log Wage")
   graph save ipc_wage, replace

*all together
coefplot (logical_vmismatch, label(Logical Edits) ) (knn_vmismatch,
label(KNN) ) (rf_vmismatch, label(Random Forest) ), bylabel(Vertical
Mismatch)  ///
||  (logical_hunder, label(Logical Edits) ) (knn_hunder, label(KNN) )
(rf_hunder, label(Random Forest) ), bylabel(Horizontal Undermatch) ///
||  (logical_wage, label(Logical Edits) ) (knn_wage, label(KNN) )
(rf_wage, label(Random Forest)), bylabel(Log Wage) ///
||, drop($covars_reghdfe vmismatched hundermatched hovermatched)
xline(0)  ///
rename(undocu_rf = "Undocumented" undocu= "Undocumented" undocu_knn=
"Undocumented" ///
undocu_inclusive= "Undocumented x Inclusive" undocu_knn_inclusive=
"Undocumented x Inclusive"  undocu_rf_inclusive= "Undocumented x
Inclusive" )
graph export ipc_coeff.png, replace



**********************************************************************
******************
******Individual Mismatch regressions with Individual Policy
Interactions ************************
**********************************************************************
******************
clear matrix
set more off
eststo clear
```

```
*Vertical Mismatch
reghdfe vmismatched hundermatched hovermatched undocu undocu_everify
undocu_license undocu_drive  $covars_reghdfe    [pweight=perwt],
absorb(statefip##year age degfield_broader) vce(cluster statefip)
estadd ysumm
eststo logical_vmismatch

reghdfe vmismatched hundermatched hovermatched undocu_knn
undocu_knn_everify undocu_knn_license undocu_knn_drive $covars_reghdfe
[pweight=perwt], absorb(statefip##year age degfield_broader )
vce(cluster statefip)
estadd ysumm
eststo knn_vmismatch

reghdfe vmismatched hundermatched hovermatched undocu_rf
undocu_rf_everify undocu_rf_license undocu_rf_drive $covars_reghdfe
[pweight=perwt], absorb(statefip##year age degfield_broader)
vce(cluster statefip)
estadd ysumm
eststo rf_vmismatch

*cd "C:
\Users\mario\Documents\GitHub\Undocu_Mismatch_Wage_Research_2024\Undoc
u Research Figures ML"
esttab logical_vmismatch knn_vmismatch rf_vmismatch using
vmismatch_regressions_policies.tex, replace label booktabs
drop($covars_reghdfe) ///
rename(undocu  "Undocumented" undocu_knn  "Undocumented" undocu_rf
"Undocumented" ///
undocu_everify "Undocumented x Inclusive Everify" undocu_knn_everify
"Undocumented x Inclusive Everify" undocu_rf_everify "Undocumented x
Inclusive Everify"  ///
undocu_license "Undocumented x Inclusive OCC" undocu_knn_license
"Undocumented x Inclusive OCC" undocu_rf_license "Undocumented x
Inclusive OCC"  ///
undocu_drive "Undocumented x Inclusive Drive" undocu_knn_drive
"Undocumented x Inclusive Drive" undocu_rf_drive "Undocumented x
Inclusive Drive" ) ///
stats( ymean r2 N  , labels(  "Mean of Dep. Var." "R-squared" N )
fmt(    %9.2f %9.2f %9.0fc ) ) ///
title("Regressions of Undocumented Status on Vmismatch (Policy
Interaction Terms)") ///
mlabel("Logical edits"  "KNN" "RF") ///
r2(4) b(4) se(4) brackets star(* .1 ** 0.05 *** 0.01) ///
note("Additional controls include:") ///
addn("dummy age indicators, gender, race/ethnicity, metropolitan
residence, statefip##year age" ///
        "government occupation, English-speaking fluency, foreign
born, immigration by age 10," ///
```

```
        "STEM degree indicators, years of schooling, state and year
interaction fixed effects." ///
        "Robust standard errors are all clustered by state.")


*Horizontal Undermatch
reghdfe  hundermatched vmismatched undocu undocu_everify
undocu_license undocu_drive  $covars_reghdfe   [pweight=perwt],
absorb(statefip##year age degfield_broader) vce(cluster statefip)
estadd ysumm
eststo logical_hunder

reghdfe  hundermatched vmismatched undocu_knn undocu_knn_everify
undocu_knn_license undocu_knn_drive $covars_reghdfe
[pweight=perwt], absorb(statefip##year age degfield_broader)
vce(cluster statefip)
estadd ysumm
eststo knn_hunder

reghdfe  hundermatched vmismatched undocu_rf undocu_rf_everify
undocu_rf_license undocu_rf_drive $covars_reghdfe   [pweight=perwt],
absorb(statefip##year age degfield_broader) vce(cluster statefip)
estadd ysumm
eststo rf_hunder

