

ECON 613 - APPLIED ECONOMETRICS
SPRING 2021
**Migration Movements Since 2000
& It's Influences**

Rene B. Ulloa
Xiaolin Ding

May 02, 2021

```
library(Hmisc)

## Loading required package: lattice
## Loading required package: survival
## Loading required package: Formula
## Loading required package: ggplot2
##
## Attaching package: 'Hmisc'
## The following objects are masked from 'package:base':
##
##     format.pval, units

library(dplyr)

##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:Hmisc':
##
##     src, summarize
## The following objects are masked from 'package:stats':
##
##     filter, lag
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

library(fastDummies)
library(ggpubr)
library(Hmisc)
library(tidyverse)

## -- Attaching packages ----- tidyverse 1.3.0 --
```

```

## v tibble 3.1.0      v purrr 0.3.4
## v tidyr 1.1.3      v stringr 1.4.0
## v readr 1.4.0      v forcats 0.5.1

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter()   masks stats::filter()
## x dplyr::lag()      masks stats::lag()
## x dplyr::src()      masks Hmisc::src()
## x dplyr::summarize() masks Hmisc::summarize()

nls = read.csv("NLS.csv")
nls = nls[-1,]
for (i in 1:ncol(nls)){
  nls[,i] = as.numeric(nls[,i])
}

for (j in 0:9){
  nls[,paste('inccat',toString(2000+2*j),sep='')] = NA
  incquant=as.numeric(quantile(subset(nls[,8+6*j],nls[,8+6*j]!=0),
                                c(0.3,0.8),na.rm=TRUE))
  for (i in 1:nrow(nls)){
    if (is.na(nls[i,8+6*j])==FALSE & nls[i,8+6*j]<=incquant[1]){
      nls[i,paste('inccat',toString(2000+2*j),sep='')] = 1
    }
    else if (is.na(nls[i,8+6*j])==FALSE & nls[i,8+6*j]>incquant[1] &
             nls[i,8+6*j]<=incquant[2]){
      nls[i,paste('inccat',toString(2000+2*j),sep='')] = 2
    }
    else if (is.na(nls[i,8+6*j])==FALSE & nls[i,8+6*j]>incquant[2]){
      nls[i,paste('inccat',toString(2000+2*j),sep='')] = 3
    }
  }
}

for (j in 0:9){
  nls[,paste('hrccat',toString(2000+2*j),sep='')] = NA
  hrquant=as.numeric(quantile(subset(nls[,10+6*j],nls[,10+6*j]!=0),
                                c(0.2,0.4,0.6,0.8),na.rm=TRUE))
  for (i in 1:nrow(nls)){
    if (is.na(nls[i,10+6*j])==FALSE & nls[i,10+6*j]<=hrquant[1]){
      nls[i,paste('hrccat',toString(2000+2*j),sep='')] = 1
    }
    else if (is.na(nls[i,10+6*j])==FALSE & nls[i,10+6*j]>hrquant[1] &
             nls[i,10+6*j]<=hrquant[2]){
      nls[i,paste('hrccat',toString(2000+2*j),sep='')] = 2
    }
    else if (is.na(nls[i,10+6*j])==FALSE & nls[i,10+6*j]>hrquant[2] &
             nls[i,10+6*j]<=hrquant[3]){
      nls[i,paste('hrccat',toString(2000+2*j),sep='')] = 3
    }
    else if (is.na(nls[i,10+6*j])==FALSE & nls[i,10+6*j]>hrquant[3] &
             nls[i,10+6*j]<=hrquant[4]){
      nls[i,paste('hrccat',toString(2000+2*j),sep='')] = 4
    }
    else if (is.na(nls[i,10+6*j])==FALSE & nls[i,10+6*j]>hrquant[4]){
      nls[i,paste('hrccat',toString(2000+2*j),sep='')] = 5
    }
  }
}

```

```

    }
  }
}

df = data.frame(matrix(ncol=12,nrow=0))
colnames(df) = c('caseid','id','race','sex',"moved",'moves',
                 'grade','inc','famsize','hrwork','inccat','hrcat')
for (i in 0:9){
  d = rbind(setNames(nls[,c(1:4,(5+6*i):(10+6*i),65+i,75+i)],colnames(df)))
  df = rbind(df,d)
}
df$year = NA
for (i in 0:9){
  df$year[(i*nrow(nls)+1):((i+1)*nrow(nls))]=rep(2000+2*i,nrow(nls))
}
df$inc = df$inc/1000

df = dummy_cols(df, select_columns=c('race','sex','year'))
for (i in unique(df$year)){
  df[,paste('moved',toString(i),sep='')] =
    df[,paste('year',toString(i),sep='_')]*df$moved
}
for (i in unique(df$year)){
  df[,paste('moves',toString(i),sep='')] =
    df[,paste('year',toString(i),sep='_')]*df$moves
}
for (i in unique(df$year)){
  df[,paste('grade',toString(i),sep='')] =
    df[,paste('year',toString(i),sep='_')]*df$grade
}
for (i in unique(df$year)){
  df[,paste('inc',toString(i),sep='')] =
    df[,paste('year',toString(i),sep='_')]*df$inc
}
for (i in unique(df$year)){
  df[,paste('famsize',toString(i),sep='')] =
    df[,paste('year',toString(i),sep='_')]*df$famsize
}
for (i in unique(df$year)){
  df[,paste('hrwork',toString(i),sep='')] =
    df[,paste('year',toString(i),sep='_')]*df$hrwork
}

```

1 Abstract

We hope to present a novel approach to assessing the movements of agents from one region to another and seeing what parameters may influence this. The parameters we choose include but are not limited to income, marriage, number of times moved before, job location, race, and age. Some of the inference we invoke in our paper includes instrumental variables, least absolute and least shrinkage operator, and other forms of regression analysis to see what perturbs movement patterns across the United States

2 Introduction

The processes delineating migration are interesting to examine as state and federal policies evolve through time. Our motivation for this paper comes from Henrik Kleven, Camille Landais, and Jakob Egholt Sogaard. Their paper, *Children and Gender Inequality: Evidence from Denmark*, presents findings that we believe we can apply to our paper. In their paper, they examine the impacts of children on gender inequality in the labor market. The authors use three different empirical strategies. The authors first use an event study approach based on labor market outcomes of women relative to men around the birth of the first child. Second, they decomposed gender inequality into what is due to children and what stems from other factors. They also showed the evolution of this composition over time. Finally, they consider the intergenerational transmission of child penalties, which is the impact of gender identity norms among maternal and paternal grandparents on mothers' child penalties. Their findings present significant evidence that gender inequality in the labor market can be mainly attributed to children, and this inequality is negatively correlated to the maternal grandmother's labor supply. We hope to use our insights from their paper with this one.

3 Impacts of Income

The first step in the analysis is to examine the impacts of income in order to see whether there is a relationship between income and migration. The analysis is restricted to survey data between 2000 and 2018 since the data on migration is incomplete before 2000. We refer to Pew Research Center's analysis of current U.S. population and define three annual income classes—low class, middle class, and high class, which accounts for 30 percent, 50 percent, and 20 percent of the working class, respectively.

We first estimate how migration status varies by these income classes. Specifically, for each individual in the data, denoting by M_{ict} the migration status, where M_{ict} is either the indicator for whether individual i in the income class c has migrated or not in year t or the number of migrations, we consider the following specification:

$$M_{ict} = \sum_t \sum_c \alpha_{ct} \cdot \mathbb{I}[c = C] \cdot \mathbb{I}[t = T] + \epsilon_{ict},$$

where we interact the income class dummies with year dummies. Therefore, the estimated year-specific coefficients $\hat{\alpha}_{ct}$ correspond to the share or the average number of migrations for income class c in year t .

```
move_ind = subset(df, is.na(df$moved)==FALSE & is.na(df$inc)==FALSE &
                  df$inc!=0)
move_ind1 = subset(move_ind, move_ind$inccat==1)
move_ind2 = subset(move_ind, move_ind$inccat==2)
move_ind3 = subset(move_ind, move_ind$inccat==3)
move_num = subset(df, is.na(df$moves)==FALSE & is.na(df$inc)==FALSE &
                  df$inc!=0)
move_num1 = subset(move_num, move_num$inccat==1)
move_num2 = subset(move_num, move_num$inccat==2)
move_num3 = subset(move_num, move_num$inccat==3)

# model 1 corresponds to low income class
# model 2 corresponds to middle income class
# model 3 corresponds to high income class

modell1a = glm(moved ~ 0 + year_2000 + year_2002 + year_2004 + year_2006 +
              year_2008 + year_2010 + year_2012 + year_2014 + year_2016 +
              year_2018,
              data = move_ind1)
modell1b = lm(moves ~ 0 + year_2000 + year_2002 + year_2004 + year_2006 +
              year_2008 + year_2010 + year_2012 + year_2014 + year_2016 +
```

```

        year_2018,
        data = move_num1)
model2a = glm(moved ~ 0 + year_2000 + year_2002 + year_2004 + year_2006 +
        year_2008 + year_2010 + year_2012 + year_2014 + year_2016 +
        year_2018,
        data = move_ind2)
model2b = lm(moves ~ 0 + year_2000 + year_2002 + year_2004 + year_2006 +
        year_2008 + year_2010 + year_2012 + year_2014 + year_2016 +
        year_2018,
        data = move_num2)
model3a = glm(moved ~ 0 + year_2000 + year_2002 + year_2004 + year_2006 +
        year_2008 + year_2010 + year_2012 + year_2014 + year_2016 +
        year_2018,
        data = move_ind3)
model3b = lm(moves ~ 0 + year_2000 + year_2002 + year_2004 + year_2006 +
        year_2008 + year_2010 + year_2012 + year_2014 + year_2016 +
        year_2018,
        data = move_num3)

coef1a = summary(model1a)$coefficient[,1]
lwr1a = confint(model1a)[,1]

## Waiting for profiling to be done...
uwr1a = confint(model1a)[,2]

## Waiting for profiling to be done...
coef1b = summary(model1b)$coefficient[,1]
lwr1b = confint(model1b)[,1]
uwr1b = confint(model1b)[,2]
coef2a = summary(model2a)$coefficient[,1]
lwr2a = confint(model2a)[,1]

## Waiting for profiling to be done...
uwr2a = confint(model2a)[,2]

## Waiting for profiling to be done...
coef2b = summary(model2b)$coefficient[,1]
lwr2b = confint(model2b)[,1]
uwr2b = confint(model2b)[,2]
coef3a = summary(model3a)$coefficient[,1]
lwr3a = confint(model3a)[,1]

## Waiting for profiling to be done...
uwr3a = confint(model3a)[,2]

## Waiting for profiling to be done...
coef3b = summary(model3b)$coefficient[,1]
lwr3b = confint(model3b)[,1]
uwr3b = confint(model3b)[,2]
years = seq(2000,2018,by=2)

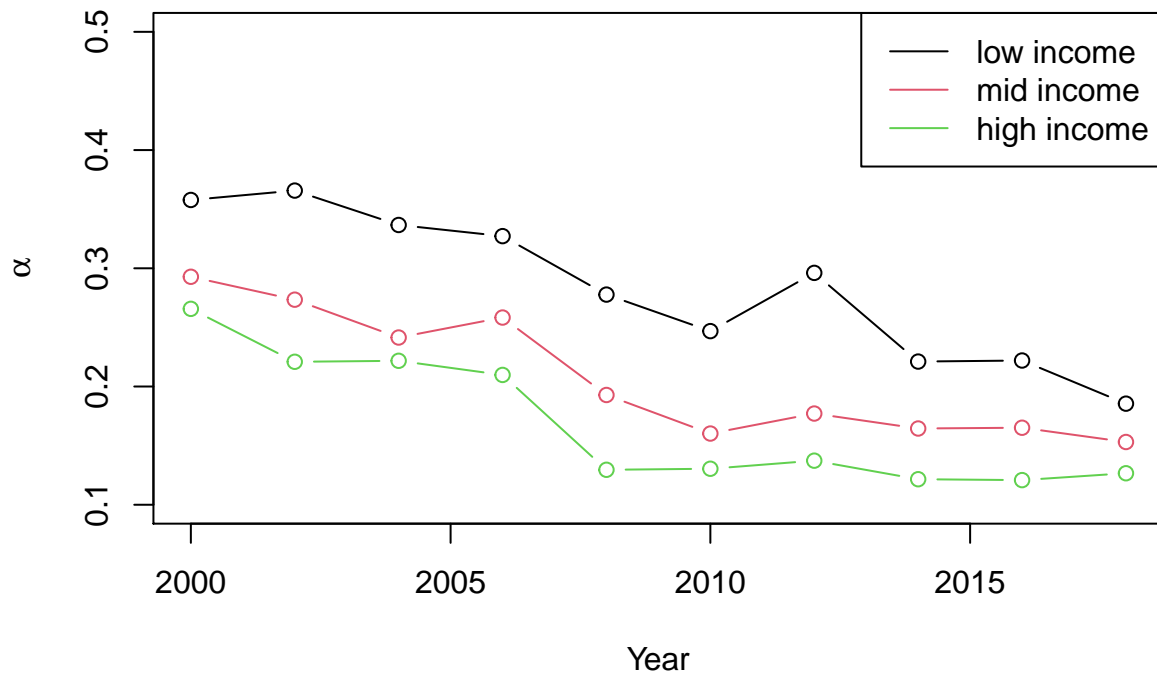
plot(years,coef1a,ylim=c(0.1,0.5),col=1,type='b',xlab='Year',
      ylab=bquote(alpha),lwd=1)

```

```

lines(years,coef2a,col=2,type='b',lwd=1)
lines(years,coef3a,col=3,type='b',lwd=1)
legend(x='topright',legend=c('low income','mid income','high income'),
      lty = c(1,1,1),col = c(1,2,3),cex=1,lwd=1)

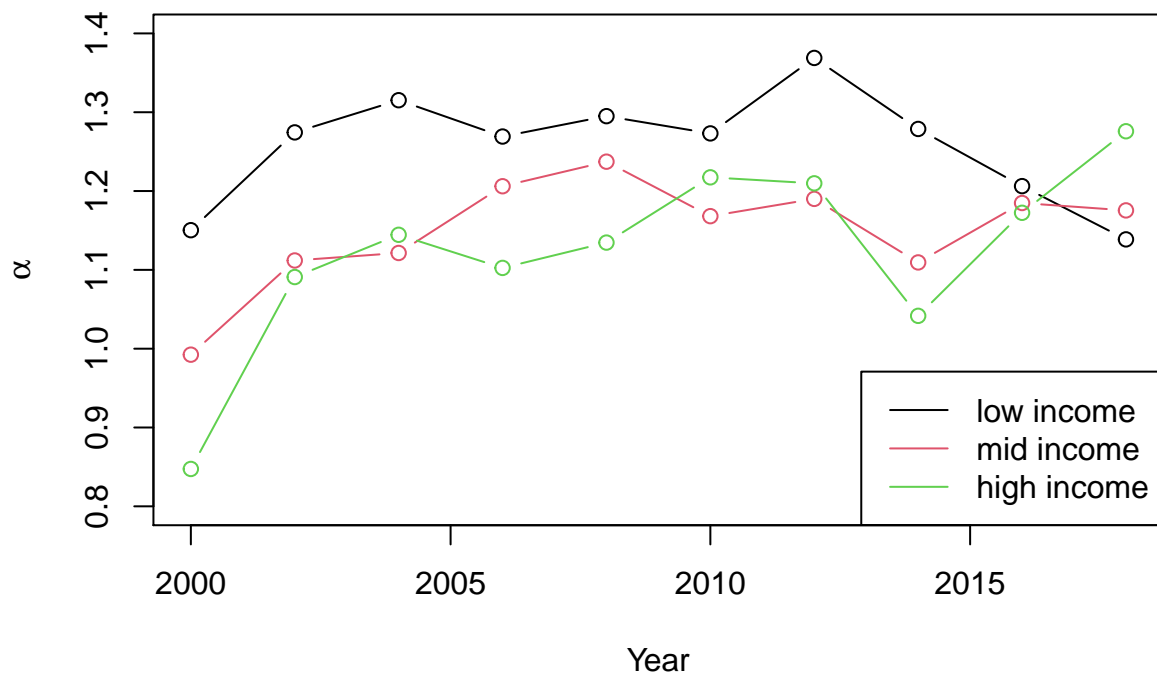
```



```

plot(years,coef1b,ylim=c(0.8,1.4),col=1,type='b',xlab='Year',
      ylab=bquote(alpha),lwd=1)
lines(years,coef2b,col=2,type='b',lwd=1)
lines(years,coef3b,col=3,type='b',lwd=1)
legend(x='bottomright',legend=c('low income','mid income','high income'),
      lty = c(1,1,1),col = c(1,2,3),cex=1,lwd=1)

```



The results are presented in Figure 1. Panel (a) plots the estimated α corresponding to the share of migrations, and we see declines in all three classes, with the low income class having the largest share and the high income class having the smallest share across the 18 year span. In each income class, the percentage of migrating population has been reduced by almost 50 percent from 2000 to 2018.

Panel (b) plots the estimated α corresponding to the average number of migrations. We see that, on average, people in the high income class have experienced an increase in the number of migrations of over 50 percent from 2000 to 2018, which far exceeds what people in the other two classes have experienced. While the numbers of migrations decreased with income class in 2000, in the long run, these numbers among the three income classes have been converging and ended up in an opposite way.

