Replicating Sustainable youth employment quality management: The impact of robotization in China by Fucheng Liang and Yi Liu

Jesse Brandt, Alyssa Langmeyer, Veena Seshadri

2025-06-10

Load packages

```
library(tidyverse)
library(readxl)
library(readr)
library(ExPanDaR)
library(lme4)
library(miceadds)
```

Load the data (cleaned and truncated)

```
robots <- read_csv("../Data/robots_cleaned.csv")
```

```
Rows: 13453 Columns: 31
-- Column specification --------------------------------------------------------
Delimiter: ","
chr  (1): province
dbl (30): year, hhid, pline, id, Quality, Robot, wage, workhours, age, gende...

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Table 2: Summary Statistics.

```
#select relevant columns
selected_variables <- robots %>%
  select(Quality, Robot, age, gender, marriage, health, lnGDP, structure, lnaverwage)
#get summary statistics
mapply(FUN = \(x) c(N = length(x), mean = mean(x), sd = sd(x), min = min(x), max = max(x)),
       selected_variables) %>%
  t()
```

```
              N       mean          sd        min        max
Quality   13453 51.9224756 105.3437388  3.3333333 649.350649
Robot     13453  2.3991275   3.4982871  0.5541379  19.844507
age       13453 28.5806140   4.2631077 17.0000000  35.000000
gender    13453  0.5768230   0.4940813  0.0000000   1.000000
marriage  13453  0.5545975   0.4970286  0.0000000   1.000000
health    13453  0.9742808   0.1583021  0.0000000   1.000000
lnGDP     13453 10.2623826   0.8267938  7.8727714  11.586836
structure 13453  1.3824509   0.6984223  0.8522082   5.169242
lnaverwage 13453 11.2073749   0.2160024 10.9240482  12.024569
```

Table 3: Linear model.

Our results do not yet match theirs. Individual FE is unclear.

Mixed effects model (column 1)

```
me_model <- lmer(data = robots, formula = Quality ~ Robot + age + gender + marriage + health
     + structure + lnaverwage + as.factor(year)+ as.factor(province) + (1|id))
summary(me_model)
```

```
Linear mixed model fit by REML ['lmerMod']
Formula: Quality ~ Robot + age + gender + marriage + health + lnGDP +
    structure + lnaverwage + as.factor(year) + as.factor(province) +
    (1 | id)
   Data: robots
```

```
REML criterion at convergence: 149795.4

Scaled residuals:
    Min      1Q  Median      3Q     Max
-5.2585 -0.1821 -0.0361  0.1148  9.8198

Random effects:
 Groups   Name        Variance Std.Dev.
 id       (Intercept)  257.6   16.05
 Residual             3811.3   61.74
Number of obs: 13453, groups:  id, 13048

Fixed effects:
                                  Estimate Std. Error t value
(Intercept)                     -8604.6932   688.2510 -12.502
Robot                              -3.8255     0.5633  -6.792
age                                 0.6979     0.1609   4.336
gender                              6.9442     1.1263   6.165
marriage                           -0.8881     1.3819  -0.643
health                              7.3841     3.4884   2.117
lnGDP                              60.2886    13.5935   4.435
structure                          66.9559     7.5638   8.852
lnaverwage                        713.8052    65.8580  10.839
as.factor(year)2019                80.0076    11.7707   6.797
as.factor(province)Beijing       -687.5745    52.2414 -13.161
as.factor(province)Chongqing      -47.8950     9.8661  -4.855
as.factor(province)Fujian         -24.0620     4.8858  -4.925
as.factor(province)Gansu           57.1007    18.5116   3.085
as.factor(province)Guangdong     -202.2901    17.0136 -11.890
as.factor(province)Guangxi         24.3649     7.4032   3.291
as.factor(province)Guizhou        -46.1658    14.1461  -3.264
as.factor(province)Hainan         -42.9159    28.9746  -1.481
as.factor(province)Hebei           15.8545     6.1043   2.597
as.factor(province)Heilongjiang    58.7453    15.0106   3.914
as.factor(province)Henan           88.4917    15.1752   5.831
as.factor(province)Hubei          -27.1182     5.6220  -4.824
as.factor(province)Hunan           -7.5222     6.2960  -1.195
as.factor(province)Inner Mongolia  10.2496    10.5174   0.975
as.factor(province)Jiangsu       -137.1632    19.6367  -6.985
as.factor(province)Jiangxi         63.5134     6.9626   9.122
as.factor(province)Jilin           71.0707     9.6616   7.356
as.factor(province)Liaoning        33.4973     6.7561   4.958
as.factor(province)Ningxia         56.8929    31.1486   1.826
```