*cd "C:
\Users\mario\Documents\GitHub\Undocu_Mismatch_Wage_Research_2024\Undoc
u Research Figures ML"
esttab logical_hunder knn_hunder rf_hunder using
hunder_regressions_policies.tex, replace label booktabs
drop($covars_reghdfe) ///
rename(undocu  "Undocumented" undocu_knn  "Undocumented" undocu_rf
"Undocumented" ///
undocu_everify "Undocumented x Inclusive Everify" undocu_knn_everify
"Undocumented x Inclusive Everify" undocu_rf_everify "Undocumented x
Inclusive Everify"  ///
undocu_license "Undocumented x Inclusive OCC" undocu_knn_license
"Undocumented x Inclusive OCC" undocu_rf_license "Undocumented x
Inclusive OCC"  ///
undocu_drive "Undocumented x Inclusive Drive" undocu_knn_drive
"Undocumented x Inclusive Drive" undocu_rf_drive "Undocumented x
Inclusive Drive" ) ///
stats( ymean r2 N  , labels(  "Mean of Dep. Var." "R-squared" N )
fmt(    %9.2f %9.2f %9.0fc ) ) ///
title("Regressions of Undocumented Status on Horizontal Undermatch
(Policy Interaction Terms)") ///
mlabel("Logical edits"  "KNN" "RF") ///
r2(4) b(4) se(4) brackets star(* .1 ** 0.05 *** 0.01) ///
note("Additional controls include:") ///
addn("dummy age indicators, gender, race/ethnicity, metropolitan
```

```
residence, statefip##year age" ///
        "government occupation, English-speaking fluency, foreign
born, immigration by age 10," ///
        "STEM degree indicators, years of schooling, state and year
interaction fixed effects." ///
        "Robust standard errors are all clustered by state.")


*Log Wages
reghdfe ln_adj vmismatched hundermatched hovermatched undocu
undocu_everify undocu_license undocu_drive  $covars_reghdfe
[pweight=perwt], absorb(statefip##year age degfield_broader)
vce(cluster statefip)
estadd ysumm
eststo logical_wage

reghdfe ln_adj vmismatched hundermatched hovermatched undocu_knn
undocu_knn_everify undocu_knn_license undocu_knn_drive $covars_reghdfe
[pweight=perwt], absorb(statefip##year age degfield_broader)
vce(cluster statefip)
estadd ysumm
eststo knn_wage

reghdfe ln_adj vmismatched hundermatched hovermatched undocu_rf
undocu_rf_everify undocu_rf_license undocu_rf_drive $covars_reghdfe
[pweight=perwt], absorb(statefip##year age degfield_broader)
vce(cluster statefip)
estadd ysumm
eststo rf_wage

*cd "C:
\Users\mario\Documents\GitHub\Undocu_Mismatch_Wage_Research_2024\Undoc
u Research Figures ML"
esttab logical_wage knn_wage rf_wage using
wage_regressions_policies.tex, replace label booktabs
drop($covars_reghdfe) ///
rename(undocu  "Undocumented" undocu_knn  "Undocumented" undocu_rf
"Undocumented" ///
undocu_everify "Undocumented x Inclusive Everify" undocu_knn_everify
"Undocumented x Inclusive Everify" undocu_rf_everify "Undocumented x
Inclusive Everify"  ///
undocu_license "Undocumented x Inclusive OCC" undocu_knn_license
"Undocumented x Inclusive OCC" undocu_rf_license "Undocumented x
Inclusive OCC"  ///
undocu_drive "Undocumented x Inclusive Drive" undocu_knn_drive
"Undocumented x Inclusive Drive" undocu_rf_drive "Undocumented x
Inclusive Drive" ) ///
stats( ymean r2 N  , labels(  "Mean of Dep. Var." "R-squared" N )
fmt(    %9.2f %9.2f %9.0fc ) ) ///
title("Regressions of Undocumented Status on Log Wages (Policy
```

```
Interaction Terms)") ///
mlabel("Logical edits"  "KNN" "RF") ///
r2(4) b(4) se(4) brackets star(* .1 ** 0.05 *** 0.01) ///
note("Additional controls include:") ///
addn("dummy age indicators, gender, race/ethnicity, metropolitan
residence, statefip##year age" ///
        "government occupation, English-speaking fluency, foreign
born, immigration by age 10," ///
        "STEM degree indicators, years of schooling, state and year
interaction fixed effects." ///
        "Robust standard errors are all clustered by state.")