We next estimate how people in different income classes response to income change. Similar to the specification in equation (1), we run the following regression:

$$M_{ict} = \sum_t \sum_c \beta_{ct} \cdot inc_{ict} \cdot \mathbb{I}[c = C] \cdot \mathbb{I}[t = T] + \epsilon_{ict},$$

where we interact the income class dummies with year dummies. Therefore, the estimated year-specific coefficients $\hat{\beta}_{ct}$ correspond to the marginal effects of income on the probability or the number of migrations for income class c in year t .

```
model1c = glm(moved ~ 0 + inc2000 + inc2002 + inc2004 + inc2006 + inc2008 +
              inc2010 + inc2012 + inc2014 + inc2016 + inc2018,
              data = move_ind1)
model1d = lm(moves ~ 0 + inc2000 + inc2002 + inc2004 + inc2006 + inc2008 +
              inc2010 + inc2012 + inc2014 + inc2016 + inc2018,
              data = move_num1)
model2c = glm(moved ~ 0 + inc2000 + inc2002 + inc2004 + inc2006 + inc2008 +
              inc2010 + inc2012 + inc2014 + inc2016 + inc2018,
              data = move_ind2)
model2d = lm(moves ~ 0 + inc2000 + inc2002 + inc2004 + inc2006 + inc2008 +
              inc2010 + inc2012 + inc2014 + inc2016 + inc2018,
              data = move_num2)
model3c = glm(moved ~ 0 + inc2000 + inc2002 + inc2004 + inc2006 + inc2008 +
              inc2010 + inc2012 + inc2014 + inc2016 + inc2018,
              data = move_ind3)
model3d = lm(moves ~ 0 + inc2000 + inc2002 + inc2004 + inc2006 + inc2008 +
              inc2010 + inc2012 + inc2014 + inc2016 + inc2018,
              data = move_num3)
```

```
coef1c = summary(model1c)$coefficient[,1]
lwr1c = confint(model1c)[,1]
```

Waiting for profiling to be done...

```
uwr1c = confint(model1c)[,2]
```

Waiting for profiling to be done...

```
coef1d = summary(model1d)$coefficient[,1]
lwr1d = confint(model1d)[,1]
uwr1d = confint(model1d)[,2]
coef2c = summary(model2c)$coefficient[,1]
lwr2c = confint(model2c)[,1]
```

Waiting for profiling to be done...

```
uwr2c = confint(model2c)[,2]
```

```
## Waiting for profiling to be done...
```

```
coef2d = summary(model2d)$coefficient[,1]
lwr2d = confint(model2d)[,1]
uwr2d = confint(model2d)[,2]
coef3c = summary(model3c)$coefficient[,1]
lwr3c = confint(model3c)[,1]
```

```
## Waiting for profiling to be done...
```

```
uwr3c = confint(model3c)[,2]
```

```
## Waiting for profiling to be done...
```

```
coef3d = summary(model3d)$coefficient[,1]
lwr3d = confint(model3d)[,1]
uwr3d = confint(model3d)[,2]
```

```
plot(years,coef1c,ylim=c(0,0.03),col=1,type='b',xlab='Year',
      ylab=bquote(beta),lwd=1)
polygon(c(rev(years),years),c(rev(lwr1c),uwr1c),
        col=adjustcolor(1,alpha.f=0.1), border=NA)
lines(years,coef2c,ylim=c(0,0.03),col=2,type='b',lwd=1)
polygon(c(rev(years),years),c(rev(lwr2c),uwr2c),
        col=adjustcolor(2,alpha.f=0.1), border=NA)
lines(years,coef3c,ylim=c(0,0.03),col=3,type='b',lwd=1)
polygon(c(rev(years),years),c(rev(lwr3c),uwr3c),
        col=adjustcolor(3,alpha.f=0.1), border=NA)
legend(x='topright',legend=c('low income','mid income','high income'),
      lty = c(1,1,1),col = c(1,2,3),cex=1,lwd=1)
```



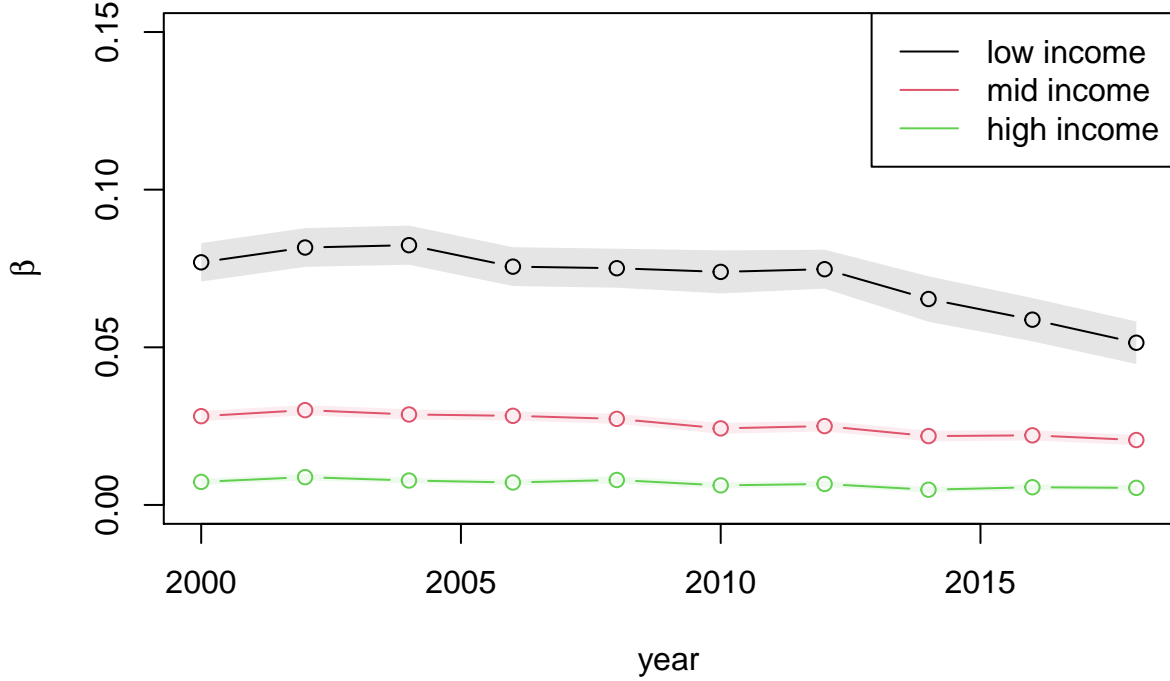
```
plot(years,coef1d,ylim=c(0,0.15),col=1,type='b',xlab='year',
      ylab=bquote(beta),lwd=1)
polygon(c(rev(years),years),c(rev(lwr1d),uwr1d),
        col=adjustcolor(1,alpha.f=0.1), border=NA)
```



```

lines(years,coef2d,ylim=c(0,0.15),col=2,type='b',lwd=1)
polygon(c(rev(years),years),c(rev(lwr2d),uwr2d),
        col=adjustcolor(2,alpha.f=0.1), border=NA)
lines(years,coef3d,ylim=c(0,0.15),col=3,type='b',lwd=1)
polygon(c(rev(years),years),c(rev(lwr3d),uwr3d),
        col=adjustcolor(3,alpha.f=0.1), border=NA)
legend(x='topright',legend=c('low income','mid income','high income'),
       lty = c(1,1,1),col = c(1,2,3),cex=1,lwd=1)

```



The results are presented in Figure 2. Panel (a) plots the estimated β for the probability of migration, and we see declines in all three classes, with the low income class having the largest marginal effects of income all the time. This suggests that for people in the low income class, income change has greater impacts on not only the frequency but also the likelihood of migration.

Panel (b) plots the estimated β for the number of migrations, with the shaded area being the 95 percent confidence interval. We see that overall, the number of migrations increases with income, as the estimated β is positive in each income class, and the marginal effects of income decrease with income class. In the long run, neither class experiences significant change in these effects.

Having estimated the impacts of income on different income classes, we define a migration surplus of lower income class relative to higher income class in year t as

$$S_t = \frac{(\hat{\alpha}_{lt} - \hat{\alpha}_{ht})}{E[\hat{M}_{ilt}|l, t]},$$

where l denotes low or middle income class, and h denotes middle or high income class. This migration surplus measures the percentage by which lower income class are leading ahead higher income class due to income effects.

Our next step is to decompose migration difference into what can be attributed to income and what can be attributed to other factors, showing how this composition has evolved over time. In particular, our decomposition approach focuses on income and the number of people in a family. With M_{ict} being the

number or indicator for migration, we consider the following specification:

$$M_{ict} = \sum_c \sum_t \alpha_{ct} \cdot \mathbb{I}[c = C] \cdot \mathbb{I}[t = T] + \sum_c \sum_t \gamma_{ct} \cdot famsize_{ict} \cdot \mathbb{I}[c = C] \cdot \mathbb{I}[t = T] + \epsilon_{ict},$$

where $famsize_{ict}$ denotes the number of people in the family of individual i who belongs to income class c in year t .

Defining the mean migration difference as $\Delta_t = \{E[M_{ilt}|l, t] - E[M_{iht}|h, t]\} / E[M_{iht}|h, t]$ and using specification (3), we can rearrange terms so as to obtain

$$\hat{\Delta}_t = \underbrace{\frac{E[S_t M_{ilt}|l, t]}{E[M_{iht}|h, t]}}_{\text{income-related migration difference}} + \underbrace{\frac{(\hat{\gamma}_{lt} - \hat{\gamma}_{ht}) E[famsize_{iht}|h, t]}{E[\hat{M}_{iht}|h, t]} + \frac{\hat{\gamma}_{lt} (E[famsize_{ilt}|l, t] - E[famsize_{iht}|h, t])}{E[\hat{M}_{iht}|h, t]}}_{\text{non-income migration difference}}.$$

```

model1e = glm(moved ~ 0 + year_2000 + year_2002 + year_2004 +
               year_2006 + year_2008 + year_2010 + year_2012 +
               year_2014 + year_2016 + year_2018 +
               famsize2000 + famsize2002 + famsize2004 + famsize2006 +
               famsize2008 + famsize2010 + famsize2012 + famsize2014 +
               famsize2016 + famsize2018,
               data = move_ind1)
model1f = lm(moves ~ 0 + year_2000 + year_2002 + year_2004 +
              year_2006 + year_2008 + year_2010 + year_2012 +
              year_2014 + year_2016 + year_2018 +
              famsize2000 + famsize2002 + famsize2004 + famsize2006 +
              famsize2008 + famsize2010 + famsize2012 + famsize2014 +
              famsize2016 + famsize2018,
              data = move_num1)
model2e = glm(moved ~ 0 + year_2000 + year_2002 + year_2004 +
               year_2006 + year_2008 + year_2010 + year_2012 +
               year_2014 + year_2016 + year_2018 +
               famsize2000 + famsize2002 + famsize2004 + famsize2006 +
               famsize2008 + famsize2010 + famsize2012 + famsize2014 +
               famsize2016 + famsize2018,
               data = move_ind2)
model2f = lm(moves ~ 0 + year_2000 + year_2002 + year_2004 +
              year_2006 + year_2008 + year_2010 + year_2012 +
              year_2014 + year_2016 + year_2018 +
              famsize2000 + famsize2002 + famsize2004 + famsize2006 +
              famsize2008 + famsize2010 + famsize2012 + famsize2014 +
              famsize2016 + famsize2018,
              data = move_num2)
model3e = glm(moved ~ 0 + year_2000 + year_2002 + year_2004 +
               year_2006 + year_2008 + year_2010 + year_2012 +
               year_2014 + year_2016 + year_2018 +
               famsize2000 + famsize2002 + famsize2004 + famsize2006 +
               famsize2008 + famsize2010 + famsize2012 + famsize2014 +
               famsize2016 + famsize2018,
               data = move_ind3)
model3f = lm(moves ~ 0 + year_2000 + year_2002 + year_2004 +
              year_2006 + year_2008 + year_2010 + year_2012 +
              year_2014 + year_2016 + year_2018 +
              famsize2000 + famsize2002 + famsize2004 + famsize2006 +
              famsize2008 + famsize2010 + famsize2012 + famsize2014 +
              famsize2016 + famsize2018,
              data = move_ind3)

```

```

famsize2016 + famsize2018,
data = move_num3)

inc_coef1e = summary(model1e)$coefficient[1:10,1]
fam_coef1e = summary(model1e)$coefficient[11:20,1]
inc_coef1f = summary(model1f)$coefficient[1:10,1]
fam_coef1f = summary(model1f)$coefficient[11:20,1]
inc_coef2e = summary(model2e)$coefficient[1:10,1]
fam_coef2e = summary(model2e)$coefficient[11:20,1]
inc_coef2f = summary(model2f)$coefficient[1:10,1]
fam_coef2f = summary(model2f)$coefficient[11:20,1]
inc_coef3e = summary(model3e)$coefficient[1:10,1]
fam_coef3e = summary(model3e)$coefficient[11:20,1]
inc_coef3f = summary(model3f)$coefficient[1:10,1]
fam_coef3f = summary(model3f)$coefficient[11:20,1]

move_ind_nobs1 = rep(0,10)
move_ind_nobs2 = rep(0,10)
move_ind_nobs3 = rep(0,10)
move_num_nobs1 = rep(0,10)
move_num_nobs2 = rep(0,10)
move_num_nobs3 = rep(0,10)
for (i in 0:9){
  move_ind_nobs1[i+1]=
  nrow(move_ind1[
    move_ind1[,paste('year',toString(2000+2*i),sep='_')==1,])
  move_ind_nobs2[i+1]=
  nrow(move_ind2[
    move_ind2[,paste('year',toString(2000+2*i),sep='_')==1,])
  move_ind_nobs3[i+1]=
  nrow(move_ind3[
    move_ind3[,paste('year',toString(2000+2*i),sep='_')==1,])
  move_num_nobs1[i+1]=
  nrow(move_num1[
    move_num1[,paste('year',toString(2000+2*i),sep='_')==1,])
  move_num_nobs2[i+1]=
  nrow(move_num2[
    move_num2[,paste('year',toString(2000+2*i),sep='_')==1,])
  move_num_nobs3[i+1]=
  nrow(move_num3[
    move_num3[,paste('year',toString(2000+2*i),sep='_')==1,])
  }

inc_diff12e = rep(0,10)
fam_diff12e = rep(0,10)
inc_diff12f = rep(0,10)
fam_diff12f = rep(0,10)
for (i in 0:9){
  inc_diff12e[i+1]=(inc_coef1e[i+1]-inc_coef2e[i+1])/
    (sum(move_ind2[,paste('moved',toString(2000+2*i),sep='')])/
      move_ind_nobs2[i+1])
}
for (i in 0:9){
  fam_diff12e[i+1]=(fam_coef1e[i+1]-fam_coef2e[i+1])*

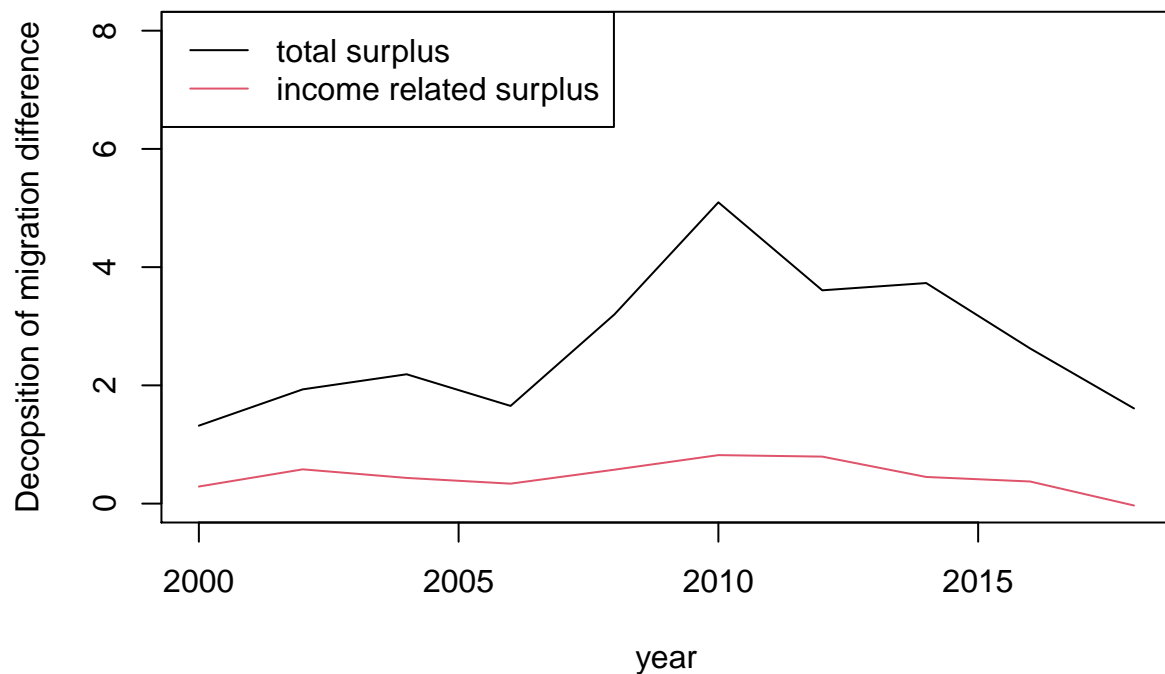
```

```

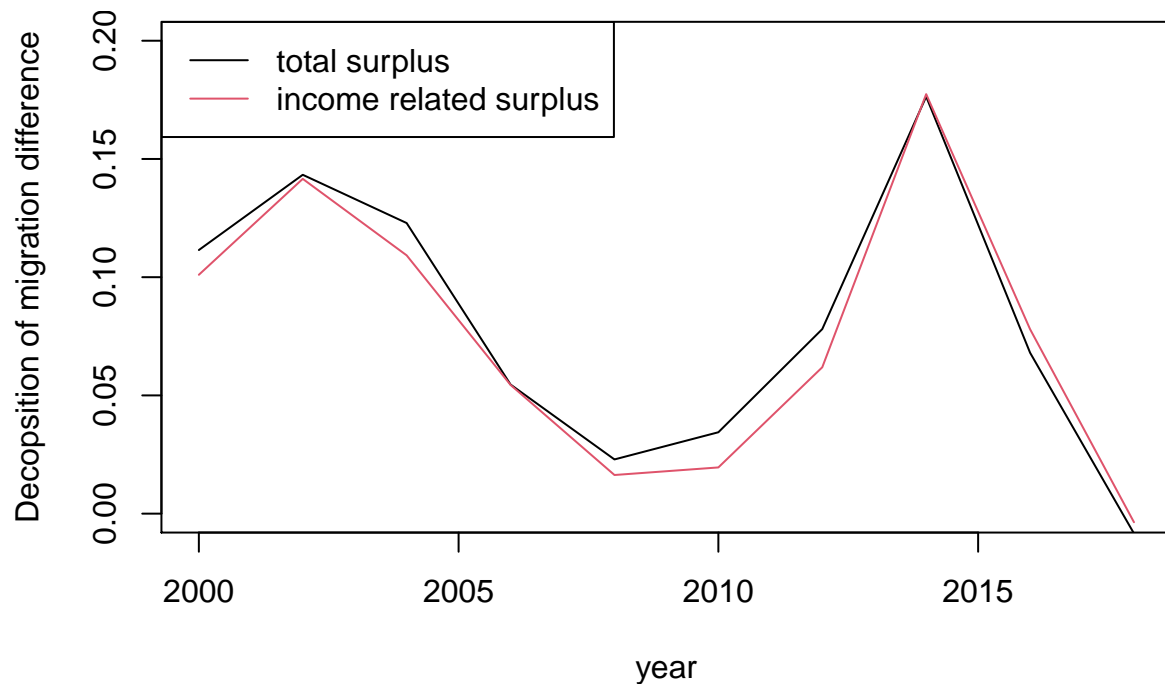
    (sum(move_ind2[,paste('famsize',toString(2000+2*i),sep='')])/
      move_ind_nobs2[i+1])/
    (sum(move_ind2[,paste('moved',toString(2000+2*i),sep='')])/
      move_ind_nobs2[i+1])+
    fam_coef1e[i+1]*
    (sum(move_ind1[,paste('famsize',toString(2000+2*i),sep='')])/
      move_ind_nobs1[i+1]-
      sum(move_ind2[,paste('famsize',toString(2000+2*i),sep='')])/
      move_num_nobs2[i+1])/
    (sum(move_ind2[,paste('moved',toString(2000+2*i),sep='')])/
      move_ind_nobs2[i+1])
  }
  for (i in 0:9){
    inc_diff12f[i+1]=(inc_coef1f[i+1]-inc_coef2f[i+1])/
      (sum(move_num2[,paste('moves',toString(2000+2*i),sep='')])/
        move_num_nobs2[i+1])
  }
  for (i in 0:9){
    fam_diff12f[i+1]=(fam_coef1f[i+1]-fam_coef2f[i+1])*
      (sum(move_num2[,paste('famsize',toString(2000+2*i),sep='')])/
        move_ind_nobs2[i+1])/
      (sum(move_num2[,paste('moves',toString(2000+2*i),sep='')])/
        move_num_nobs2[i+1])+
    fam_coef1f[i+1]*
    (sum(move_num1[,paste('famsize',toString(2000+2*i),sep='')])/
      move_num_nobs1[i+1]-
      sum(move_num2[,paste('famsize',toString(2000+2*i),sep='')])/
      move_num_nobs2[i+1])/
    (sum(move_num2[,paste('moves',toString(2000+2*i),sep='')])/
      move_num_nobs2[i+1])
  }
  diff12e = inc_diff12e + fam_diff12e
  diff12f = inc_diff12f + fam_diff12f

plot(years,diff12e,ylim=c(0,8),col=1,type='l',xlab='year',
      ylab='Decoposition of migration difference',lwd=1)
lines(years,inc_diff12e,col=2,type='l',lwd=1)
legend(x='topleft',legend=c('total surplus','income related surplus'),
      lty = c(1,1),col = c(1,2),cex=1,lwd=1)

```



```
plot(years,diff12f,ylim=c(0,0.2),col=1,type='l',xlab='year',
      ylab='Decoposition of migration difference',lwd=1)
lines(years,inc_diff12f,col=2,type='l',lwd=1)
legend(x='topleft',legend=c('total surplus','income related surplus'),
      lty = c(1,1),col = c(1,2),cex=1,lwd=1)
```