```
as.factor(province)Qinghai           23.8517    37.2808   0.640
as.factor(province)Shaanxi           21.3271     5.5129   3.869
as.factor(province)Shandong         -95.4285    12.8547  -7.424
as.factor(province)Shanghai        -537.7085    45.2211 -11.891
as.factor(province)Shanxi            76.9749     9.2641   8.309
as.factor(province)Sichuan          -74.3120     6.6920 -11.105
as.factor(province)Tianjin         -252.8187    27.4769  -9.201
as.factor(province)Yunnan           -44.0869    10.7465  -4.102
as.factor(province)Zhejiang        -177.2084    14.5148 -12.209


Correlation matrix not shown by default, as p = 38 > 12.
Use print(x, correlation=TRUE)  or
    vcov(x)        if you need it
```

Fixed effects model (column 2)

```
fe_model <- lm.cluster(data = robots, formula = Quality ~ Robot + age + gender + marriage + h
                       + structure + lnaverwage + as.factor(year) + as.factor(province),
                       cluster = "id")
```

```
Loading required namespace: sandwich
```

```
summary(fe_model)
```

```
R^2= 0.63415

                                      Estimate   Std. Error    t value
(Intercept)                     -8620.6252003 1813.8072265 -4.7527792
Robot                              -3.8171637    1.3484827 -2.8307101
age                                 0.6989876    0.1591414  4.3922422
gender                              6.9856033    1.0382601  6.7281821
marriage                           -0.9062716    1.3545836 -0.6690407
health                              7.4792741    3.8457341  1.9448235
lnGDP                              60.2330785   30.4254062  1.9796968
structure                          66.6028212   19.8915604  3.3482955
lnaverwage                        715.2955378  171.3881732  4.1735408
as.factor(year)2019                79.8049820   30.6506040  2.6037001
as.factor(province)Beijing       -687.0428949  135.4882279 -5.0708678
```

as.factor(province)Chongqing	-47.9839975	21.2443447	-2.2586716
as.factor(province)Fujian	-23.9231023	7.1833913	-3.3303354
as.factor(province)Gansu	57.4032589	39.9495824	1.4368926
as.factor(province)Guangdong	-202.2594181	41.0144712	-4.9314160
as.factor(province)Guangxi	24.8132167	12.5388306	1.9789099
as.factor(province)Guizhou	-46.0016580	31.2279508	-1.4730924
as.factor(province)Hainan	-42.4734188	65.3901324	-0.6495387
as.factor(province)Hebei	16.0768812	11.0963291	1.4488468
as.factor(province)Heilongjiang	59.5886896	35.8777184	1.6608829
as.factor(province)Henan	88.8770984	36.0922030	2.4625013
as.factor(province)Hubei	-26.9179671	9.1754481	-2.9336951
as.factor(province)Hunan	-7.1855497	11.5081881	-0.6243858
as.factor(province)Inner Mongolia	10.3090860	20.3616418	0.5062993
as.factor(province)Jiangsu	-137.3900135	48.5560997	-2.8295109
as.factor(province)Jiangxi	63.6988212	11.7681722	5.4128050
as.factor(province)Jilin	71.2446431	18.4929666	3.8525265
as.factor(province)Liaoning	33.9691070	13.7684284	2.4671739
as.factor(province)Ningxia	56.8598084	68.5704193	0.8292177
as.factor(province)Qinghai	23.8332566	83.6703109	0.2848472
as.factor(province)Shaanxi	21.3805143	7.3464556	2.9103169
as.factor(province)Shandong	-95.2878400	28.6242821	-3.3289163
as.factor(province)Shanghai	-537.6056329	120.3142511	-4.4683454
as.factor(province)Shanxi	77.3796375	19.2151925	4.0270030
as.factor(province)Sichuan	-74.2003997	14.3043603	-5.1872575
as.factor(province)Tianjin	-253.0439632	68.9790143	-3.6684195
as.factor(province)Yunnan	-44.0282438	22.7034732	-1.9392735
as.factor(province)Zhejiang	-177.1215701	36.7959089	-4.8136213

	Pr(>|t|)
(Intercept)	2.006395e-06
Robot	4.644479e-03
age	1.121876e-05
gender	1.717959e-11
marriage	5.034695e-01
health	5.179622e-02
lnGDP	4.773761e-02
structure	8.131028e-04
lnaverwage	2.999019e-05
as.factor(year)2019	9.222340e-03
as.factor(province)Beijing	3.960057e-07
as.factor(province)Chongqing	2.390382e-02
as.factor(province)Fujian	8.674143e-04
as.factor(province)Gansu	1.507485e-01
as.factor(province)Guangdong	8.163569e-07

