```
log using "D:\SkintoneOutput_ver3_clean.log", replace

/*
Replication code and output for:

Devaraj, S., Quigley, N., and Patel, P.C. The Effects of Skin Tone, Height, and Gender on Objective Career Success. PLOS ONE.

Data availability:
NLSY 1997 data has two versions – one version that is publicly available and the other with geocode information available through special request to the Bureau of Labor Statistics (BLS) in the US. We have used the latter version to control for location fixed-effects. Therefore we cannot publicly post the data, however, it is available from BLS upon request. All requests related to NLSY geocode data should be directed to: NLSYGeocode@bls.gov .

Replication code:
To facilitate replication, we have uploaded the following on Harvard Dataverse:
(i) An end-to-end syntax file in Stata 15 for replication using NLSY 1997 data provided by the BLS to us on October 28, 2014 by Employment Research and Program Development Staff.
(ii) an end-to-end raw output file from the analysis completed in Stata 15.

Devaraj, Srikant; Quigley, Narda R.; Patel, Pankaj C., 2017, "Replication files for "The Effects of Skin Tone, Height, and Gender on Earnings", doi:10.7910/DVN/XGNHSA, Harvard Dataverse,V1/2

*/

use Finaldataset.dta

xtset PUBID_1997 year

by PUBID_1997: gen race_eth = KEY_RACE_ETHNICITY_[1]
count if race_eth ==.
gen color = skincolor
replace color=0 if race_eth ==4


by PUBID_1997 : gen gender = KEY_SEX_[1]
clonevar female = gender
recode female (1=0) (2=1)

gen male=female
recode male (0=1) (1=0)

gen white=.
replace white=1 if race_eth ==4
replace white=0 if race_eth !=4 & race_eth !=.

gen anyblack=.
replace anyblack=1 if race_eth==1
replace anyblack=0 if race_eth!=1 & race_eth !=.

gen lightblack=.
replace lightblack=1 if race_eth ==1 & (skincolor >=1 & skincolor <=5)
replace lightblack=0 if lightblack ==. & race_eth!=.

gen medblack=.
replace medblack=1 if race_eth ==1 & (skincolor >=6 & skincolor <=7)
replace medblack =0 if medblack==.  & race_eth!=.

gen darkblack=.
replace darkblack=1 if race_eth ==1 & (skincolor >=8 & skincolor <=10)
replace darkblack=0 if darkblack ==.  & race_eth!=.

clonevar totincome = YINC_1700_
gen lntotinc = ln(totincome+1)

gen cpiadj = .
replace cpiadj =1.40121 if year==1997
replace cpiadj =1.37987 if year==1998
```

```
  63   replace cpiadj =1.35025 if year==1999
  64   replace cpiadj =1.30627 if year==2000
  65   replace cpiadj =1.27049 if year==2001
  66   replace cpiadj =1.25054 if year==2002
  67   replace cpiadj =1.22244 if year==2003
  68   replace cpiadj =1.19068 if year==2004
  69   replace cpiadj =1.15191 if year==2005
  70   replace cpiadj =1.11595 if year==2006
  71   replace cpiadj =1.08481 if year==2007
  72   replace cpiadj =1.04495 if year==2008
  73   replace cpiadj =1.04831 if year==2009
  74   replace cpiadj =1.03143 if year==2010
  75   replace cpiadj =1 if year==2011
  76
  77   gen totincomeadj = totincome * cpiadj
  78   gen lntotincadj = ln(totincomeadj +1)
  79
  80
  81   clonevar heightfta=YSAQ_000A000001
  82   egen heightft = rowtotal (YHEA_2050 YSAQ_000A_000001 YSAQ_000A000001), missing
  83   recode heightft (-1=.) (-2=.) (-3=.) (-4=.) (-5=.)
  84
  85   egen heightinch = rowtotal (YHEA_2100 YSAQ_000A_000002 YSAQ_000A000002), missing
  86   recode heightinch (-1=.) (-2=.) (-3=.) (-4=.) (-5=.)
  87   gen height_in_inches = (heightft *12)+ heightinch
  88
  89   egen weightpd = rowtotal(YHEA_2200 YHEA_2300 YSAQ_000B)
  90   recode weightpd (-5/-1=.) (999=.)