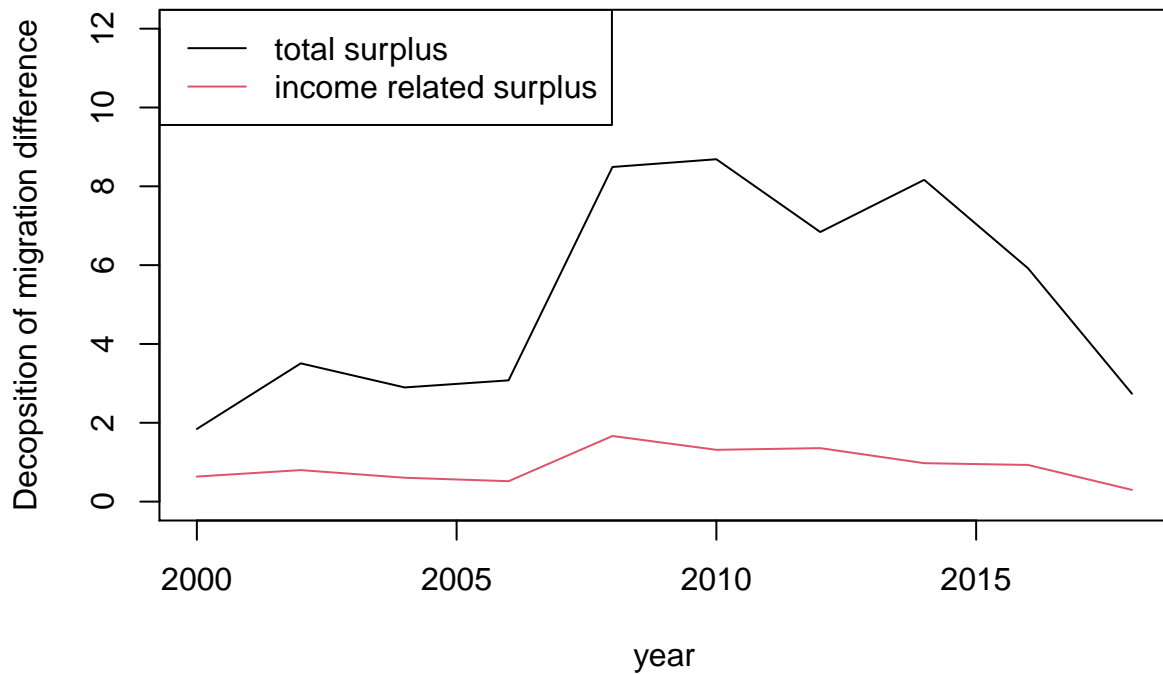
```
inc_diff13e = rep(0,10)
fam_diff13e = rep(0,10)
inc_diff13f = rep(0,10)
fam_diff13f = rep(0,10)
for (i in 0:9){
```

```

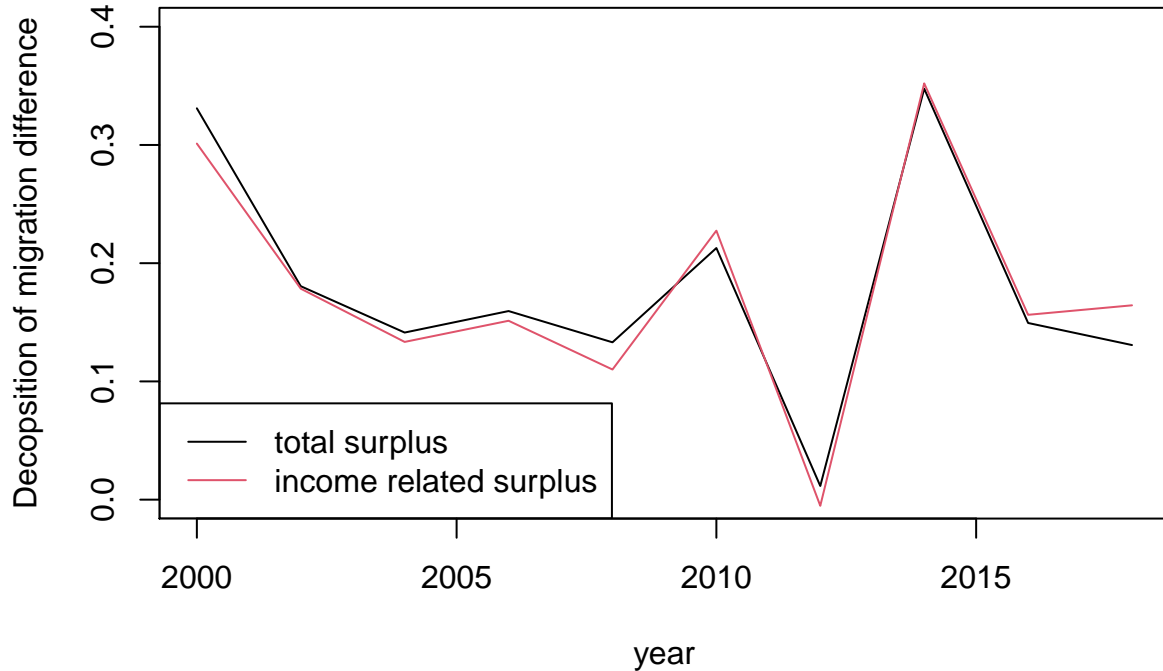
inc_diff13e[i+1]=(inc_coef1e[i+1]-inc_coef3e[i+1])/
  (sum(move_ind3[,paste('moved',toString(2000+2*i),sep='')])/
    move_ind_nobs3[i+1])
}
for (i in 0:9){
  fam_diff13e[i+1]=(fam_coef1e[i+1]-fam_coef3e[i+1])*
    (sum(move_ind3[,paste('famsize',toString(2000+2*i),sep='')])/
      move_ind_nobs3[i+1])/
    (sum(move_ind3[,paste('moved',toString(2000+2*i),sep='')])/
      move_ind_nobs3[i+1])+
  fam_coef1e[i+1]*
    (sum(move_ind1[,paste('famsize',toString(2000+2*i),sep='')])/
      move_ind_nobs1[i+1]-
      sum(move_ind3[,paste('famsize',toString(2000+2*i),sep='')])/
      move_num_nobs3[i+1])/
    (sum(move_ind3[,paste('moved',toString(2000+2*i),sep='')])/
      move_ind_nobs3[i+1])
}
for (i in 0:9){
  inc_diff13f[i+1]=(inc_coef1f[i+1]-inc_coef3f[i+1])/
    (sum(move_num3[,paste('moves',toString(2000+2*i),sep='')])/
      move_num_nobs3[i+1])
}
for (i in 0:9){
  fam_diff13f[i+1]=(fam_coef1f[i+1]-fam_coef3f[i+1])*
    (sum(move_num3[,paste('famsize',toString(2000+2*i),sep='')])/
      move_ind_nobs3[i+1])/
    (sum(move_num3[,paste('moves',toString(2000+2*i),sep='')])/
      move_num_nobs3[i+1])+
  fam_coef1f[i+1]*
    (sum(move_num1[,paste('famsize',toString(2000+2*i),sep='')])/
      move_num_nobs1[i+1]-
      sum(move_num3[,paste('famsize',toString(2000+2*i),sep='')])/
      move_num_nobs3[i+1])/
    (sum(move_num3[,paste('moves',toString(2000+2*i),sep='')])/
      move_num_nobs3[i+1])
}
diff13e = inc_diff13e + fam_diff13e
diff13f = inc_diff13f + fam_diff13f

plot(years,diff13e,ylim=c(0,12),col=1,type='l',xlab='year',
      ylab='Decoposition of migration difference',lwd=1)
lines(years,inc_diff13e,col=2,type='l',lwd=1)
legend(x='topleft',legend=c('total surplus','income related surplus'),
      lty = c(1,1),col = c(1,2),cex=1,lwd=1)

```



```
plot(years,diff13f,ylim=c(0,0.4),col=1,type='l',xlab='year',
      ylab='Decoposition of migration difference',lwd=1)
lines(years,inc_diff13f,col=2,type='l',lwd=1)
legend(x='bottomleft',legend=c('total surplus','income related surplus'),
      lty = c(1,1),col = c(1,2),cex=1,lwd=1)
```



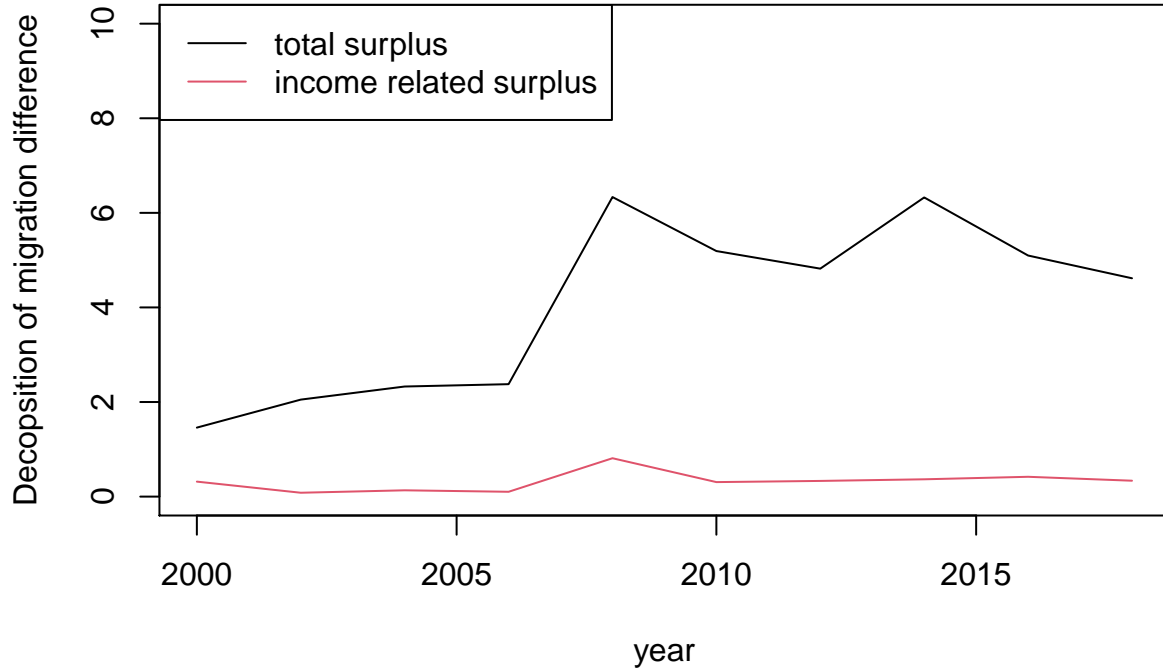
```
inc_diff23e = rep(0,10)
fam_diff23e = rep(0,10)
inc_diff23f = rep(0,10)
fam_diff23f = rep(0,10)
for (i in 0:9){
```

```

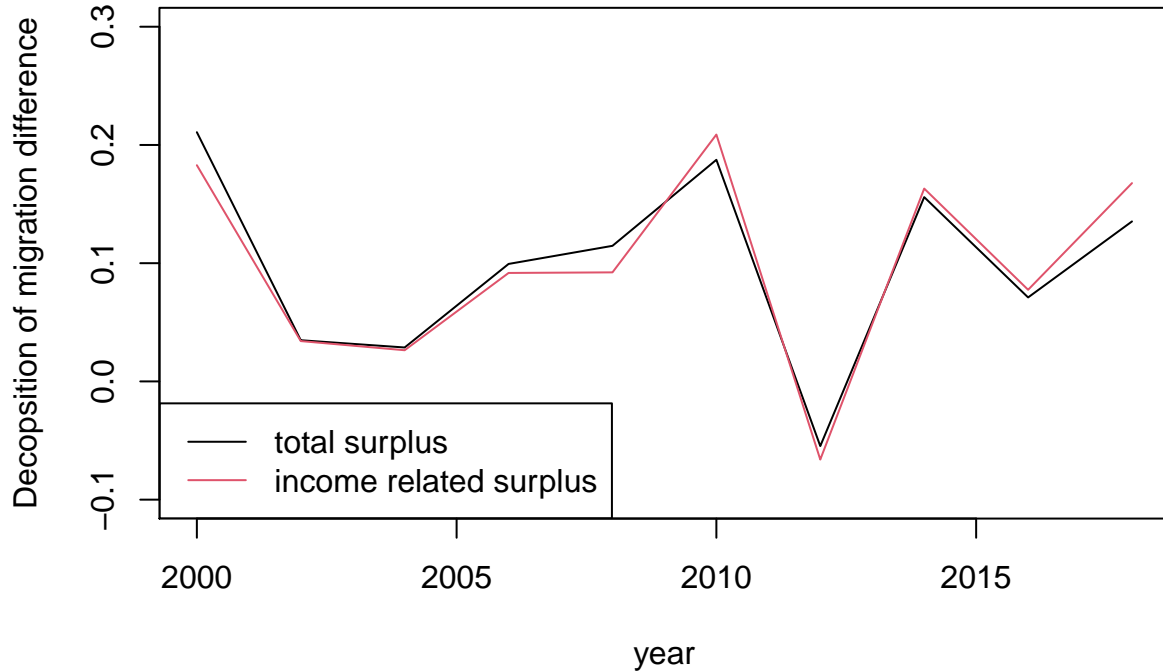
inc_diff23e[i+1]=(inc_coef2e[i+1]-inc_coef3e[i+1])/
  (sum(move_ind3[,paste('moved',toString(2000+2*i),sep='')])/
    move_ind_nobs3[i+1])
}
for (i in 0:9){
  fam_diff23e[i+1]=(fam_coef2e[i+1]-fam_coef3e[i+1])*
    (sum(move_ind3[,paste('famsize',toString(2000+2*i),sep='')])/
      move_ind_nobs3[i+1])/
    (sum(move_ind3[,paste('moved',toString(2000+2*i),sep='')])/
      move_ind_nobs3[i+1])+
    fam_coef2e[i+1]*
    (sum(move_ind2[,paste('famsize',toString(2000+2*i),sep='')])/
      move_ind_nobs2[i+1]-
      sum(move_ind3[,paste('famsize',toString(2000+2*i),sep='')])/
        move_num_nobs3[i+1])/
    (sum(move_ind3[,paste('moved',toString(2000+2*i),sep='')])/
      move_ind_nobs3[i+1])
}
for (i in 0:9){
  inc_diff23f[i+1]=(inc_coef2f[i+1]-inc_coef3f[i+1])/
    (sum(move_num3[,paste('moves',toString(2000+2*i),sep='')])/
      move_num_nobs3[i+1])
}
for (i in 0:9){
  fam_diff23f[i+1]=(fam_coef2f[i+1]-fam_coef3f[i+1])*
    (sum(move_num3[,paste('famsize',toString(2000+2*i),sep='')])/
      move_ind_nobs3[i+1])/
    (sum(move_num3[,paste('moves',toString(2000+2*i),sep='')])/
      move_num_nobs3[i+1])+
    fam_coef2f[i+1]*
    (sum(move_num2[,paste('famsize',toString(2000+2*i),sep='')])/
      move_num_nobs2[i+1]-
      sum(move_num3[,paste('famsize',toString(2000+2*i),sep='')])/
        move_num_nobs3[i+1])/
    (sum(move_num3[,paste('moves',toString(2000+2*i),sep='')])/
      move_num_nobs3[i+1])
}
diff23e = inc_diff23e + fam_diff23e
diff23f = inc_diff23f + fam_diff23f

plot(years,diff23e,ylim=c(0,10),col=1,type='l',xlab='year',
      ylab='Decopsition of migration difference',lwd=1)
lines(years,inc_diff23e,col=2,type='l',lwd=1)
legend(x='topleft',legend=c('total surplus','income related surplus'),
      lty = c(1,1),col = c(1,2),cex=1,lwd=1)

```

```
plot(years,diff23f,ylim=c(-0.1,0.3),col=1,type='l',xlab='year',
      ylab='Decomposition of migration difference',lwd=1)
lines(years,inc_diff23f,col=2,type='l',lwd=1)
legend(x='bottomleft',legend=c('total surplus','income related surplus'),
      lty = c(1,1),col = c(1,2),cex=1,lwd=1)
```



The estimated mean migration difference $\hat{\Delta}_t$ are presented in Figure 3. Panels (a), (c), and (e) plot the decomposition of the difference in migration's probability, while panels (b), (d), and (e) plot the decomposition of the difference in migration's number between low and middle income classes, low and high income classes, as well as middle and high income classes, respectively.

Three points are worth highlighting from our results. First, the trends in all three class pairs evolve in similar

fashions. Second, the total surplus in the migration's probability is much larger than the income-related surplus, meaning that the number of people in a family has much greater impacts on the surplus in migration's probability than income, and that people living in large families in higher income class are more unwilling to migrate than those in lower income class. Second, there is a sharp increase in the gaps between the two lines around 2006-2008 in panels (a), (c), and (e), indicating that either people in lower income class become less responsive to family size or people in higher income class become more responsive to family size, and these gaps gradually shrink after 2015. Third, there is almost no difference between the total surplus and the income-related surplus in the migration's number, and these surpluses dramatically increase around 2012-2015. While in panels (d) and (f), the surpluses of low income class relative to high income class and of middle income class relative to high income class experiences a big drop between 2010 and 2012, in panel (b), the surplus of low income class relative to middle income class experiences an increase in this period. This is an interesting find because consider the following property:

$$A \sim C \cap B \sim C \implies A \sim C.$$

Therefore, and the trends of low and high income classes and of middle and high income classes are similar, then theoretically the trend of low and middle income classes should not be too different. This inconsistency suggests the need for further work.

In this section, we have tried to explore how migration status and the marginal effects of income vary by income classes. We also decompose the migration difference into income-related difference and non-income difference. The finding is that the probability and the number of migration increase with income class, and the marginal effects of income are positive in each class, meaning that people becomes more likely to migrate and migrate more frequently as their income rises. Furthermore, we find that the number of people in a family has much larger impacts on the difference in migration's probability than income. However, due to limited amount of time, we have not explored the inconsistency between the differences in migration's number from 2010 to 2012. We hope that either we can continue to investigate this research limit or that future researchers could take this as their topic.

4 Income Effects by Race and Gender

Census 2000 Special Reports found that the likelihood that people move varies by race and Hispanic origin, and the redistribution of people by race and Hispanic origin is largely a function of domestic and international migration, which strongly influence the population growth and decline of particular racial and ethnic groups in particular locations. Given these proven differences, our goal is to further investigate whether the marginal effects of income on migration differ by race.