```
as.factor(province)Guangxi         4.782615e-02
as.factor(province)Guizhou         1.407261e-01
as.factor(province)Hainan          5.159903e-01
as.factor(province)Hebei           1.473804e-01
as.factor(province)Heilongjiang    9.673696e-02
as.factor(province)Henan           1.379717e-02
as.factor(province)Hubei           3.349531e-03
as.factor(province)Hunan           5.323742e-01
as.factor(province)Inner Mongolia 6.126465e-01
as.factor(province)Jiangsu         4.661921e-03
as.factor(province)Jiangxi         6.204504e-08
as.factor(province)Jilin           1.169053e-04
as.factor(province)Liaoning        1.361842e-02
as.factor(province)Ningxia         4.069812e-01
as.factor(province)Qinghai         7.757612e-01
as.factor(province)Shaanxi         3.610624e-03
as.factor(province)Shandong        8.718459e-04
as.factor(province)Shanghai        7.882694e-06
as.factor(province)Shanxi          5.649232e-05
as.factor(province)Sichuan         2.134136e-07
as.factor(province)Tianjin         2.440545e-04
as.factor(province)Yunnan          5.246804e-02
as.factor(province)Zhejiang        1.482196e-06
```

```r
model1 <- lm(data = robots, formula = Quality ~ Robot + age + gender + marriage + health + lr
   + structure + lnaverwage + as.factor(year) + as.factor(province))
model2 <- lm(data = robots, formula = Quality ~ Robot + age + gender + marriage + health + lr

model.1.summary <- summary(model1)
model.2.summary <- summary(model2)

#model including year FE and region FE
model.1.summary
#model 1 with province and year FE not shown
model.1.summary$coefficients[1:9,1:4]
#model without year FE and region FE
model.2.summary

#model including id - filtered to include only ids with more than one observation
duplicate_ids <- robots %>%
  group_by(id) %>%
  mutate(count = length(id)) %>%
```

```
  filter(count>1)

fe_model3 <- lm(data = duplicate_ids, formula = Quality ~ Robot + age + gender + marriage + h
                + structure + lnaverwage + as.factor(year) + as.factor(province) + as.factor

#Next model takes a long time to run. Is it equivalent to fe_model3?
start_time <- Sys.time()
# fe_model4 <- lm(data = robots, formula = Quality ~ Robot + age + gender + marriage + healt
#                 + structure + lnaverwage + as.factor(year) + as.factor(province) + as.facto
end_time <- Sys.time()

elapsed_time <- end_time-start_time

cat("Time to run model:", elapsed_time)
model.3.summary <- summary(fe_model3)
model.3.summary$coefficients[1:9,1:4]
```

Random effects model (column 3)

Note: in the paper, column 3 is the same as column 1, but the text indicates that it should be the random effects model.