  91
  92   by PUBID_1997 : gen poverty1997 =  CV_HH_POV_RATIO_[1]
  93   recode poverty1997 (-1=.) (-2=.) (-3=.) (-4=.) (-5=.)
  94   gen lnpoverty1997 = ln(poverty1997+1)
  95
  96
  97   clonevar msa= CV_MSA_
  98   recode msa (1=0) (2/4=1) (5=.) (-1=.) (-2=.) (-3=.) (-4=.) (-5=.)
  99   label define msa1 0 "Not in MSA" 1 "In MSA"
 100   label values msa msa1
 101
 102
 103   clonevar age = CV_AGE_INT_DATE_
 104   recode age (-5=.)
 105
 106   by PUBID_1997 : gen resid_mothereduc= CV_HGC_RES_MOM_[1]
 107   recode resid_mothereduc (-4=.) (-3=.) (95=.)
 108   gen mom_hsgrad = .
 109   replace mom_hsgrad=1 if resid_mothereduc >=12 & resid_mothereduc !=.
 110   replace mom_hsgrad=0 if resid_mothereduc <12 & resid_mothereduc !=.
 111   rename mom_hsgrad resmom_hsgrad
 112
 113
 114   by PUBID_1997 : gen resid_fathereduc= CV_HGC_RES_DAD_[1]
 115   recode resid_fathereduc (-4=.) (-3=.) (95=.)
 116   gen dad_hsgrad = .
 117   replace dad_hsgrad=1 if resid_fathereduc >=12 & resid_fathereduc !=.
 118   replace dad_hsgrad=0 if resid_fathereduc <12 & resid_fathereduc !=.
 119   rename dad_hsgrad resdad_hsgrad
 120
 121   by PUBID_1997 : gen mothereduc= CV_HGC_BIO_MOM_[1]
 122   recode mothereduc (-4=.) (-3=.) (95=.)
 123   gen mom_hsgrad = .
 124   replace mom_hsgrad=1 if mothereduc >=12 & mothereduc !=.
 125   replace mom_hsgrad=0 if mothereduc <12 & mothereduc !=.
 126
 127   by PUBID_1997 : gen fathereduc= CV_HGC_BIO_DAD_[1]
 128   recode fathereduc (-4=.) (-3=.) (95=.)
 129   gen dad_hsgrad = .
 130   replace dad_hsgrad=1 if fathereduc >=12 & fathereduc !=.
 131   replace dad_hsgrad=0 if fathereduc <12 & fathereduc !=.
 132
```

```
  gen biores_dad_hs = .
  replace biores_dad_hs = 1 if (resdad_hsgrad==1 | dad_hsgrad==1)
  replace biores_dad_hs = 0 if (resdad_hsgrad==0 & dad_hsgrad==0)|(resdad_hsgrad==. &
  dad_hsgrad==0)|(resdad_hsgrad==0 & dad_hsgrad==.)

  gen biores_mom_hs = .
  replace biores_mom_hs = 1 if (resmom_hsgrad==1 | mom_hsgrad==1)
  replace biores_mom_hs = 0 if (resmom_hsgrad==0 & mom_hsgrad==0)|(resmom_hsgrad==. &
  mom_hsgrad==0)|(resmom_hsgrad==0 & mom_hsgrad==.)

  merge 1:1 PUBID_1997 year using "D:\NLSY-Panel-1997-2011-v1-education.dta"

  clonevar educ_hga =  YSCH_2857_
  clonevar educ_hgc =  YSCH_3112_
  recode  educ_hga (-5=.) (-4=.) (-2=.) (-1=.) (95=.)
  recode  educ_hgc (-5=.) (-4=.) (-2=.) (-1=.) (95=.)
  rename _merge _mergeeducation


  by PUBID_1997 : gen educ_hgc2 = educ_hgc[11] if educ_hgc==.
  egen educ_hgc3 = rowtotal (educ_hgc educ_hgc2), missing

  gen hsgrad = .
  replace hsgrad=1 if educ_hgc3>=12 & educ_hgc3!=.
  replace hsgrad=0 if educ_hgc3<12 & educ_hgc3!=.