Our sample are divided into three race groups: Hispanic, Black, and non-Hispanic, non-Black. We start by estimating the year coefficients $\hat{\beta}_{rt}$ similar to the specification in equation (1) for women and men, where r denotes race.

```
move_ind_hisp = subset(move_ind, move_ind$race==1)
move_ind_blk = subset(move_ind, move_ind$race==2)
move_ind_non = subset(move_ind, move_ind$race==3)
move_num_hisp = subset(move_num, move_num$race==1)
move_num_blk = subset(move_num, move_num$race==2)
move_num_non = subset(move_num, move_num$race==3)

# model 4 corresponds to hispanic
# model 5 corresponds to black
# model 6 corresponds to non-hispanic, non-black

model4a = glm(moved ~ 0 + inc2000 + inc2002 + inc2004 + inc2006 + inc2008 +
              inc2010 + inc2012 + inc2014 + inc2016 + inc2018,
              data = move_ind_hisp)
```

```

model4b = lm(moves ~ 0 + inc2000 + inc2002 + inc2004 + inc2006 + inc2008 +
             inc2010 + inc2012 + inc2014 + inc2016 + inc2018,
             data = move_num_hisp)
model5a = glm(moved ~ 0 + inc2000 + inc2002 + inc2004 + inc2006 + inc2008 +
             inc2010 + inc2012 + inc2014 + inc2016 + inc2018,
             data = move_ind_blk)
model5b = lm(moves ~ 0 + inc2000 + inc2002 + inc2004 + inc2006 + inc2008 +
             inc2010 + inc2012 + inc2014 + inc2016 + inc2018,
             data = move_num_blk)
model6a = glm(moved ~ 0 + inc2000 + inc2002 + inc2004 + inc2006 + inc2008 +
             inc2010 + inc2012 + inc2014 + inc2016 + inc2018,
             data = move_ind_non)
model6b = lm(moves ~ 0 + inc2000 + inc2002 + inc2004 + inc2006 + inc2008 +
             inc2010 + inc2012 + inc2014 + inc2016 + inc2018,
             data = move_num_non)

coef4a = summary(model4a)$coefficient[,1]
lwr4a = confint(model4a)[,1]

## Waiting for profiling to be done...
uwr4a = confint(model4a)[,2]

## Waiting for profiling to be done...
coef4b = summary(model4b)$coefficient[,1]
lwr4b = confint(model4b)[,1]
uwr4b = confint(model4b)[,2]
coef5a = summary(model5a)$coefficient[,1]
lwr5a = confint(model5a)[,1]

## Waiting for profiling to be done...
uwr5a = confint(model5a)[,2]

## Waiting for profiling to be done...
coef5b = summary(model5b)$coefficient[,1]
lwr5b = confint(model5b)[,1]
uwr5b = confint(model5b)[,2]
coef6a = summary(model6a)$coefficient[,1]
lwr6a = confint(model6a)[,1]

## Waiting for profiling to be done...
uwr6a = confint(model6a)[,2]

## Waiting for profiling to be done...
coef6b = summary(model6b)$coefficient[,1]
lwr6b = confint(model6b)[,1]
uwr6b = confint(model6b)[,2]

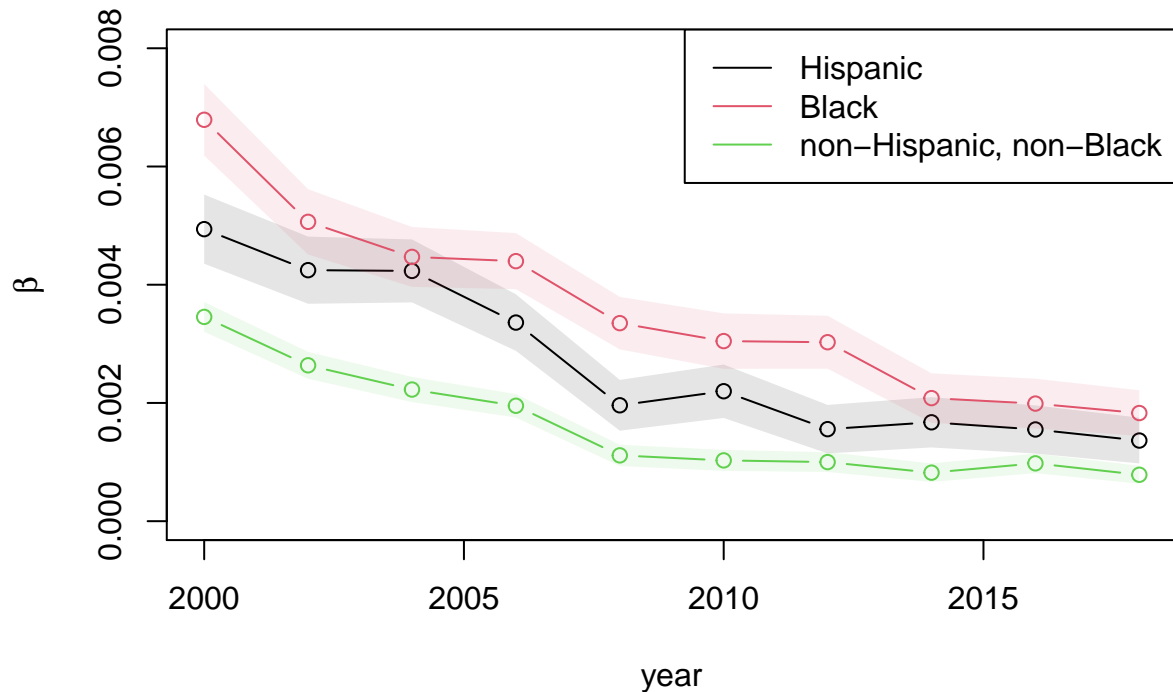
plot(years,coef4a,ylim=c(0,0.008),col=1,type='b',xlab='year',
     ylab=bquote(beta),lwd=1)
polygon(c(rev(years),years),c(rev(lwr4a),uwr4a),
       col=adjustcolor(1,alpha.f=0.1), border=NA)
lines(years,coef5a,col=2,type='b',lwd=1)
polygon(c(rev(years),years),c(rev(lwr5a),uwr5a),

```

```

col=adjustcolor(2,alpha.f=0.1), border=NA)
lines(years,coef6a,col=3,type='b',lwd=1)
polygon(c(rev(years),years),c(rev(lwr6a),uwr6a),
        col=adjustcolor(3,alpha.f=0.1), border=NA)
legend(x='topright',legend=c('Hispanic','Black','non-Hispanic, non-Black'),
       lty = c(1,1,1),col = c(1,2,3),cex=1,lwd=1)

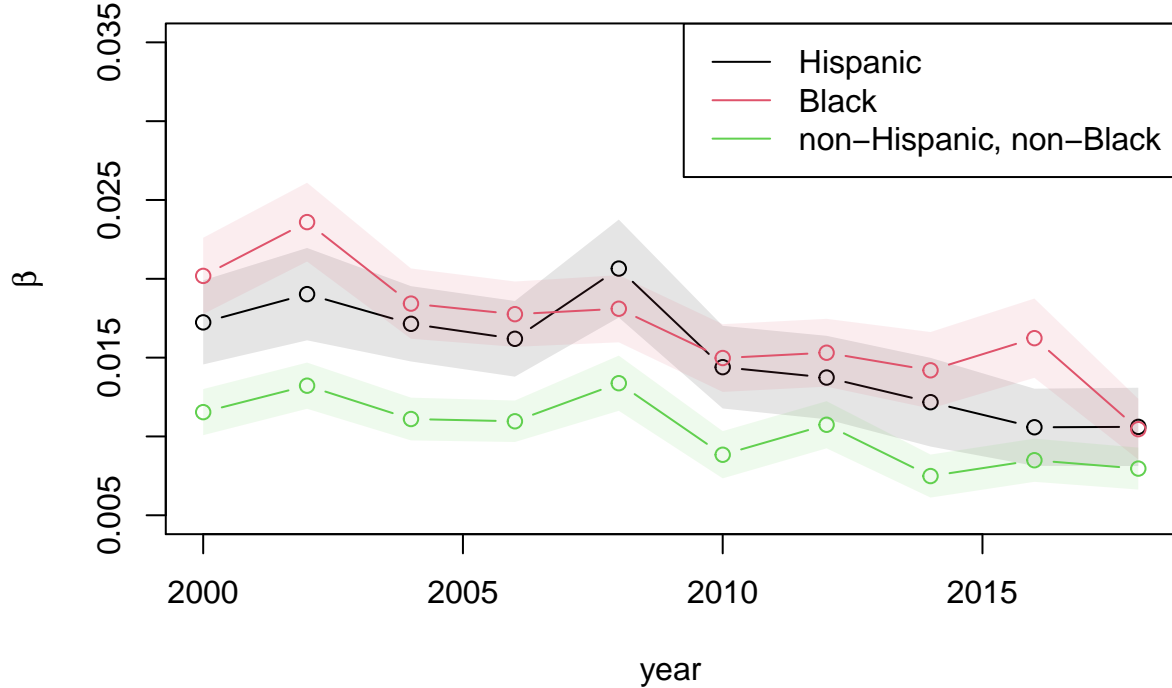
```



```

plot(years,coef4b,ylim=c(0.005,0.035),col=1,type='b',xlab='year',
      ylab=bquote(beta),lwd=1)
polygon(c(rev(years),years),c(rev(lwr4b),uwr4b),
        col=adjustcolor(1,alpha.f=0.1), border=NA)
lines(years,coef5b,col=2,type='b',lwd=1)
polygon(c(rev(years),years),c(rev(lwr5b),uwr5b),
        col=adjustcolor(2,alpha.f=0.1), border=NA)
lines(years,coef6b,col=3,type='b',lwd=1)
polygon(c(rev(years),years),c(rev(lwr6b),uwr6b),
        col=adjustcolor(3,alpha.f=0.1), border=NA)
legend(x='topright',legend=c('Hispanic','Black','non-Hispanic, non-Black'),
       lty = c(1,1,1),col = c(1,2,3),cex=1,lwd=1)

```



The results are presented in Figure 4. Panel (a) plots the estimated β for the probability of migrations, with the shaded area being the 95 percent confidence interval, and we see that the marginal effects of income are the largest among Hispanic and are the smallest among non-Hispanic, non-Black. While the marginal effects of income have been reduced by more than 50 percent from 2000 to 2018 for each race group, these effects among the three groups have been converging over time.

Panel (b) plots the estimated β for the number of migration, and we see that while the marginal effects of income are the smallest among non-Hispanic, non-Black, these effects have been fluctuating in degree among Hispanic and Black, and the general trend is downward for each race group.

We next consider whether migration status varies by gender. Early literature have found that for a range of labor market outcomes, women and men evolve in parallel until the birth of their first child, diverge sharply immediately after childbirth, and do not converge again (Kleven, Landais, and Sogaard* 2019). In particular, just after the birth of the first child, women start falling behind men in terms of their earnings and wage rate. Given these proven differences, our goal is to further investigate whether the marginal effects of income differ by gender.

We start by estimating the year coefficients $\hat{\beta}_{gt}$ similar to the specification in equation (1) for women and men, where g denotes gender.

```
move_ind_m = subset(move_ind, move_ind$sex==1)
move_ind_w = subset(move_ind, move_ind$sex==2)
move_num_m = subset(move_num, move_num$sex==1)
move_num_w = subset(move_num, move_num$sex==2)

# model 7 corresponds to male
# model 8 corresponds to female

model7a = glm(moved ~ 0 + inc2000 + inc2002 + inc2004 + inc2006 + inc2008 +
              inc2010 + inc2012 + inc2014 + inc2016 + inc2018,
              data = move_ind_m)
model7b = lm(moves ~ 0 + inc2000 + inc2002 + inc2004 + inc2006 + inc2008 +
             inc2010 + inc2012 + inc2014 + inc2016 + inc2018,
             data = move_num_m)
```

```

model8a = glm(moved ~ 0 + inc2000 + inc2002 + inc2004 + inc2006 + inc2008 +
              inc2010 + inc2012 + inc2014 + inc2016 + inc2018,
              data = move_ind_w)
model8b = lm(moves ~ 0 + inc2000 + inc2002 + inc2004 + inc2006 + inc2008 +
             inc2010 + inc2012 + inc2014 + inc2016 + inc2018,
             data = move_num_w)

coef7a = summary(model7a)$coefficient[,1]
lwr7a = confint(model7a)[,1]

## Waiting for profiling to be done...
uwr7a = confint(model7a)[,2]

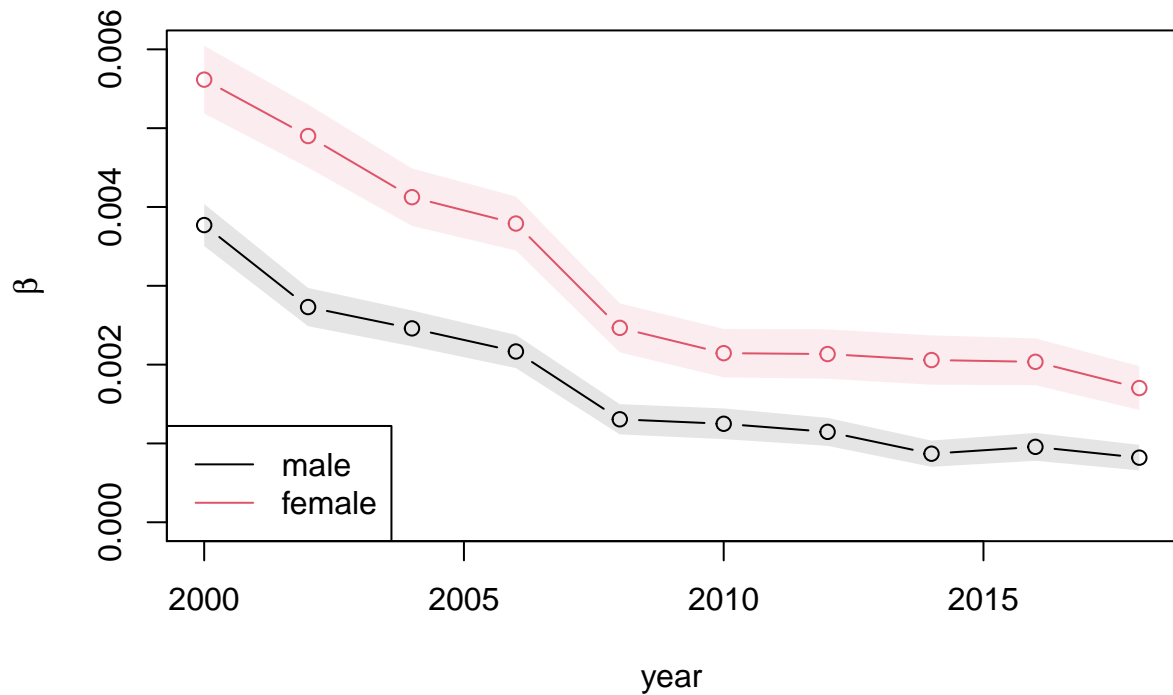
## Waiting for profiling to be done...
coef7b = summary(model7b)$coefficient[,1]
lwr7b = confint(model7b)[,1]
uwr7b = confint(model7b)[,2]
coef8a = summary(model8a)$coefficient[,1]
lwr8a = confint(model8a)[,1]

## Waiting for profiling to be done...
uwr8a = confint(model8a)[,2]

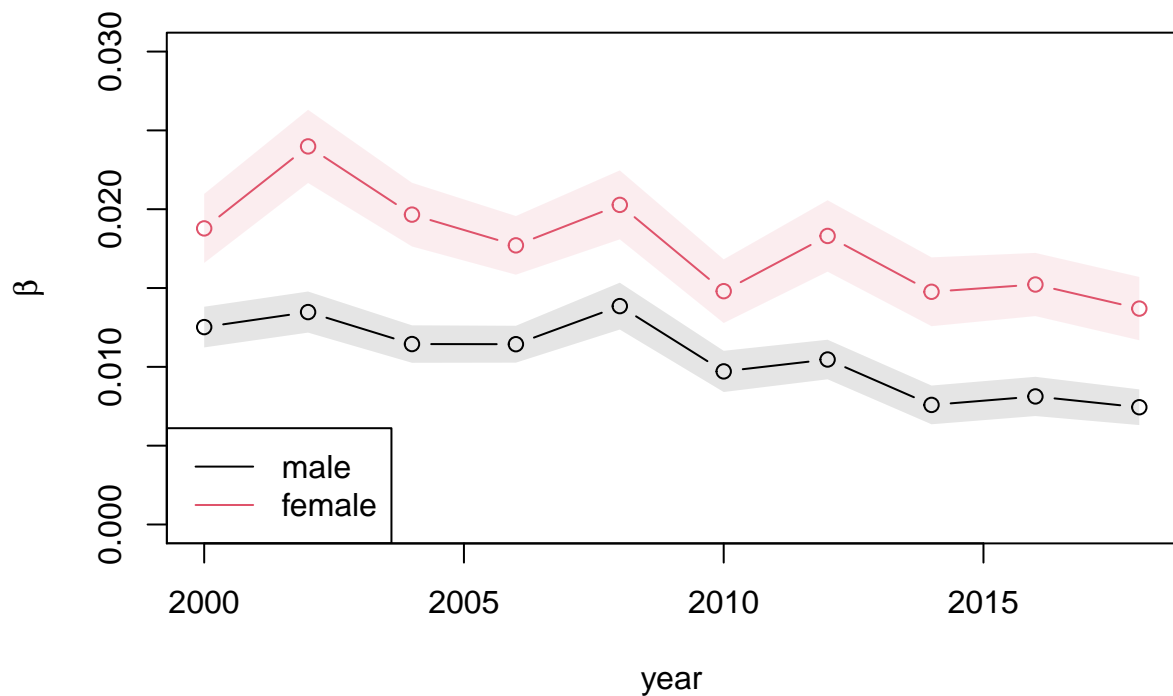
## Waiting for profiling to be done...
coef8b = summary(model8b)$coefficient[,1]
lwr8b = confint(model8b)[,1]
uwr8b = confint(model8b)[,2]

plot(years,coef7a,ylim=c(0,0.006),col=1,type='b',xlab='year',
      ylab=bquote(beta),lwd=1)
polygon(c(rev(years),years),c(rev(lwr7a),uwr7a),
        col=adjustcolor(1,alpha.f=0.1), border=NA)
lines(years,coef8a,col=2,type='b',lwd=1)
polygon(c(rev(years),years),c(rev(lwr8a),uwr8a),
        col=adjustcolor(2,alpha.f=0.1), border=NA)
legend(x='bottomleft',legend=c('male','female'),
       lty = c(1,1),col = c(1,2),cex=1,lwd=1)

```



```
plot(years,coef7b,ylim=c(0,0.03),col=1,type='b',xlab='year',
      ylab=bquote(beta),lwd=1)
polygon(c(rev(years),years),c(rev(lwr7b),uwr7b),
        col=adjustcolor(1,alpha.f=0.1), border=NA)
lines(years,coef8b,col=2,type='b',lwd=1)
polygon(c(rev(years),years),c(rev(lwr8b),uwr8b),
        col=adjustcolor(2,alpha.f=0.1), border=NA)
legend(x='bottomleft',legend=c('male','female'),
       lty = c(1,1),col = c(1,2),cex=1,lwd=1)
```



The results are presented in Figure 4. Panel (a) plots the estimated β for the probability of migrations, with the shaded area being the 95 percent confidence interval, and we see that the marginal effects of income are significant larger for women than for men, while these effects for men and women evolve in an almost parallel fashion, with no significant change in the long run.

Panel (b) plots the estimated β for the number of migration, and we see that the marginal effects of income are significant larger for women than for men, while these effects for men and women evolve in an almost parallel fashion and decline over time, with the higher decline rate in the early 2000s.