```
re_model <- lmer(data = robots, formula = Quality ~ Robot + age + gender + marriage + health
summary(re_model)
```

```
Linear mixed model fit by REML ['lmerMod']
Formula: Quality ~ Robot + age + gender + marriage + health + lnGDP +
    structure + lnaverwage + (1 | year) + (1 | province) + (1 |      id)
   Data: robots

REML criterion at convergence: 150175.7

Scaled residuals:
    Min      1Q  Median      3Q     Max
-5.1468 -0.1832 -0.0342  0.1156  9.8156

Random effects:
 Groups   Name        Variance Std.Dev.
 id       (Intercept)   257.6   16.05
 province (Intercept) 22576.7  150.26
 year     (Intercept)  5474.8   73.99
```

```
 Residual              3812.5   61.75
Number of obs: 13453, groups:  id, 13048; province, 29; year, 2

Fixed effects:
              Estimate Std. Error t value
(Intercept) -7215.5171   637.4763 -11.319
Robot          -4.0014     0.5605  -7.140
age             0.6906     0.1610   4.290
gender          6.9675     1.1265   6.185
marriage       -0.8925     1.3820  -0.646
health          7.3161     3.4889   2.097
lnGDP          62.1642    12.5228   4.964
structure      62.8623     7.4621   8.424
lnaverwage    587.3647    59.4018   9.888

Correlation of Fixed Effects:
           (Intr) Robot  age    gender marrig health lnGDP  strctr
Robot      -0.179
age        -0.016 -0.008
gender     -0.001  0.014 -0.058
marriage    0.001 -0.001 -0.575  0.071
health     -0.010 -0.006  0.020 -0.001  0.005
lnGDP       0.171 -0.007 -0.011  0.000  0.017  0.009
structure   0.103 -0.241  0.017 -0.018  0.003  0.004  0.015
lnaverwage -0.978  0.175  0.011  0.000 -0.002  0.002 -0.351 -0.117
```

Table 4: Results of endogeneity test and robustness test.

Table 5: Heterogeneity analysis results.

```r
#Gender
man <- filter(robots, gender == 1)
woman <- filter(robots, gender ==0)

#Education
unique(robots$education)
count(robots, education)
690 + 4763 # 0 and 1 are primary
# 2 is middle
# 3 is university
```

```r
primary <- filter(robots, education == 0 | education ==1)
middle <- filter(robots, education == 2)
university <- filter(robots, education ==3)

#Age groups
age_1 <- filter(robots, age >= 17 & age <= 25)
age_2 <- filter(robots, age >= 26 & age <= 35)

# Regional employment
large <- filter(robots, reg_labor > mean(reg_labor))
small <- filter(robots, reg_labor <= mean(reg_labor))

#temp


fit_fe_model <- function(dataset){
  model <- lm.cluster(data = dataset, formula = Quality ~ Robot + age + gender + marriage + l
             + structure + lnaverwage + as.factor(year) + as.factor(province),
             cluster = "id")
  model_summary <- summary(model)
  return(model_summary[2,1:4])
}

fit_fe_model(large)
tbl <- map(.x = list(man,woman,primary,middle,university,large,small), .f = fit_fe_model) %>%
  data.frame()
names(tbl) <- c("Man", "Woman", "Primary", "Middle", "University", "Large", "Small")

rownames(tbl)[1] <- "Estimate (Robot)"
```

```r
tbl
```

```
                          Man      Woman     Primary      Middle    University
Estimate (Robot) -4.793580808 -0.6970551 -2.03857422 -4.44314393 -1.125973e+01
Std. Error        1.619199830  1.8647680  1.23102065  1.77098300  2.720375e+00
t value          -2.960462767 -0.3738026 -1.65600327 -2.50885747 -4.139036e+00
Pr(>|t|)          0.003071772  0.7085512  0.09772116  0.01211223  3.487684e-05
                     Large         Small
Estimate (Robot) 1.4763263 -27.255197402
Std. Error       1.5054416   9.955646390
t value          0.9806600  -2.737662261
Pr(>|t|)         0.3267605   0.006187759
```

Table 6: Results of mechanism analysis.

Table 7: Results of regulatory effect analysis.

Table 8: Results of he effectiveness nalysis of employment promotion measures.