  gen avgmaleht = 69.5 if female ==0
  gen avgfemaleht = 64 if female ==1
  gen deviation_ht_male = height_in_inches - avgmaleht if female ==0
  gen deviation_ht_female = height_in_inches - avgfemaleht if female ==1
  egen deviation_ht = rowtotal (deviation_ht_male deviation_ht_female ), missing


  by PUBID_1997 : gen asvab_score= ASVAB_MATH_VERBAL_SCORE_PCT_[3]
  recode asvab_score (-5=.) (-4=.) (-2=.) (-1=.)
  gen lnasvab_score = ln(asvab_score +1)

  gen fedminwage = .
  replace fedminwage =5.15 if year>=1997 & year<2007
  replace fedminwage =5.85 if year==2007
  replace fedminwage =6.55 if year==2008
  replace fedminwage =7.25 if year>2008
  gen fedminwageadj = fedminwage * cpiadj

  gen recession=.
  replace recession =1 if year==2001
  replace recession =1 if year==2008
  replace recession =1 if year==2009
  replace recession =0 if recession !=1


  merge 1:1 PUBID_1997 year using "D:\Geocode-stateofresid-1997-2011.dta"

  gen health = YHEA_100_
  recode health (-1=.) (-2=.) (-3=.) (-4=.) (-5=.) (1=5) (2=4) (4=2) (5=1)

  rename _merge _mergestateofresid
  merge 1:1 PUBID_1997 year using "D:\NLSY-1997-2011-marital_spinc_children-panel.dta"
  rename _merge _mergemaritalspincchild


  /* Spousal income*/
  clonevar totspincome = YINC_2600_
  recode totspincome (-5=.) (-4=0) (-2=.) (-1=.)
  gen totspincomeadj = totspincome * cpiadj
  gen lntotspincadj = ln(totspincomeadj +1)

  clonevar num_biochild=CV_BIO_CHILD_HH_
  recode num_biochild (-5=.) (-4=.) (-3=.)
```

```
201
202   clonevar marital_status = YMAR_620_
203   recode marital_status (-5=.) (-4=.) (-2=.) (-1=.)
204   tab marital_status, gen(marital)
205
206   clonevar occupation_code =  YEMP_OCCODE_2002_01_
207   recode occupation_code (-5=.) (-4=.) (-3=.)
208   recode occupation_code (10/430=1) (500/950=2) (1000/1240=3) (1300/1530=4) (1540/1560=5) (
      1600/1760=6) (1800/1860=7) (1900/1960=8) (2000/2060=9) (2100/2150=10) (2200/2340=11) (2400/
      2550=12) (2600/2760=13) (2800/2960=14) (3000/3260=15) (3300/3650=16) (3700/3950=17) (4000/
      4160=18) (4200/4250=19) (4300/4430=20) (4460=21) (4500/4650=22) (4700/4960=23) (5000/5930=24
      ) (6000/6130=25) (6200/6940=26) (7000/7620=27) (7700/7750=28) (7800/7850=29) (7900/8960=30)
      (9000/9750=31) (9800/9840=32) (9950/9990=33)
209
210   merge m:1 stateofresid year using "D:\state_median_income.dta"
211   rename _merge _merge_statehhinc
212   gen deviation_hhinc = totincomeadj - statemedhhincadj
213
214   merge m:1 stateofresid year using "D:\state_shareofblack.dta"
215   rename _merge _merge_stateshareblack
216
217
218   *S1 Table. Pairwise Correlations and descriptive statistics*
219   summ lntotincadj age hsgrad health weightpd i.marital_status lnasvab_score lntotspincadj msa
         biores_dad_hs biores_mom_hs  fedminwageadj recession white color deviation_ht female if
      race_eth !=2 & race_eth !=3 & lntotincadj !=. & color !=. & deviation_ht!=. & hsgrad !=.  &
      health !=.  & lnasvab_score !=. & lntotspincadj !=. & marital_status !=.  & msa!=. &
      stateofresid !=. & occupation_code !=.  & biores_dad_hs  !=. & biores_mom_hs !=. & weightpd
      !=.