We now consider whether gender difference varies by race or race difference varies by gender and run the following regression:

$$M_{irgt} = \sum_r \sum_g \sum_t \beta_{rgt} \cdot inc_{irgt} \cdot \mathbb{I}[r = R] \cdot \mathbb{I}[g = G] \cdot \mathbb{I}[t = T] + \epsilon_{irgt},$$

where $\mathbb{I}[r = R]$ is an indicator for race being R , and $\mathbb{I}[g = G]$ is an indicator for gender being G .

```
move_ind_hisp_m = subset(move_ind, move_ind$race==1 &
                          move_ind$sex==1)
move_ind_hisp_w = subset(move_ind, move_ind$race==1 &
                          move_ind$sex==2)
move_ind_blk_m = subset(move_ind, move_ind$race==2 &
                          move_ind$sex==1)
move_ind_blk_w = subset(move_ind, move_ind$race==2 &
                          move_ind$sex==2)
move_ind_non_m = subset(move_ind, move_ind$race==3 &
                          move_ind$sex==1)
move_ind_non_w = subset(move_ind, move_ind$race==3 &
                          move_ind$sex==2)
move_num_hisp_m = subset(move_num, move_num$race==1 &
                          move_num$sex==1)
move_num_hisp_w = subset(move_num, move_num$race==1 &
                          move_num$sex==2)
move_num_blk_m = subset(move_num, move_num$race==2 &
                          move_num$sex==1)
move_num_blk_w = subset(move_num, move_num$race==2 &
                          move_num$sex==2)
move_num_non_m = subset(move_num, move_num$race==3 &
                          move_num$sex==1)
move_num_non_w = subset(move_num, move_num$race==3 &
                          move_num$sex==2)

# model 47 corresponds to hispanic male
# model 48 corresponds to hispanic female
# model 57 corresponds to black male
# model 58 corresponds to black female
# model 67 corresponds to non-hispanic, non-black male
# model 68 corresponds to non-hispanic, non-black female

model47a = glm(moved ~ 0 + inc2000 + inc2002 + inc2004 + inc2006 + inc2008 +
               inc2010 + inc2012 + inc2014 + inc2016 + inc2018,
               data = move_ind_hisp_m)
model47b = lm(moves ~ 0 + inc2000 + inc2002 + inc2004 + inc2006 + inc2008 +
               inc2010 + inc2012 + inc2014 + inc2016 + inc2018,
               data = move_num_hisp_m)
model48a = glm(moved ~ 0 + inc2000 + inc2002 + inc2004 + inc2006 + inc2008 +
               inc2010 + inc2012 + inc2014 + inc2016 + inc2018,
```



```

      data = move_ind_hisp_w)
model48b = lm(moves ~ 0 + inc2000 + inc2002 + inc2004 + inc2006 + inc2008 +
      inc2010 + inc2012 + inc2014 + inc2016 + inc2018,
      data = move_num_hisp_w)
model57a = glm(moved ~ 0 + inc2000 + inc2002 + inc2004 + inc2006 + inc2008 +
      inc2010 + inc2012 + inc2014 + inc2016 + inc2018,
      data = move_ind_blk_m)
model57b = lm(moves ~ 0 + inc2000 + inc2002 + inc2004 + inc2006 + inc2008 +
      inc2010 + inc2012 + inc2014 + inc2016 + inc2018,
      data = move_num_blk_m)
model58a = glm(moved ~ 0 + inc2000 + inc2002 + inc2004 + inc2006 + inc2008 +
      inc2010 + inc2012 + inc2014 + inc2016 + inc2018,
      data = move_ind_blk_w)
model58b = lm(moves ~ 0 + inc2000 + inc2002 + inc2004 + inc2006 + inc2008 +
      inc2010 + inc2012 + inc2014 + inc2016 + inc2018,
      data = move_num_blk_w)
model67a = glm(moved ~ 0 + inc2000 + inc2002 + inc2004 + inc2006 + inc2008 +
      inc2010 + inc2012 + inc2014 + inc2016 + inc2018,
      data = move_ind_non_m)
model67b = lm(moves ~ 0 + inc2000 + inc2002 + inc2004 + inc2006 + inc2008 +
      inc2010 + inc2012 + inc2014 + inc2016 + inc2018,
      data = move_num_non_m)
model68a = glm(moved ~ 0 + inc2000 + inc2002 + inc2004 + inc2006 + inc2008 +
      inc2010 + inc2012 + inc2014 + inc2016 + inc2018,
      data = move_ind_non_w)
model68b = lm(moves ~ 0 + inc2000 + inc2002 + inc2004 + inc2006 + inc2008 +
      inc2010 + inc2012 + inc2014 + inc2016 + inc2018,
      data = move_num_non_w)

```

```

coef47a = summary(model47a)$coefficient[,1]
lwr47a = confint(model47a)[,1]

```

Waiting for profiling to be done...

```

uwr47a = confint(model47a)[,2]

```

Waiting for profiling to be done...

```

coef47b = summary(model47b)$coefficient[,1]
lwr47b = confint(model47b)[,1]
uwr47b = confint(model47b)[,2]
coef48a = summary(model48a)$coefficient[,1]
lwr48a = confint(model48a)[,1]

```

Waiting for profiling to be done...

```

uwr48a = confint(model48a)[,2]

```

Waiting for profiling to be done...

```

coef48b = summary(model48b)$coefficient[,1]
lwr48b = confint(model48b)[,1]
uwr48b = confint(model48b)[,2]
coef57a = summary(model57a)$coefficient[,1]
lwr57a = confint(model57a)[,1]

```

Waiting for profiling to be done...

```

uwr57a = confint(model57a)[,2]

## Waiting for profiling to be done...
coef57b = summary(model57b)$coefficient[,1]
lwr57b = confint(model57b)[,1]
uwr57b = confint(model57b)[,2]
coef58a = summary(model58a)$coefficient[,1]
lwr58a = confint(model58a)[,1]

## Waiting for profiling to be done...
uwr58a = confint(model58a)[,2]

## Waiting for profiling to be done...
coef58b = summary(model58b)$coefficient[,1]
lwr58b = confint(model58b)[,1]
uwr58b = confint(model58b)[,2]
coef67a = summary(model67a)$coefficient[,1]
lwr67a = confint(model67a)[,1]

## Waiting for profiling to be done...
uwr67a = confint(model67a)[,2]

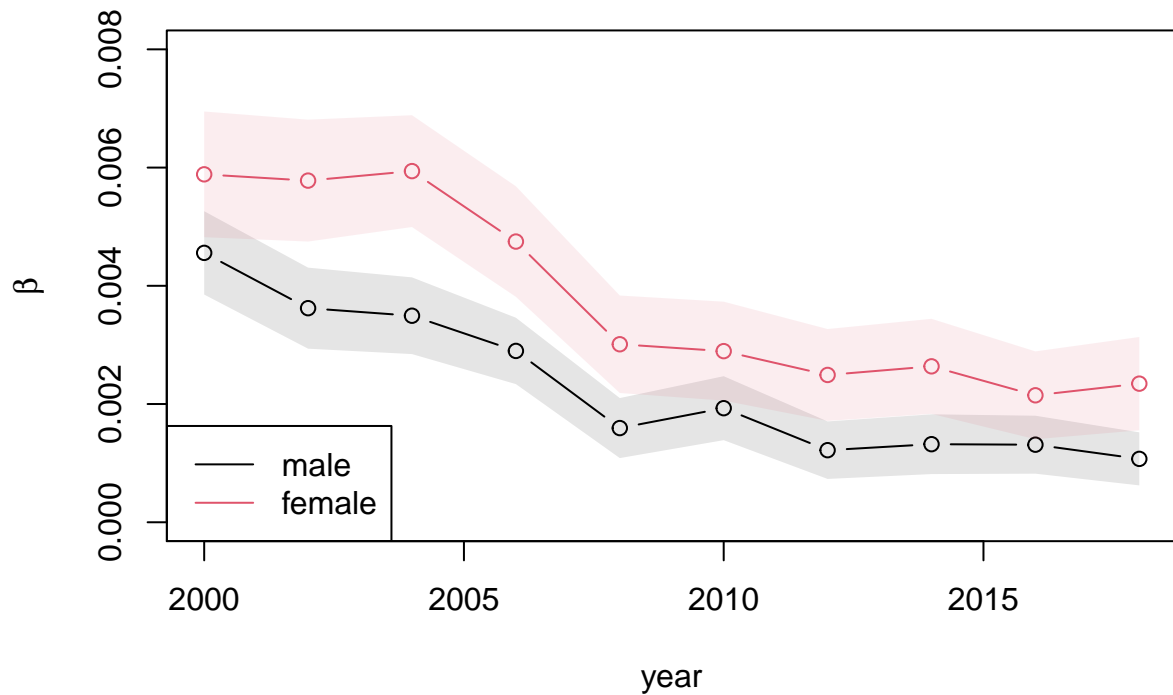
## Waiting for profiling to be done...
coef67b = summary(model67b)$coefficient[,1]
lwr67b = confint(model67b)[,1]
uwr67b = confint(model67b)[,2]
coef68a = summary(model68a)$coefficient[,1]
lwr68a = confint(model68a)[,1]

## Waiting for profiling to be done...
uwr68a = confint(model68a)[,2]

## Waiting for profiling to be done...
coef68b = summary(model68b)$coefficient[,1]
lwr68b = confint(model68b)[,1]
uwr68b = confint(model68b)[,2]

plot(years,coef47a,ylim=c(0,0.008),col=1,type='b',xlab='year',
      ylab=bquote(beta),lwd=1)
polygon(c(rev(years),years),c(rev(lwr47a),uwr47a),
        col=adjustcolor(1,alpha.f=0.1), border=NA)
lines(years,coef48a,col=2,type='b',lwd=1)
polygon(c(rev(years),years),c(rev(lwr48a),uwr48a),
        col=adjustcolor(2,alpha.f=0.1), border=NA)
legend(x='bottomleft',legend=c('male','female'),
       lty = c(1,1),col = c(1,2),cex=1,lwd=1)

```



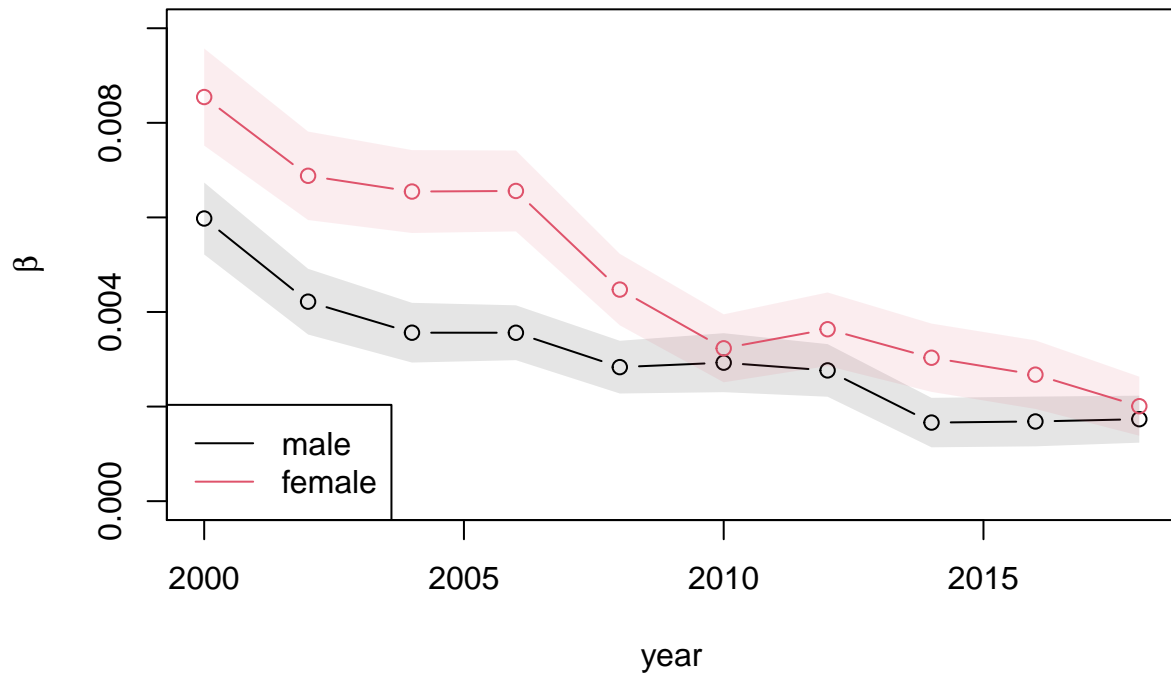
```
plot(years,coef47b,ylim=c(0,0.04),col=1,type='b',xlab='year',
      ylab=bquote(beta),lwd=1)
polygon(c(rev(years),years),c(rev(lwr47b),uwr47b),
        col=adjustcolor(1,alpha.f=0.1), border=NA)
lines(years,coef48b,col=2,type='b',lwd=1)
polygon(c(rev(years),years),c(rev(lwr48b),uwr48b),
        col=adjustcolor(2,alpha.f=0.1), border=NA)
legend(x='bottomleft',legend=c('male','female'),
       lty = c(1,1),col = c(1,2),cex=1,lwd=1)
```



```

plot(years,coef57a,ylim=c(0,0.01),col=1,type='b',xlab='year',
      ylab=bquote(beta),lwd=1)
polygon(c(rev(years),years),c(rev(lwr57a),uwr57a),
        col=adjustcolor(1,alpha.f=0.1), border=NA)
lines(years,coef58a,col=2,type='b',lwd=1)
polygon(c(rev(years),years),c(rev(lwr58a),uwr58a),
        col=adjustcolor(2,alpha.f=0.1), border=NA)
legend(x='bottomleft',legend=c('male','female'),
       lty = c(1,1),col = c(1,2),cex=1,lwd=1)

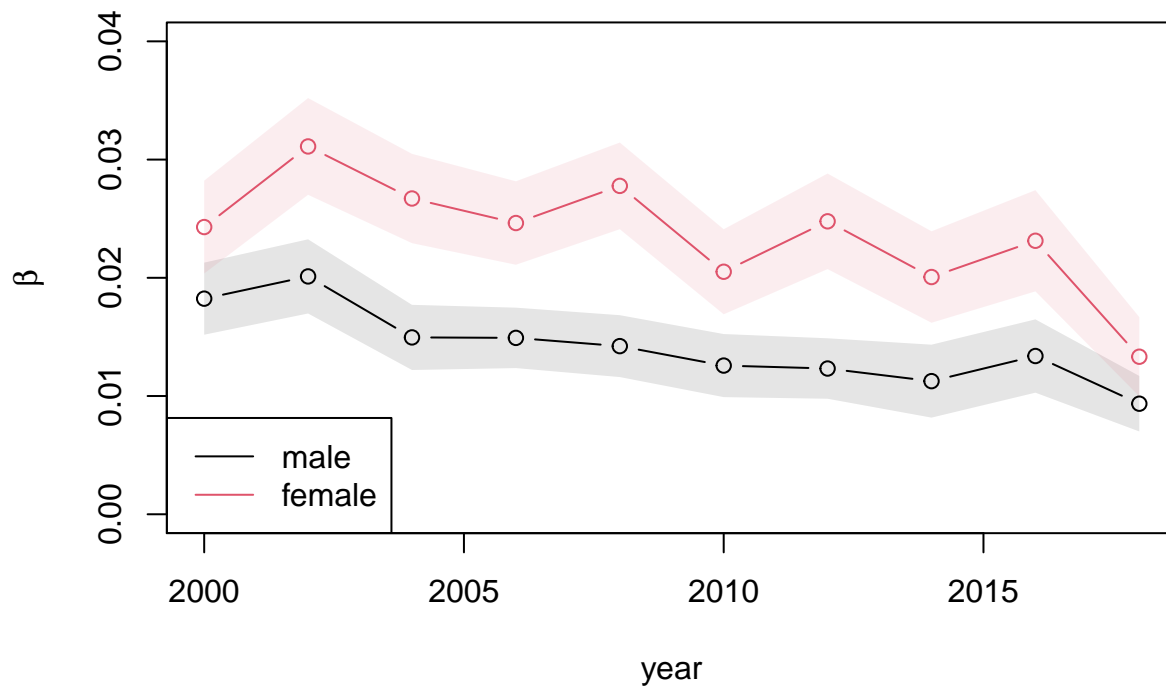
```



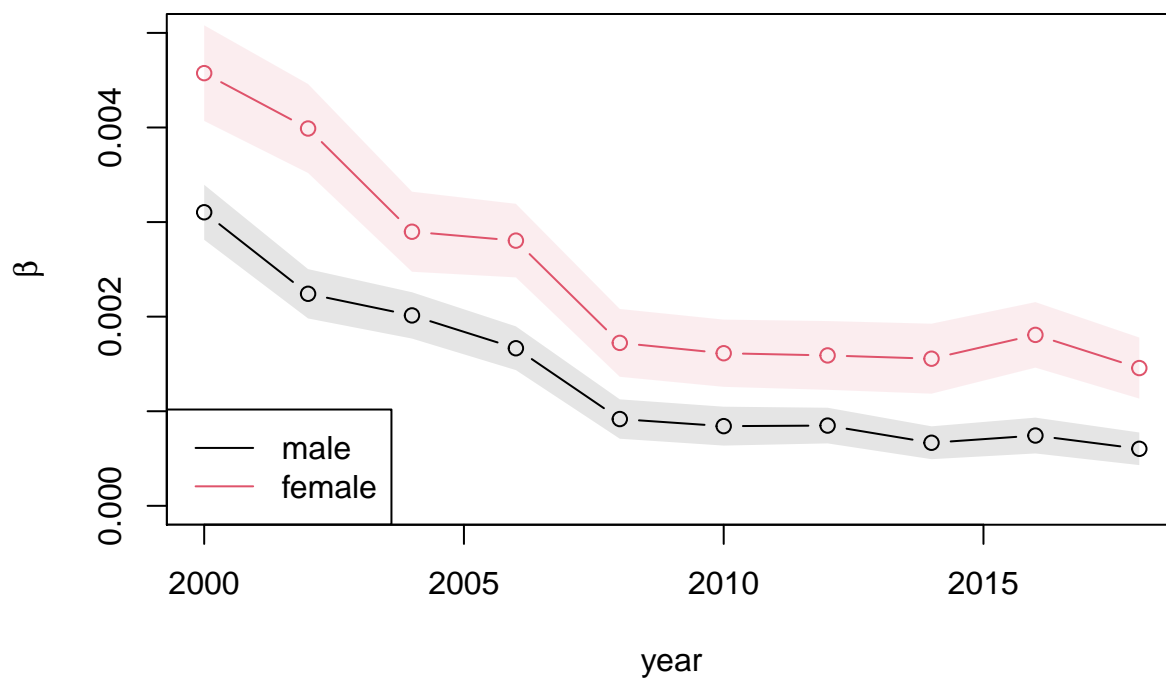
```

plot(years,coef57b,ylim=c(0,0.04),col=1,type='b',xlab='year',
      ylab=bquote(beta),lwd=1)
polygon(c(rev(years),years),c(rev(lwr57b),uwr57b),
        col=adjustcolor(1,alpha.f=0.1), border=NA)
lines(years,coef58b,col=2,type='b',lwd=1)
polygon(c(rev(years),years),c(rev(lwr58b),uwr58b),
        col=adjustcolor(2,alpha.f=0.1), border=NA)
legend(x='bottomleft',legend=c('male','female'),
       lty = c(1,1),col = c(1,2),cex=1,lwd=1)

```



```
plot(years,coef67a,ylim=c(0,0.005),col=1,type='b',xlab='year',
      ylab=bquote(beta),lwd=1)
polygon(c(rev(years),years),c(rev(lwr67a),uwr67a),
        col=adjustcolor(1,alpha.f=0.1), border=NA)
lines(years,coef68a,col=2,type='b',lwd=1)
polygon(c(rev(years),years),c(rev(lwr68a),uwr68a),
        col=adjustcolor(2,alpha.f=0.1), border=NA)
legend(x='bottomleft',legend=c('male','female'),
       lty = c(1,1),col = c(1,2),cex=1,lwd=1)
```



```

plot(years,coef67b,ylim=c(0,0.025),col=1,type='b',xlab='year',
      ylab=bquote(beta),lwd=1)
polygon(c(rev(years),years),c(rev(lwr67b),uwr67b),
        col=adjustcolor(1,alpha.f=0.1), border=NA)
lines(years,coef68b,col=2,type='b',lwd=1)
polygon(c(rev(years),years),c(rev(lwr68b),uwr68b),
        col=adjustcolor(2,alpha.f=0.1), border=NA)
legend(x='bottomleft',legend=c('male','female'),
       lty = c(1,1),col = c(1,2),cex=1,lwd=1)

```



We first present the estimated marginal effects β between men and women for each race in Figure 6 to see how gender difference varies by race. Panels (a), (c), and (e) plot separately for male and female the marginal effects of income on migration's probability, while panels (b), (d), and (e) plot the marginal effects of income on migration's number for Hispanic, Black, and non-Hispanic, non-Black, respectively, with the shaded area being the 95 percent confidence interval.