**Coefficient Plots****


*vertical mismatch
coefplot (logical_vmismatch, label(Logical Edits) ) (knn_vmismatch,
label(KNN) ) (rf_vmismatch, label(Random Forest) ) ///
 ||, drop($covars_reghdfe hundermatched hovermatched )  xline(0)   ///
rename(undocu_rf = "Undocumented" undocu= "Undocumented" undocu_knn=
"Undocumented" ///
undocu_everify= "Undocumented x Inclusive Everify" undocu_knn_everify=
"Undocumented x Inclusive Everify" undocu_rf_everify= "Undocumented x
Inclusive Everify"  ///
undocu_license= "Undocumented x Inclusive OCC" undocu_knn_license=
"Undocumented x Inclusive OCC" undocu_rf_license= "Undocumented x
Inclusive OCC"  ///
undocu_drive= "Undocumented x Inclusive Drive" undocu_knn_drive=
"Undocumented x Inclusive Drive" undocu_rf_drive= "Undocumented x
Inclusive Drive" ) ///
 xline(0) title("Vertical Mismatch")
  graph save policy_vmismatch, replace


 *horizontal undermatch
coefplot (logical_hunder, label(Logical Edits) ) (knn_hunder,
label(KNN) ) (rf_hunder, label(Random Forest) ) ///
 ||, drop($covars_reghdfe vmismatched)  xline(0)   ///
rename(undocu_rf = "Undocumented" undocu= "Undocumented" undocu_knn=
"Undocumented" ///
undocu_everify= "Undocumented x Inclusive Everify" undocu_knn_everify=
"Undocumented x Inclusive Everify" undocu_rf_everify= "Undocumented x
Inclusive Everify"  ///
undocu_license= "Undocumented x Inclusive OCC" undocu_knn_license=
"Undocumented x Inclusive OCC" undocu_rf_license= "Undocumented x
Inclusive OCC"  ///
undocu_drive= "Undocumented x Inclusive Drive" undocu_knn_drive=
"Undocumented x Inclusive Drive" undocu_rf_drive= "Undocumented x
Inclusive Drive" ) ///
```

```
 xline(0) title("Horizontal Undermatch")
   graph save policy_hunder, replace

 *wages
coefplot (logical_wage, label(Logical Edits) ) (knn_wage, label(KNN) )
(rf_wage, label(Random Forest)) ///
||, drop($covars_reghdfe vmismatched hundermatched hovermatched)
xline(0)   ///
rename(undocu_rf = "Undocumented" undocu= "Undocumented" undocu_knn=
"Undocumented" ///
undocu_everify= "Undocumented x Inclusive Everify" undocu_knn_everify=
"Undocumented x Inclusive Everify" undocu_rf_everify= "Undocumented x
Inclusive Everify"  ///
undocu_license= "Undocumented x Inclusive OCC" undocu_knn_license=
"Undocumented x Inclusive OCC" undocu_rf_license= "Undocumented x
Inclusive OCC"  ///
undocu_drive= "Undocumented x Inclusive Drive" undocu_knn_drive=
"Undocumented x Inclusive Drive" undocu_rf_drive= "Undocumented x
Inclusive Drive" ) ///
 xline(0) title("Log Wage")
   graph save policy_wage, replace

*all together
coefplot (logical_vmismatch, label(Logical Edits) ) (knn_vmismatch,
label(KNN) ) (rf_vmismatch, label(Random Forest) ), bylabel(Vertical
Mismatch)  ///
||  (logical_hunder, label(Logical Edits) ) (knn_hunder, label(KNN) )
(rf_hunder, label(Random Forest) ), bylabel(Horizontal Undermatch) ///
||  (logical_wage, label(Logical Edits) ) (knn_wage, label(KNN) )
(rf_wage, label(Random Forest)), bylabel(Log Wage) ///
||, drop($covars_reghdfe vmismatched hundermatched hovermatched)
xline(0)  ///
rename(undocu_rf = "Undocumented" undocu= "Undocumented" undocu_knn=
"Undocumented" ///
undocu_everify= "Undocumented x Inclusive Everify" undocu_knn_everify=
"Undocumented x Inclusive Everify" undocu_rf_everify= "Undocumented x
Inclusive Everify"  ///
undocu_license= "Undocumented x Inclusive OCC" undocu_knn_license=
"Undocumented x Inclusive OCC" undocu_rf_license= "Undocumented x
Inclusive OCC"  ///
undocu_drive= "Undocumented x Inclusive Drive" undocu_knn_drive=
"Undocumented x Inclusive Drive" undocu_rf_drive= "Undocumented x
Inclusive Drive" )
graph export policy_coeff.png, replace
```