220   pwcorr lntotincadj age hsgrad health weightpd marital2 marital3 marital4 marital5
      lnasvab_score lntotspincadj msa   biores_dad_hs biores_mom_hs  fedminwageadj recession white
       color deviation_ht female if race_eth !=2 & race_eth !=3 & lntotincadj !=. & color !=. &
      deviation_ht!=. & hsgrad !=.  & health !=.  & lnasvab_score !=. & lntotspincadj !=. &
      marital_status !=.  & msa!=. &  stateofresid !=. & occupation_code !=.  & biores_dad_hs  !=.
       & biores_mom_hs !=. & weightpd!=., sig
221
222   * Table 1: Pooled OLS regression results*
223   * model 1
224   reg  lntotincadj female white age hsgrad health weightpd i.marital_status  lnasvab_score
      lntotspincadj msa   biores_dad_hs biores_mom_hs  fedminwageadj recession i.year  i.state3 i.
      occupation_code    if race_eth !=2 & race_eth !=3 & color!=. & deviation_ht!=. &
      stateofresid!=., robust
225
226   * model 2
227   reg  lntotincadj color  female white age hsgrad health weightpd i.marital_status
      lnasvab_score lntotspincadj msa   biores_dad_hs biores_mom_hs  fedminwageadj recession i.
      year  i.state3 i.occupation_code    if race_eth !=2 & race_eth !=3 & deviation_ht!=. &
      stateofresid!=., robust
228
229   * model 3
230   reg  lntotincadj color  deviation_ht  female white age hsgrad health weightpd i.
      marital_status  lnasvab_score lntotspincadj msa   biores_dad_hs biores_mom_hs  fedminwageadj
       recession i.year  i.state3 i.occupation_code    if race_eth !=2 & race_eth !=3 &
      stateofresid!=., robust
231
232   * model 4
233   reg  lntotincadj c.color##c.deviation_ht female white age hsgrad health weightpd i.
      marital_status  lnasvab_score lntotspincadj msa   biores_dad_hs biores_mom_hs  fedminwageadj
       recession i.year  i.state3 i.occupation_code    if race_eth !=2 & race_eth !=3 &
      stateofresid!=., robust
234
235   * model 5
236   reg  lntotincadj c.color##c.deviation_ht##female white age hsgrad health weightpd i.
      marital_status  lnasvab_score lntotspincadj msa   biores_dad_hs biores_mom_hs  fedminwageadj
       recession i.year  i.state3 i.occupation_code    if race_eth !=2 & race_eth !=3 &
      stateofresid!=., robust
237
238
239   * Table 2: Robustness Tests*
240   * deviation from state median household income as dependent variable
```

```
241   reg  deviation_hhinc  c.color##c.deviation_ht##female white age hsgrad health weightpd i.
      marital_status  lnasvab_score lntotspincadj msa   biores_dad_hs biores_mom_hs  fedminwageadj
       recession i.year  i.state3 i.occupation_code    if race_eth !=2 & race_eth !=3 &
      lntotincadj !=.& stateofresid!=., robust
242
243   * mixed effects model
244   mixed  lntotincadj c.color##c.deviation_ht##female age hsgrad health weightpd i.
      marital_status  lnasvab_score lntotspincadj msa   biores_dad_hs biores_mom_hs  fedminwageadj
       recession i.year  i.state3 i.occupation_code    if race_eth !=2 & race_eth !=3 &
      stateofresid!=. || race_eth:  || color:
245
246   * addition of share of Black population as additional control
247   reg  lntotincadj c.color##c.deviation_ht##female white age hsgrad health weightpd i.
      marital_status  lnasvab_score lntotspincadj msa   biores_dad_hs biores_mom_hs  fedminwageadj
       recession shareofblack i.year  i.state3 i.occupation_code    if race_eth !=2 & race_eth !=3
       & stateofresid!=., robust
248
249
250
251   * Table 3: Additional test with “blurred” categorical color line - Pooled OLS regression
      results*
252   * model 1
253   reg  lntotincadj lightblack medblack darkblack deviation_ht female age hsgrad health
      weightpd i.marital_status  lnasvab_score lntotspincadj msa   biores_dad_hs biores_mom_hs
      fedminwageadj recession i.year  i.state3 i.occupation_code    if race_eth !=2 & race_eth !=3
       & color!=.& stateofresid!=., robust
254
255   * model 2
256   reg  lntotincadj lightblack##c.deviation_ht medblack##c.deviation_ht darkblack##c.