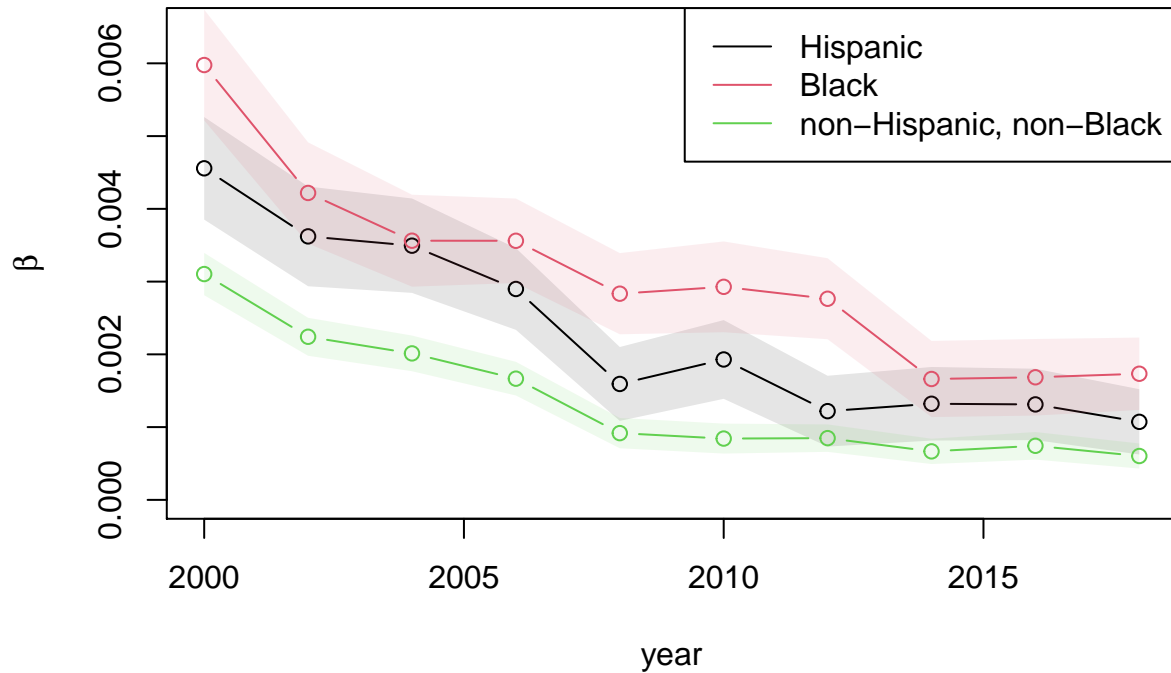
Four points are worth highlighting from our results. First, the marginal effects of income on migration's probability and number are larger for women than for men in all race groups, with some insignificance. Second, the marginal effects of income on migration's probability decline over time for both men and women in all race groups. Third, the marginal effects of income on migration's number decline slightly with fluctuation in general. Fourth, Hispanic and Black are more likely to migrate and migrate more frequently than non-Hispanic, non-Black, as the limits of the horizontal axes in panels (a) and (c) (panels (b) and (d)) are twice as high as that in panel (e) (panel (f)).

```

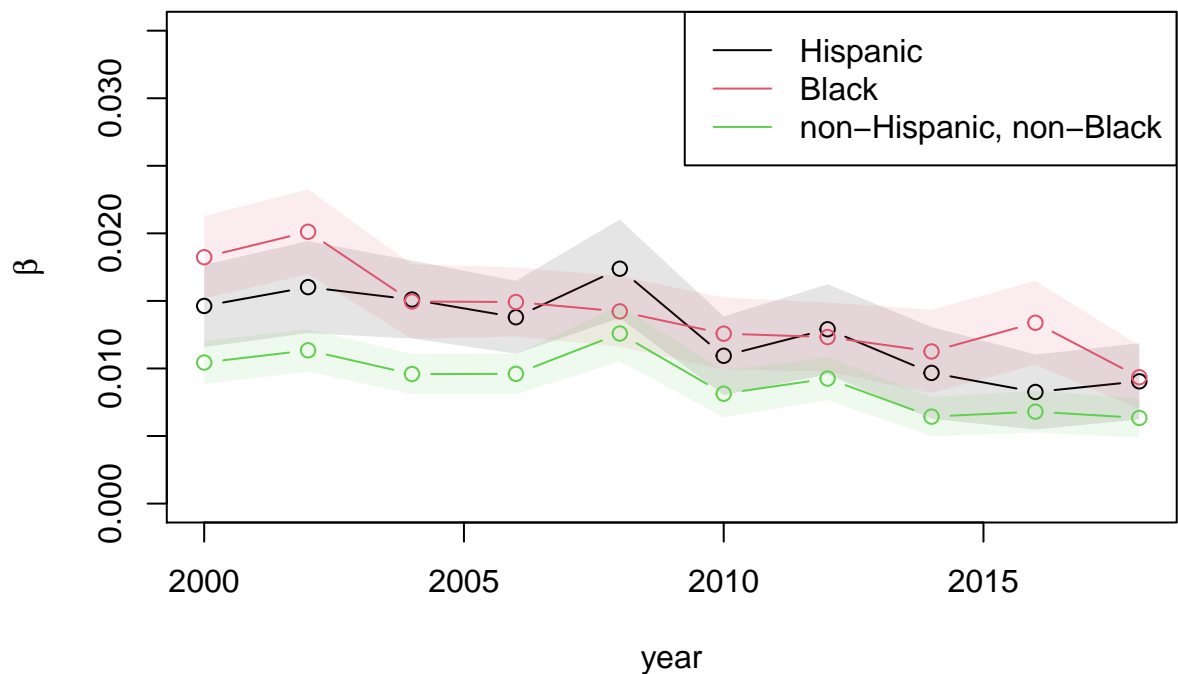
plot(years,coef47a,ylim=c(0,0.0065),col=1,type='b',xlab='year',
      ylab=bquote(beta),lwd=1)
polygon(c(rev(years),years),c(rev(lwr47a),uwr47a),
        col=adjustcolor(1,alpha.f=0.1), border=NA)
lines(years,coef57a,col=2,type='b',lwd=1)
polygon(c(rev(years),years),c(rev(lwr57a),uwr57a),
        col=adjustcolor(2,alpha.f=0.1), border=NA)
lines(years,coef67a,col=3,type='b',lwd=1)
polygon(c(rev(years),years),c(rev(lwr67a),uwr67a),

```

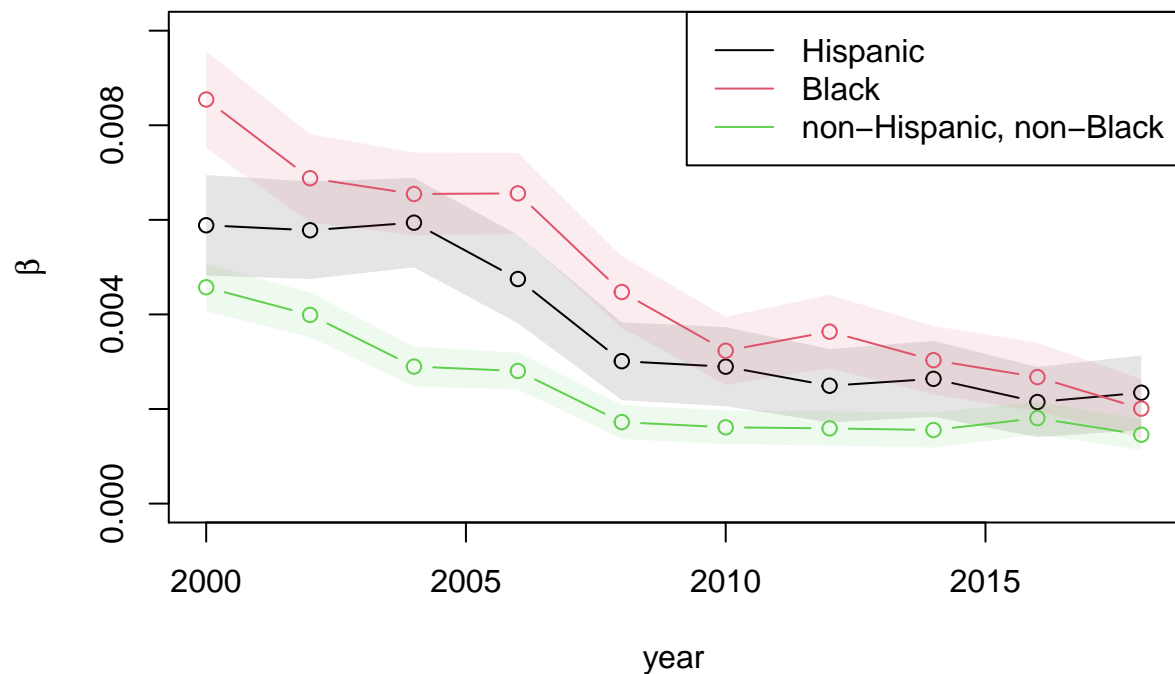
```
col=adjustcolor(3,alpha.f=0.1), border=NA)
legend(x='topright',legend=c('Hispanic','Black','non-Hispanic, non-Black'),
      lty = c(1,1,1),col = c(1,2,3),cex=1,lwd=1)
```



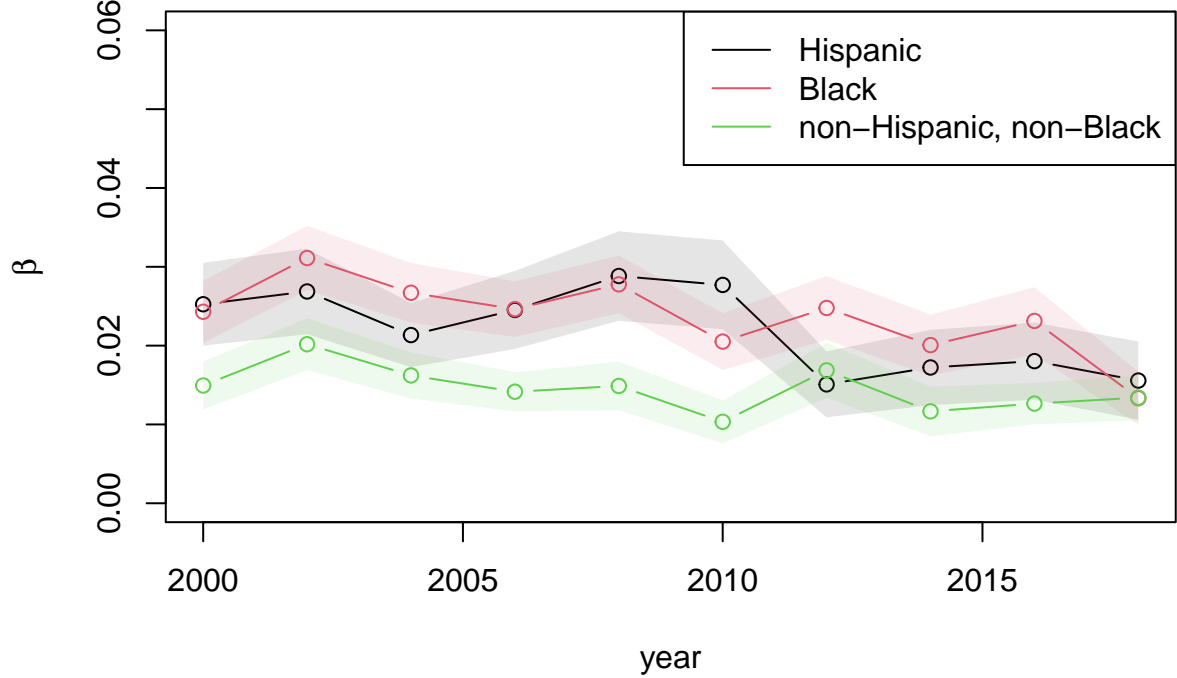
```
plot(years,coef47b,ylim=c(0,0.035),col=1,type='b',xlab='year',
      ylab=bquote(beta),lwd=1)
polygon(c(rev(years),years),c(rev(lwr47b),uwr47b),
        col=adjustcolor(1,alpha.f=0.1), border=NA)
lines(years,coef57b,col=2,type='b',lwd=1)
polygon(c(rev(years),years),c(rev(lwr57b),uwr57b),
        col=adjustcolor(2,alpha.f=0.1), border=NA)
lines(years,coef67b,col=3,type='b',lwd=1)
polygon(c(rev(years),years),c(rev(lwr67b),uwr67b),
        col=adjustcolor(3,alpha.f=0.1), border=NA)
legend(x='topright',legend=c('Hispanic','Black','non-Hispanic, non-Black'),
      lty = c(1,1,1),col = c(1,2,3),cex=1,lwd=1)
```



```
plot(years,coef48a,ylim=c(0,0.01),col=1,type='b',xlab='year',
      ylab=bquote(beta),lwd=1)
polygon(c(rev(years),years),c(rev(lwr48a),uwr48a),
        col=adjustcolor(1,alpha.f=0.1), border=NA)
lines(years,coef58a,col=2,type='b',lwd=1)
polygon(c(rev(years),years),c(rev(lwr58a),uwr58a),
        col=adjustcolor(2,alpha.f=0.1), border=NA)
lines(years,coef68a,col=3,type='b',lwd=1)
polygon(c(rev(years),years),c(rev(lwr68a),uwr68a),
        col=adjustcolor(3,alpha.f=0.1), border=NA)
legend(x='topright',legend=c('Hispanic','Black','non-Hispanic, non-Black'),
       lty = c(1,1,1),col = c(1,2,3),cex=1,lwd=1)
```

```
plot(years,coef48b,ylim=c(0,0.06),col=1,type='b',xlab='year',
      ylab=bquote(beta),lwd=1)
polygon(c(rev(years),years),c(rev(lwr48b),uwr48b),
        col=adjustcolor(1,alpha.f=0.1), border=NA)
lines(years,coef58b,col=2,type='b',lwd=1)
polygon(c(rev(years),years),c(rev(lwr58b),uwr58b),
        col=adjustcolor(2,alpha.f=0.1), border=NA)
lines(years,coef68b,col=3,type='b',lwd=1)
polygon(c(rev(years),years),c(rev(lwr68b),uwr68b),
        col=adjustcolor(3,alpha.f=0.1), border=NA)
legend(x='topright',legend=c('Hispanic','Black','non-Hispanic, non-Black'),
       lty = c(1,1,1),col = c(1,2,3),cex=1,lwd=1)
```



We next present the estimated marginal effects β between race groups for each gender in Figure 7 to see how race difference varies by gender. Panels (a) and (c) plot separately for each race group the marginal effects of income on migration's probability, while panels (b) and (d) plot the marginal effects of income on migration's number for male and female, respectively, with the shaded area being the 95 percent confidence interval.

Four points are worth highlighting from our results. First, the marginal effects of income on migration's probability are largest among Black male and female, except for Black female in 2018, and are smallest among non-Hispanic, non-Black male and female, with some insignificance. Second, the marginal effects of income on migration's probability decline and converge over time for both men and women in all race groups. Third, the marginal effects of income on migration's number are generally smaller among non-Hispanic, non-Black male and female, these effects have been fluctuating in degree among Hispanic and Black, for both male and female, and the general trend is slight downward for both men and women in all race groups. Fourth, women are more likely to migrate and migrate more frequently than men, as the limit of the horizontal axis in panel (c) (panel (d)) is higher than that in panel (a) (panel (b)).

In this section, we have tried to explore how the marginal effects of income vary by race and gender. There are three main findings: (i), non-Hispanic, non-Black are least affected by income change, and men are less affected than women by income change, in terms of their migration's probability and number, (ii), the probability of migration declines over time for both men and women in all race groups, with men and women experiencing convergence in their migration's probability, and (iii), the number of migrations declines slightly with fluctuation over time for both men and women in all race groups.

5 Working Hours and Migration Difference

In this section, we investigate the relationship between migration differences and working hours. Specifically, we rank our sample by quantiles of the distribution of working hour. A higher rank in these distributions implies that the individual had worked for more hours in the past calendar year.

We estimate how migration differences vary by these working hours rank measures. Because here we are not primarily interested in the exact dynamic path of migration differences, we exclude the full set of year dummies and conduct a pooled estimation. In other words, we are considering average migration differences over the 18-year period from 2000 to 2018. To estimate the effect of working hours, we consider specifications

of the following type:

$$M_{icq} = \sum_c \sum_q \alpha_{cq} \cdot \mathbb{I}[c = C] \cdot \mathbb{I}[q = Q] + \epsilon_{icq},$$

where $\mathbb{I}[q = Q]$ is an indicator for working hours being in the quantile Q .

*# The first number indicate income class, and the second number indicates
working hours quantile rank. For example, move_ind11 represent people
in the low income class with working hours below 20 percentile.*

```
move_ind11 = subset(move_ind1, move_ind1$hrcat==1)
move_ind12 = subset(move_ind1, move_ind1$hrcat==2)
move_ind13 = subset(move_ind1, move_ind1$hrcat==3)
move_ind14 = subset(move_ind1, move_ind1$hrcat==4)
move_ind15 = subset(move_ind1, move_ind1$hrcat==5)
move_num11 = subset(move_num1, move_num1$hrcat==1)
move_num12 = subset(move_num1, move_num1$hrcat==2)
move_num13 = subset(move_num1, move_num1$hrcat==3)
move_num14 = subset(move_num1, move_num1$hrcat==4)
move_num15 = subset(move_num1, move_num1$hrcat==5)
```

```
model11a = glm(moved ~ 1, data = move_ind11)
model11b = lm(moves ~ 1, data = move_num11)
model12a = glm(moved ~ 1, data = move_ind12)
model12b = lm(moves ~ 1, data = move_num12)
model13a = glm(moved ~ 1, data = move_ind13)
model13b = lm(moves ~ 1, data = move_num13)
model14a = glm(moved ~ 1, data = move_ind14)
model14b = lm(moves ~ 1, data = move_num14)
model15a = glm(moved ~ 1, data = move_ind15)
model15b = lm(moves ~ 1, data = move_num15)
```

```
move_ind21 = subset(move_ind2, move_ind2$hrcat==1)
move_ind22 = subset(move_ind2, move_ind2$hrcat==2)
move_ind23 = subset(move_ind2, move_ind2$hrcat==3)
move_ind24 = subset(move_ind2, move_ind2$hrcat==4)
move_ind25 = subset(move_ind2, move_ind2$hrcat==5)
move_num21 = subset(move_num2, move_num2$hrcat==1)
move_num22 = subset(move_num2, move_num2$hrcat==2)
move_num23 = subset(move_num2, move_num2$hrcat==3)
move_num24 = subset(move_num2, move_num2$hrcat==4)
move_num25 = subset(move_num2, move_num2$hrcat==5)
```

```
model21a = glm(moved ~ 1, data = move_ind21)
model21b = lm(moves ~ 1, data = move_num21)
model22a = glm(moved ~ 1, data = move_ind22)
model22b = lm(moves ~ 1, data = move_num22)
model23a = glm(moved ~ 1, data = move_ind23)
model23b = lm(moves ~ 1, data = move_num23)
model24a = glm(moved ~ 1, data = move_ind24)
model24b = lm(moves ~ 1, data = move_num24)
model25a = glm(moved ~ 1, data = move_ind25)
model25b = lm(moves ~ 1, data = move_num25)
```

```
move_ind31 = subset(move_ind3, move_ind3$hrcat==1)
move_ind32 = subset(move_ind3, move_ind3$hrcat==2)
```

```

move_ind33 = subset(move_ind3, move_ind3$hrcat==3)
move_ind34 = subset(move_ind3, move_ind3$hrcat==4)
move_ind35 = subset(move_ind3, move_ind3$hrcat==5)
move_num31 = subset(move_num3, move_num3$hrcat==1)
move_num32 = subset(move_num3, move_num3$hrcat==2)
move_num33 = subset(move_num3, move_num3$hrcat==3)
move_num34 = subset(move_num3, move_num3$hrcat==4)
move_num35 = subset(move_num3, move_num3$hrcat==5)

```

```

model31a = glm(moved ~ 1, data = move_ind31)
model31b = lm(moves ~ 1, data = move_num31)
model32a = glm(moved ~ 1, data = move_ind32)
model32b = lm(moves ~ 1, data = move_num32)
model33a = glm(moved ~ 1, data = move_ind33)
model33b = lm(moves ~ 1, data = move_num33)
model34a = glm(moved ~ 1, data = move_ind34)
model34b = lm(moves ~ 1, data = move_num34)
model35a = glm(moved ~ 1, data = move_ind35)
model35b = lm(moves ~ 1, data = move_num35)

```

```

coef11a = summary(model11a)$coefficient[1,1]
coef11b = summary(model11b)$coefficient[1,1]
coef12a = summary(model12a)$coefficient[1,1]
coef12b = summary(model12b)$coefficient[1,1]
coef13a = summary(model13a)$coefficient[1,1]
coef13b = summary(model13b)$coefficient[1,1]
coef14a = summary(model14a)$coefficient[1,1]
coef14b = summary(model14b)$coefficient[1,1]
coef15a = summary(model15a)$coefficient[1,1]
coef15b = summary(model15b)$coefficient[1,1]
coef21a = summary(model21a)$coefficient[1,1]
coef21b = summary(model21b)$coefficient[1,1]
coef22a = summary(model22a)$coefficient[1,1]
coef22b = summary(model22b)$coefficient[1,1]
coef23a = summary(model23a)$coefficient[1,1]
coef23b = summary(model23b)$coefficient[1,1]
coef24a = summary(model24a)$coefficient[1,1]
coef24b = summary(model24b)$coefficient[1,1]
coef25a = summary(model25a)$coefficient[1,1]
coef25b = summary(model25b)$coefficient[1,1]
coef31a = summary(model31a)$coefficient[1,1]
coef31b = summary(model31b)$coefficient[1,1]
coef32a = summary(model32a)$coefficient[1,1]
coef32b = summary(model32b)$coefficient[1,1]
coef33a = summary(model33a)$coefficient[1,1]
coef33b = summary(model33b)$coefficient[1,1]