      deviation_ht female age hsgrad health weightpd i.marital_status  lnasvab_score lntotspincadj
       msa   biores_dad_hs biores_mom_hs  fedminwageadj recession i.year  i.state3 i.
      occupation_code    if race_eth !=2 & race_eth !=3  & color!=. & stateofresid!=., robust
257
258   * model 3
259   reg  lntotincadj lightblack##c.deviation_ht##female medblack##c.deviation_ht##female
      darkblack##c.deviation_ht##female age hsgrad health weightpd i.marital_status  lnasvab_score
       lntotspincadj msa   biores_dad_hs biores_mom_hs  fedminwageadj recession i.year  i.state3 i
      .occupation_code    if race_eth !=2 & race_eth !=3  & color!=. & stateofresid!=., robust
260
261
262   * Table 4: Sub-group analysis by education, ability, and income*
263   * not a high school grad
264   reg  lntotincadj c.color##c.deviation_ht##female white age health weightpd i.marital_status
        lnasvab_score lntotspincadj msa   biores_dad_hs biores_mom_hs  fedminwageadj recession i.
      year  i.state3 i.occupation_code    if hsgrad==0 & race_eth !=2 & race_eth !=3 &
      stateofresid!=., robust
265
266   * high school grad
267   reg  lntotincadj c.color##c.deviation_ht##female white age health weightpd i.marital_status
        lnasvab_score lntotspincadj msa   biores_dad_hs biores_mom_hs  fedminwageadj recession i.
      year  i.state3 i.occupation_code    if hsgrad==1 & race_eth !=2 & race_eth !=3 &
      stateofresid!=., robust
268
269   * Dividing the ASVAB score into quartiles
270   xtile asvabquart = asvab_score   if race_eth !=2 & race_eth !=3 & lntotincadj !=. & color
      !=. & deviation_ht!=. & hsgrad !=.  & health !=.  & lnasvab_score !=. & lntotspincadj !=. &
      marital_status !=.  & msa!=. &  stateofresid !=. & occupation_code !=.  & biores_dad_hs  !=.
       & biores_mom_hs !=. & weightpd!=. & stateofresid!=., nq(4)
271
272   *xtile 1
273   reg  lntotincadj c.color##c.deviation_ht##female white age hsgrad health weightpd i.
      marital_status  lntotspincadj msa   biores_dad_hs biores_mom_hs  fedminwageadj recession i.
      year  i.state3 i.occupation_code    if asvabquart==1  & race_eth !=2 & race_eth !=3 &
      stateofresid!=., robust
274
275   *xtile 4
276   reg  lntotincadj c.color##c.deviation_ht##female white age hsgrad health weightpd i.
      marital_status  lntotspincadj msa   biores_dad_hs biores_mom_hs  fedminwageadj recession i.
      year  i.state3 i.occupation_code    if asvabquart==4  & race_eth !=2 & race_eth !=3 &
      stateofresid!=., robust
```

```
277
278   *75th percentile of income
279   reg  lntotincadj c.color##c.deviation_ht##female white age hsgrad health weightpd i.
      marital_status  lnasvab_score lntotspincadj msa   biores_dad_hs biores_mom_hs  fedminwageadj
       recession i.year  i.state3 i.occupation_code    if race_eth !=2 & race_eth !=3 &
      totincomeadj >=31246.6 & stateofresid!=., robust
280
281
282
283   * Figures - 1 and 2*
284   reg  lntotincadj c.color##c.deviation_ht##female age hsgrad health weightpd marital2
      lnasvab_score lntotspincadj  biores_dad_hs biores_mom_hs  fedminwageadj recession i.year  i.
      state3 i.occupation_code    if race_eth !=2 & race_eth !=3 & stateofresid!=., robust
285
286   * female
287   margins, at(deviation_ht= (-24(10)24) color = (2(3)8) female=(1)) plot
288
289   * male
290   margins, at(deviation_ht= (-24(10)24) color = (2(3)8) female=(0)) plot
291
292
293   log close
294
```