```

```

## Warning in summary.lm(model33b): essentially perfect fit: summary may be
## unreliable

```

```

coef34a = summary(model34a)$coefficient[1,1]
coef34b = summary(model34b)$coefficient[1,1]
coef35a = summary(model35a)$coefficient[1,1]
coef35b = summary(model35b)$coefficient[1,1]

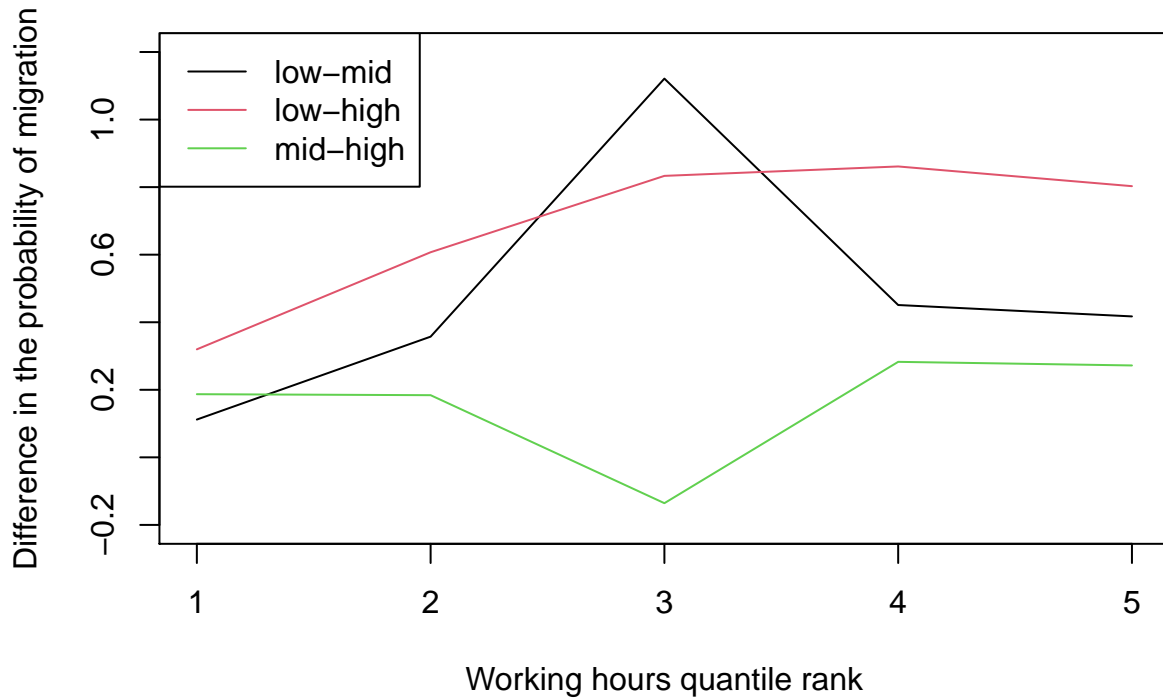
```

*# The first number indicate lower income class, the second number indicates
higher income class, and the third number indicates working hours quantile
rank. For example, diff121 represents migration difference between people in # and middle income class*

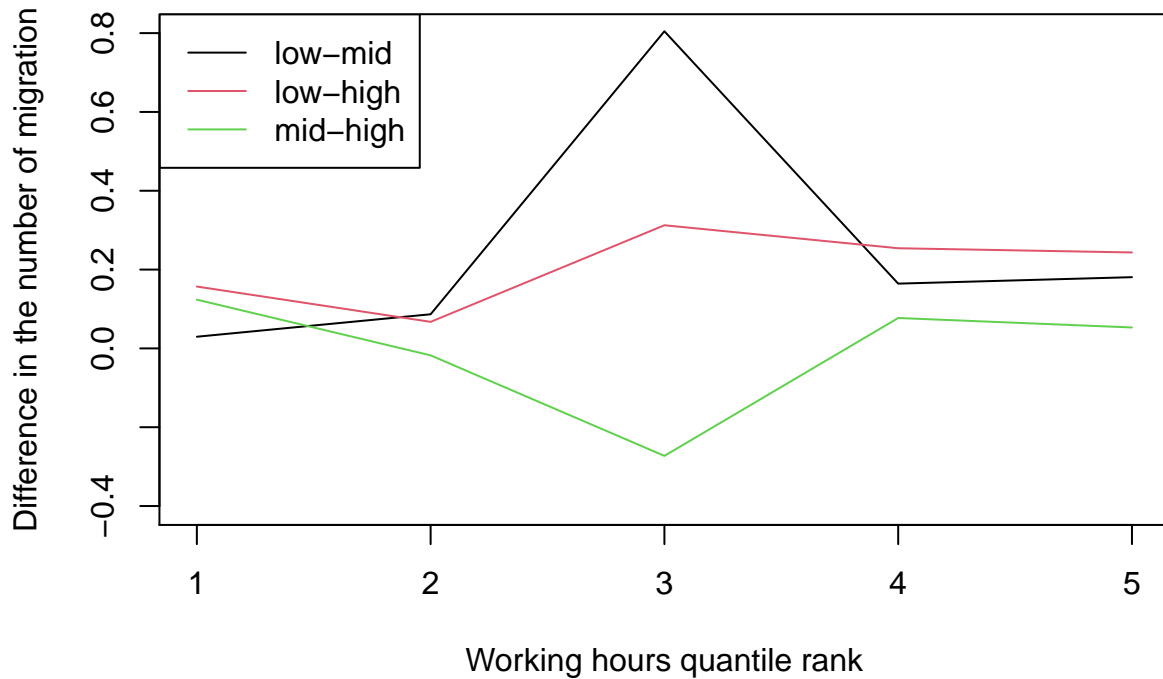
```
diff12a = c((coef11a-coef21a)/mean(move_ind21$moved),
            (coef12a-coef22a)/mean(move_ind22$moved),
            (coef13a-coef23a)/mean(move_ind23$moved),
            (coef14a-coef24a)/mean(move_ind24$moved),
            (coef15a-coef25a)/mean(move_ind25$moved))
diff13a = c((coef11a-coef31a)/mean(move_ind31$moved),
            (coef12a-coef32a)/mean(move_ind32$moved),
            (coef13a-coef33a)/mean(move_ind33$moved),
            (coef14a-coef34a)/mean(move_ind34$moved),
            (coef15a-coef35a)/mean(move_ind35$moved))
diff23a = c((coef21a-coef31a)/mean(move_ind31$moved),
            (coef22a-coef32a)/mean(move_ind32$moved),
            (coef23a-coef33a)/mean(move_ind33$moved),
            (coef24a-coef34a)/mean(move_ind34$moved),
            (coef25a-coef35a)/mean(move_ind35$moved))
```

```
diff12b = c((coef11b-coef21b)/mean(move_num21$moves),
            (coef12b-coef22b)/mean(move_num22$moves),
            (coef13b-coef23b)/mean(move_num23$moves),
            (coef14b-coef24b)/mean(move_num24$moves),
            (coef15b-coef25b)/mean(move_num25$moves))
diff13b = c((coef11b-coef31b)/mean(move_num31$moves),
            (coef12b-coef32b)/mean(move_num32$moves),
            (coef13b-coef33b)/mean(move_num33$moves),
            (coef14b-coef34b)/mean(move_num34$moves),
            (coef15b-coef35b)/mean(move_num35$moves))
diff23b = c((coef21b-coef31b)/mean(move_num31$moves),
            (coef22b-coef32b)/mean(move_num32$moves),
            (coef23b-coef33b)/mean(move_num33$moves),
            (coef24b-coef34b)/mean(move_num34$moves),
            (coef25b-coef35b)/mean(move_num35$moves))
```

```
hrcat = 1:5
plot(hrcat,diff12a,ylim=c(-0.2,1.2),col=1,type='l',
     xlab='Working hours quantile rank',
     ylab='Difference in the probability of migration',lwd=1)
lines(hrcat,diff13a,col=2,type='l',lwd=1)
lines(hrcat,diff23a,col=3,type='l',lwd=1)
legend(x='topleft',legend=c('low-mid','low-high','mid-high'),
      lty=c(1,1,1),col=c(1,2,3),cex=1,lwd=1)
```



```
plot(hrcat,diff12b,ylim=c(-0.4,0.8),col=1,type='l',
     xlab='Working hours quantile rank',
     ylab='Difference in the number of migration',lwd=1)
lines(hrcat,diff13b,col=2,type='l',lwd=1)
lines(hrcat,diff23b,col=3,type='l',lwd=1)
legend(x='topleft',legend=c('low-mid','low-high','mid-high'),
      lty=c(1,1,1),col=c(1,2,3),cex=1,lwd=1)
```



The results are presented in Figure 5. This figure plots the difference in the probability of migration (Panel (a)) and the number of migrations (Panel (b)) against quintiles of the working hours distribution. In this

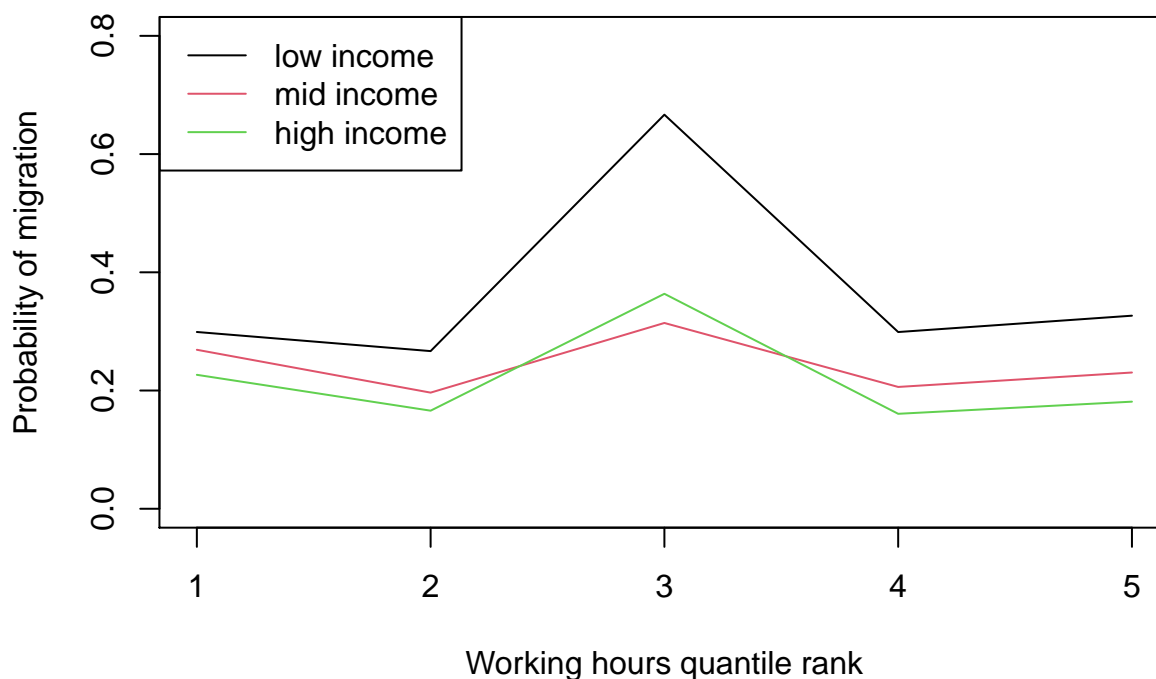
figure, we are not including controls for family size and gender.

Panel (a) of the figure shows that the migration difference between low and middle income classes reaches its maximum for people with working hours being in the third quantile, while the difference between middle and high income classes reaches its minimum for these people. In fact, by our specification, the green line in panel (a) is the difference between the black and the red lines so it is negative at the third quantile, implying that people with middle income and median working hours migrate with a lower probability than those with high income. The effects of going from the the second quantile to the third quantile and from the the third quantile to the fourth quantile are quite large for the low-middle and the middle-high income class pairs, with the former being opposite to the latter. This suggests that either people with low (middle) income and median working hours are much more (less) likely to migrate than those who work slightly more or less, or the contrary for people in middle (high) income class. Similar results are shown in Panel (b).

Our next step is then to investigate whether the effects of going from the the second quantile to the third quantile and from the the third quantile to the fourth quantile are caused by low (middle) income class or middle (high) income class.

```
move_prob1a = c(coef11a,coef12a,coef13a,coef14a,coef15a)
move_number1b = c(coef11b,coef12b,coef13b,coef14b,coef15b)
move_prob2a = c(coef21a,coef22a,coef23a,coef24a,coef25a)
move_number2b = c(coef21b,coef22b,coef23b,coef24b,coef25b)
move_prob3a = c(coef31a,coef32a,coef33a,coef34a,coef35a)
move_number3b = c(coef31b,coef32b,coef33b,coef34b,coef35b)
```

```
plot(hrcat,move_prob1a,ylim=c(0,0.8),col=1,type='l',
     xlab='Working hours quantile rank',
     ylab='Probability of migration',lwd=1)
lines(hrcat,move_prob2a,col=2,type='l',lwd=1)
lines(hrcat,move_prob3a,col=3,type='l',lwd=1)
legend(x='topleft',legend=c('low income','mid income','high income'),
      lty=c(1,1,1),col=c(1,2,3),cex=1,lwd=1)
```

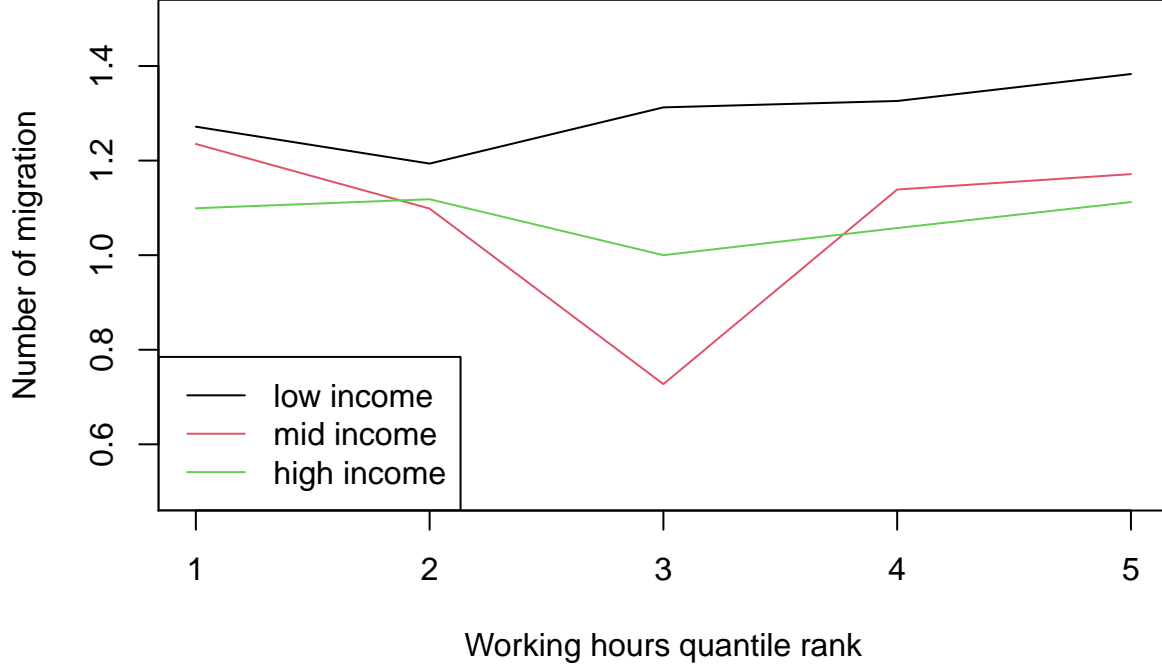


```
plot(hrcat,move_number1b,ylim=c(0.5,1.5),col=1,type='l',
     xlab='Working hours quantile rank',
```

```

ylab='Number of migration',lwd=1)
lines(hrcat,move_number2b,col=2,type='l',lwd=1)
lines(hrcat,move_number3b,col=3,type='l',lwd=1)
legend(x='bottomleft',legend=c('low income','mid income','high income'),
      lty=c(1,1,1),col=c(1,2,3),cex=1,lwd=1)

```



The results are presented in Figure 6. This figure plots the probability of migration (Panel (a)) and the number of migrations (Panel (b)) against quintiles of the working hours distribution. In this figure, we are not including controls for family size and gender.

Panel (a) of the figure shows that for the probability of migration, people in low income class experience a sharp rise as they go from the second quantile to the third quantile and a sharp fall as they go from the third quantile to the fourth quantile, while people in middle income class only experience a slight rise and a slight fall, implying that the fluctuation in the difference between low and middle income class is mainly caused by people in low income class. As for the difference between middle and high income classes, we see that people in high income class experience a larger magnitude of increase and decrease relative to people in middle income class as they go from the second quantile to the third quantile and from the third quantile to the fourth quantile, implying that the fluctuation in the difference between middle and high income class is attributed more to people in high income class.

Panel (b) of the figure shows that for the number of migration, people in middle income class experience a sharp fall as they go from the second quantile to the third quantile and a sharp rise as they go from the third quantile to the fourth quantile, while people in low income class only experience a slight rise and no visible fall, implying that the fluctuation in the difference between low and middle income class is mainly caused by people in middle income class. An analogous argument applies for the fluctuation in the difference between middle and high income class.

We now included controls for family size and gender and consider specifications of the following type:

$$\begin{aligned}
M_{icq} = & \sum_c \sum_q \alpha_{cq} \cdot \mathbb{I}[c = C] \cdot \mathbb{I}[q = Q] + \sum_c \sum_q \beta_{cq} \cdot famsize_{icq} \cdot \mathbb{I}[c = C] \cdot \mathbb{I}[q = Q] \\
& + \sum_c \sum_q \gamma_{cq} \cdot \mathbb{I}[c = C] \cdot \mathbb{I}[q = Q] \cdot \mathbb{I}[g = G] + \epsilon_{icq},
\end{aligned}$$

where $\mathbb{I}[g = G]$ is an indicator for gender being G .

```
model11c = glm(moved ~ 1 + famsize + sex_1, data = move_ind11)
model11d = lm(moves ~ 1 + famsize + sex_1, data = move_num11)
model12c = glm(moved ~ 1 + famsize + sex_1, data = move_ind12)
model12d = lm(moves ~ 1 + famsize + sex_1, data = move_num12)
model13c = glm(moved ~ 1 + famsize + sex_1, data = move_ind13)
model13d = lm(moves ~ 1 + famsize + sex_1, data = move_num13)
model14c = glm(moved ~ 1 + famsize + sex_1, data = move_ind14)
model14d = lm(moves ~ 1 + famsize + sex_1, data = move_num14)
model15c = glm(moved ~ 1 + famsize + sex_1, data = move_ind15)
model15d = lm(moves ~ 1 + famsize + sex_1, data = move_num15)
```

```
model21c = glm(moved ~ 1 + famsize + sex_1, data = move_ind21)
model21d = lm(moves ~ 1 + famsize + sex_1, data = move_num21)
model22c = glm(moved ~ 1 + famsize + sex_1, data = move_ind22)
model22d = lm(moves ~ 1 + famsize + sex_1, data = move_num22)
model23c = glm(moved ~ 1 + famsize + sex_1, data = move_ind23)
model23d = lm(moves ~ 1 + famsize + sex_1, data = move_num23)
model24c = glm(moved ~ 1 + famsize + sex_1, data = move_ind24)
model24d = lm(moves ~ 1 + famsize + sex_1, data = move_num24)
model25c = glm(moved ~ 1 + famsize + sex_1, data = move_ind25)
model25d = lm(moves ~ 1 + famsize + sex_1, data = move_num25)
```

```
model31c = glm(moved ~ 1 + famsize + sex_1, data = move_ind31)
model31d = lm(moves ~ 1 + famsize + sex_1, data = move_num31)
model32c = glm(moved ~ 1 + famsize + sex_1, data = move_ind32)
model32d = lm(moves ~ 1 + famsize + sex_1, data = move_num32)
model33c = glm(moved ~ 1 + famsize + sex_1, data = move_ind33)
model33d = lm(moves ~ 1 + famsize + sex_1, data = move_num33)
model34c = glm(moved ~ 1 + famsize + sex_1, data = move_ind34)
model34d = lm(moves ~ 1 + famsize + sex_1, data = move_num34)
model35c = glm(moved ~ 1 + famsize + sex_1, data = move_ind35)
model35d = lm(moves ~ 1 + famsize + sex_1, data = move_num35)
```

```
coef11c = summary(model11c)$coefficient[1,1]
coef11d = summary(model11d)$coefficient[1,1]
coef12c = summary(model12c)$coefficient[1,1]
coef12d = summary(model12d)$coefficient[1,1]
coef13c = summary(model13c)$coefficient[1,1]
coef13d = summary(model13d)$coefficient[1,1]
coef14c = summary(model14c)$coefficient[1,1]
coef14d = summary(model14d)$coefficient[1,1]
coef15c = summary(model15c)$coefficient[1,1]
coef15d = summary(model15d)$coefficient[1,1]
coef21c = summary(model21c)$coefficient[1,1]
coef21d = summary(model21d)$coefficient[1,1]
coef22c = summary(model22c)$coefficient[1,1]
coef22d = summary(model22d)$coefficient[1,1]
coef23c = summary(model23c)$coefficient[1,1]
coef23d = summary(model23d)$coefficient[1,1]
coef24c = summary(model24c)$coefficient[1,1]
coef24d = summary(model24d)$coefficient[1,1]
coef25c = summary(model25c)$coefficient[1,1]
coef25d = summary(model25d)$coefficient[1,1]
```

```
coef31c = summary(model31c)$coefficient[1,1]
coef31d = summary(model31d)$coefficient[1,1]
coef32c = summary(model32c)$coefficient[1,1]
coef32d = summary(model32d)$coefficient[1,1]
coef33c = summary(model33c)$coefficient[1,1]
coef33d = summary(model33d)$coefficient[1,1]
```

```
## Warning in summary.lm(model33d): essentially perfect fit: summary may be
## unreliable
```

```
coef34c = summary(model34c)$coefficient[1,1]
coef34d = summary(model34d)$coefficient[1,1]
coef35c = summary(model35c)$coefficient[1,1]
coef35d = summary(model35d)$coefficient[1,1]
```

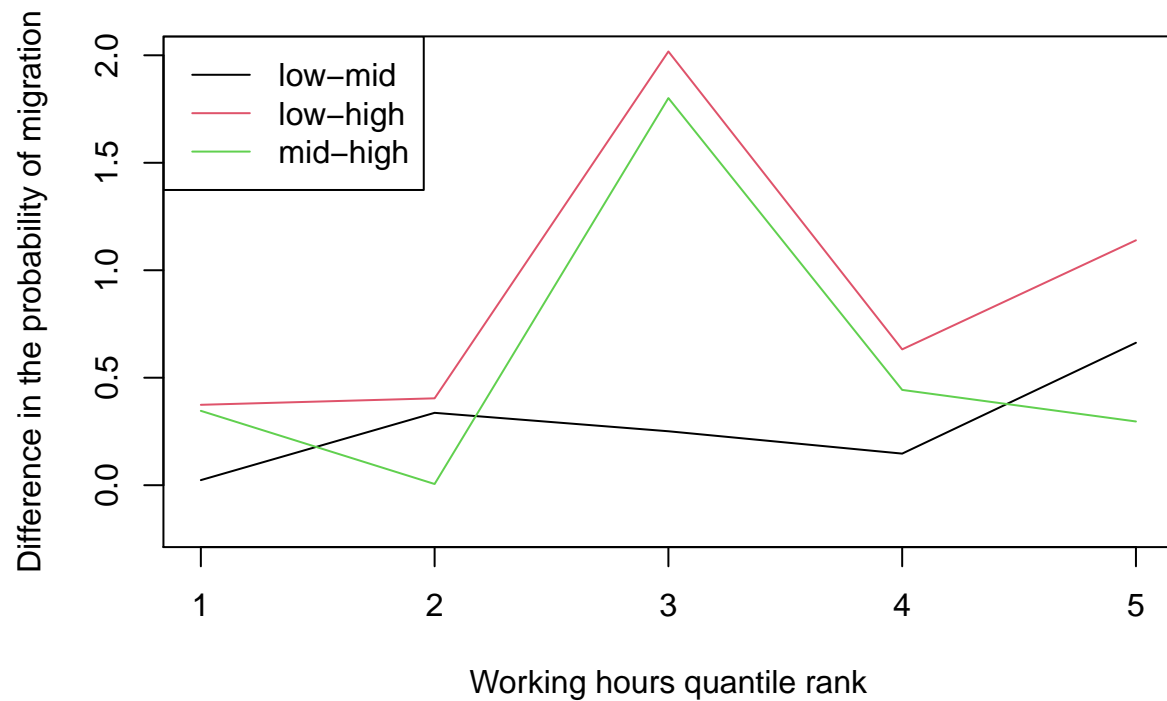
*# The first number indicate lower income class, the second number indicates
higher income class, and the third number indicates working hours quantile
rank. For example, diff121 represents migration difference between people in # and middle income class*

```
diff12c = c((coef11c-coef21c)/mean(move_ind21$moved),
            (coef12c-coef22c)/mean(move_ind22$moved),
            (coef13c-coef23c)/mean(move_ind23$moved),
            (coef14c-coef24c)/mean(move_ind24$moved),
            (coef15c-coef25c)/mean(move_ind25$moved))
diff13c = c((coef11c-coef31c)/mean(move_ind31$moved),
            (coef12c-coef32c)/mean(move_ind32$moved),
            (coef13c-coef33c)/mean(move_ind33$moved),
            (coef14c-coef34c)/mean(move_ind34$moved),
            (coef15c-coef35c)/mean(move_ind35$moved))
diff23c = c((coef21c-coef31c)/mean(move_ind31$moved),
            (coef22c-coef32c)/mean(move_ind32$moved),
            (coef23c-coef33c)/mean(move_ind33$moved),
            (coef24c-coef34c)/mean(move_ind34$moved),
            (coef25c-coef35c)/mean(move_ind35$moved))
```

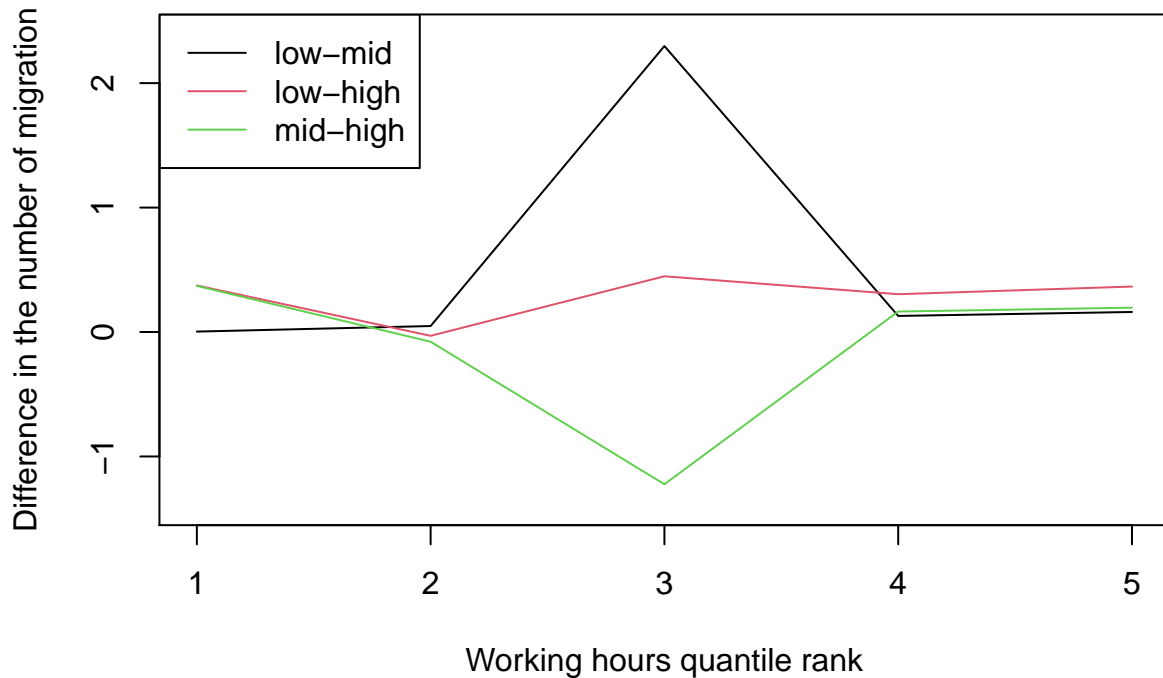
```
diff12d = c((coef11d-coef21d)/mean(move_num21$moves),
            (coef12d-coef22d)/mean(move_num22$moves),
            (coef13d-coef23d)/mean(move_num23$moves),
            (coef14d-coef24d)/mean(move_num24$moves),
            (coef15d-coef25d)/mean(move_num25$moves))
diff13d = c((coef11d-coef31d)/mean(move_num31$moves),
            (coef12d-coef32d)/mean(move_num32$moves),
            (coef13d-coef33d)/mean(move_num33$moves),
            (coef14d-coef34d)/mean(move_num34$moves),
            (coef15d-coef35d)/mean(move_num35$moves))
diff23d = c((coef21d-coef31d)/mean(move_num31$moves),
            (coef22d-coef32d)/mean(move_num32$moves),
            (coef23d-coef33d)/mean(move_num33$moves),
            (coef24d-coef34d)/mean(move_num34$moves),
            (coef25d-coef35d)/mean(move_num35$moves))
```

```
plot(hrcat,diff12c,ylim=c(-0.2,2),col=1,type='l',
     xlab='Working hours quantile rank',
     ylab='Difference in the probability of migration',lwd=1)
lines(hrcat,diff13c,col=2,type='l',lwd=1)
```

```
lines(hrcat,diff23c,col=3,type='l',lwd=1)
legend(x='topleft',legend=c('low-mid','low-high','mid-high'),
      lty=c(1,1,1),col=c(1,2,3),cex=1,lwd=1)
```



```
plot(hrcat,diff12d,ylim=c(-1.4,2.4),col=1,type='l',
     xlab='Working hours quantile rank',
     ylab='Difference in the number of migration',lwd=1)
lines(hrcat,diff13d,col=2,type='l',lwd=1)
lines(hrcat,diff23d,col=3,type='l',lwd=1)
legend(x='topleft',legend=c('low-mid','low-high','mid-high'),
      lty=c(1,1,1),col=c(1,2,3),cex=1,lwd=1)
```



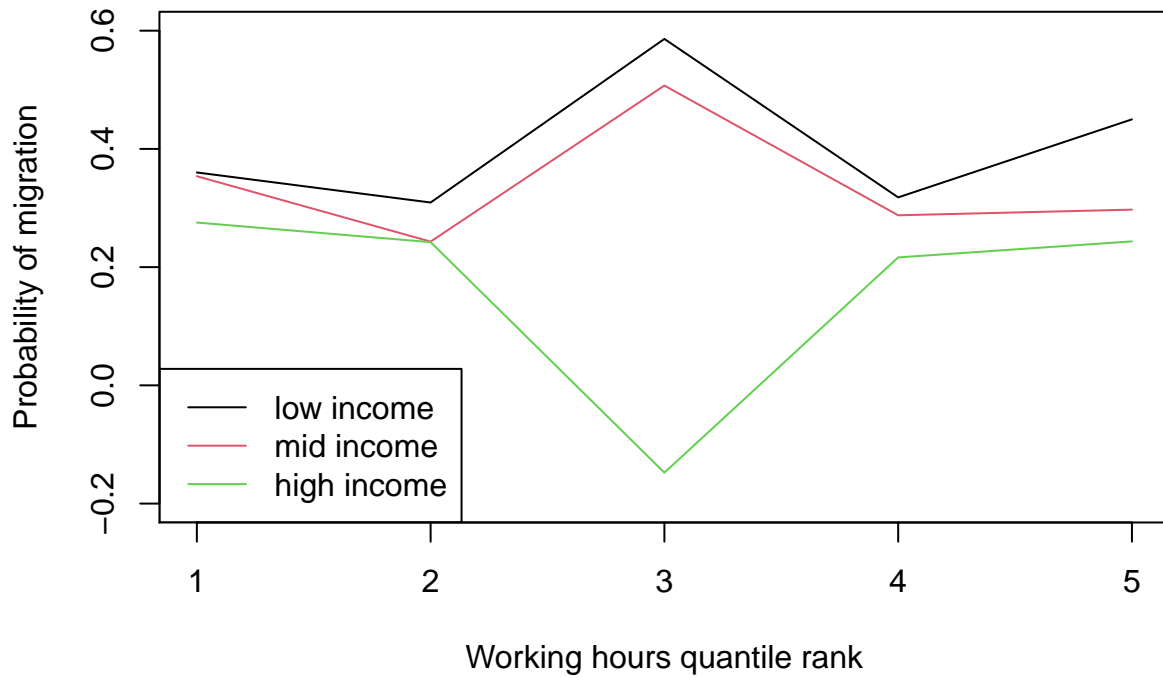
The results are presented in Figure 6. This figure plots the difference in the probability of migration (Panel (a)) and the number of migrations (Panel (b)) against quintiles of the working hours distribution.

Panel (a) of the figure shows that after controlling for family size and gender, both the migration difference between low and high income class and the difference between middle and high income class reach their maximum for people with working hours being in the third quantile. The result for the middle-high income class pair is the opposite to what we find when family size and gender are not included. The effects of going from the the second quantile to the third quantile and from the the third quantile to the fourth quantile are quite large for the low-high and the middle-high income class pairs, with the former exceeding the latter. This suggests that either people with low or middle income and median working hours are much more likely to migrate than those who work slightly more or less, or the contrary for people in high income class. Panel (b) is consistent with what we find when family size and gender are not included, with increased limit of the horizontal axis, implying greater difference in the number of migrations.

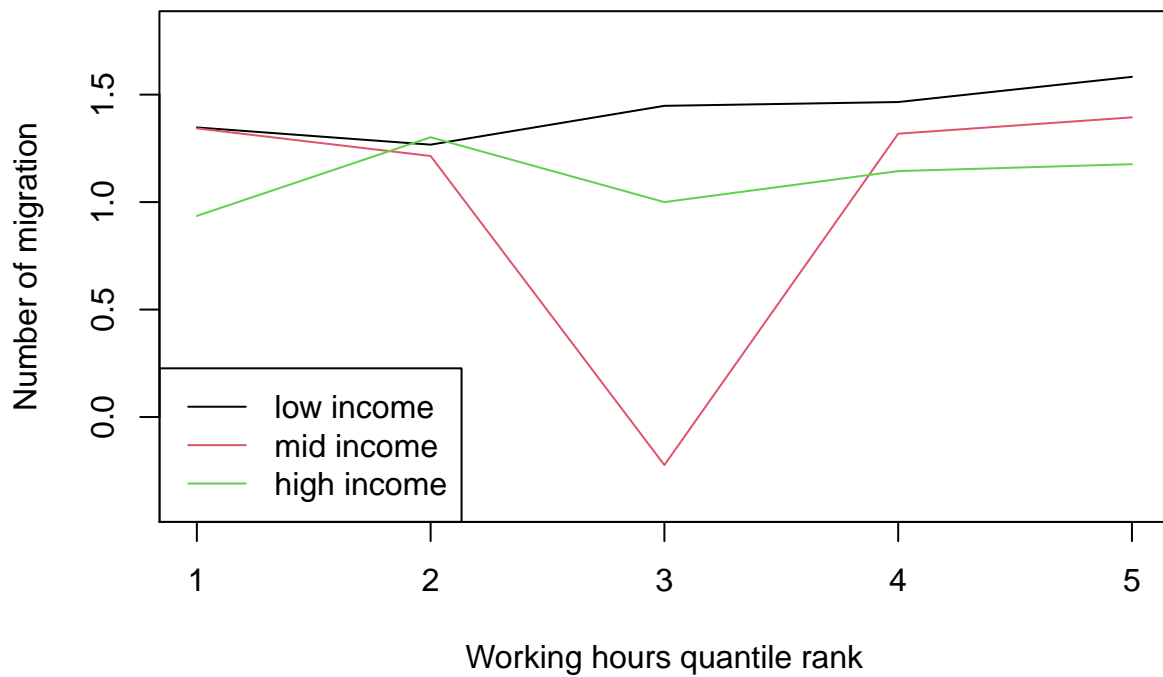
Our next step is again to investigate whether the effects of going from the the second quantile to the third quantile and from the the third quantile to the fourth quantile are caused by low and middle income classes or high income class.

```
move_prob1c = c(coef11c,coef12c,coef13c,coef14c,coef15c)
move_number1d = c(coef11d,coef12d,coef13d,coef14d,coef15d)
move_prob2c = c(coef21c,coef22c,coef23c,coef24c,coef25c)
move_number2d = c(coef21d,coef22d,coef23d,coef24d,coef25d)
move_prob3c = c(coef31c,coef32c,coef33c,coef34c,coef35c)
move_number3d = c(coef31d,coef32d,coef33d,coef34d,coef35d)

plot(hrcat,move_prob1c,ylim=c(-0.2,0.6),col=1,type='l',
     xlab='Working hours quantile rank',
     ylab='Probability of migration',lwd=1)
lines(hrcat,move_prob2c,col=2,type='l',lwd=1)
lines(hrcat,move_prob3c,col=3,type='l',lwd=1)
legend(x='bottomleft',legend=c('low income','mid income','high income'),
      lty=c(1,1,1),col=c(1,2,3),cex=1,lwd=1)
```



```
plot(hrcat,move_number1d,ylim=c(-0.4,1.8),col=1,type='l',
     xlab='Working hours quantile rank',
     ylab='Number of migration',lwd=1)
lines(hrcat,move_number2d,col=2,type='l',lwd=1)
lines(hrcat,move_number3d,col=3,type='l',lwd=1)
legend(x='bottomleft',legend=c('low income','mid income','high income'),
      lty=c(1,1,1),col=c(1,2,3),cex=1,lwd=1)
```



The results are presented in Figure 6. This figure plots the probability of migration (Panel (a)) and the number of migrations (Panel (b)) against quintiles of the working hours distribution.

Panel (a) of the figure shows that for the probability of migration, people in low and middle income classes experience sharp rises as they go from the second quantile to the third quantile and sharp falls as they go from the third quantile to the fourth quantile, while people in high income class experience the opposite, implying that the fluctuation in the differences between low-high and middle-high income class pairs can be attributed to both sides.

Panel (b) of the figure shows that for the number of migration, people in middle income class experience a sharp fall as they go from the second quantile to the third quantile and a sharp rise as they go from the third quantile to the fourth quantile, while people in low (high) income class only experience a slight rise (slight fall) and no visible change (slight rise), implying that the fluctuation in the difference between low (high) and middle income class is mainly caused by people in middle income class.

In this section, we have tried to explore how migration difference varies by working hours. We also decompose the migration difference into what can be attributed to the lower income class and what can be attributed to the higher income class. The findings depend on whether we control for family size and gender or not, but both approaches agree upon three things: (i), the effects of going from the second quantile to the third quantile and from the third quantile to the fourth quantile are large in general, (ii), there is almost no instance where people in low income class are more likely to migrate or migrate more frequently than people in the other two income classes, except for the number of migration between people in low income and in high income classes with working hours being in the second quantile, and (iii), the difference in migration's number between low and middle income classes reaches its maximum and is positive while this difference between middle and high income classes reaches its minimum and is negative for people with working hours being in the third quantile.

6 Conclusion

We hope to continue exploring other avenues to augment our findings. Perhaps the introduction of additional variables such as the number of children per family or marital status could change the probabilities of migration or average migration and other decision paths. The use of instrumental variables also needs further exploration as over or under estimates can produce unfortunate consequences in our model. In our paper, we first estimated how migration status varies by these income classes. We see that the share of migrations declines for all three classes, with the low-income class having the largest share and the high-income class having the smallest share across our timeframe. Also, the average number of migrations for people in the high-income class has experienced an increase in the number of migrations of over 50 percent from 2000 to 2018, far greater than the other two classes. We also found how significant family size factors into the decision-making process of migration. Overall, we hope to have presented a unique approach in estimating the parameters that influence migration movement across the United States and provide new paths for further exploration.