MLaE: Whether WFH affect well-being

Boyu Chen, Yihuai Chang

2023-05-29

Research Motivation

- Under the pandemic, there were many inconveniences.
- We try to figure out how working from home affect workers' well-being.
- Our assumption is that working from home has negative impact on workers' well-being, since they might feel socially isolated.

Literature

- Marco Bertoni,Danilo Cavapozzi et al.*(2022), “Remote Working and Mental Health During the First Wave of the COVID-19 Pandemic”

Data Source

IPUMS Time Use 2021

- D_i: Distance working binary variable
- Y_i: Well-being ladder (0-10)
- X_i: Control variables, including statefip, age, gender, have kids, occupation, earning per week, race, fullpart.
- # of observations: 3,281, # of variables: ???

Assumptions

- We assume that unconfoundedness is satisfied, which is:

$$(Y_{i(0)}, Y_{i(1)}) \perp D_i | X_i$$

- The sparsity assumption holds

Model

XPOPOSSION

Cross-fit partialing-out lasso Poisson regression, the model is:

$$E(y|D, X) = exp(D\alpha + X^T\beta)$$

where

- y is the dep. variable.
- D is treatment, which is a scalar.
- X is the control variable matrix, which is a $n \times p$ matrix.
- β is a $p \times 1$ vector.

XPOLPR algorithm

Step 1

Randomly Partition the sample to K folds.

Step 2

Define two sets:

- I_k: the obs. in fold k
- IC_k: the obs. not in fold k

XPOLPR algorithm

Step 3

Run Double Selection poisson lasso For $k = 1, ..., K$

1. Run poisson lasso for the following model

$$y = exp(D\alpha_k + X'\beta_k)$$

   and we get the non-zero covariates, denoted by $\tilde{X}_{k,y}$.

2. Run poisson regression for the following model

$$y = exp(D\alpha_k + \tilde{X}'_{k,y}\beta_k)$$

   and we get the estimated coefficients $\tilde{\alpha}_k$ and $\tilde{\delta}_k$.

XPOLPR algorithm

3. For the obs. $i \in I_k$, fill in the prediction for the high-dimensional component using the out-of-sample estimate $\tilde{\delta}_k$.

$$\tilde{s}_i = \tilde{X}'_{k,y,i}\tilde{\delta}_k$$

4. Using the observations $i \in IC_k$, perform a linear lasso of D on X using observation-level weights, w_i.

$$w_i = exp'(D\tilde{\alpha}_k + \tilde{s}_i)$$

5.

Run OLS

Consider the following model:

$$Y_i = \tau_{ols} D_i + X_i^T \beta + \epsilon_i, \forall i = 1, ..., n$$

Where τ_{ols}, D_i are scalar, X_i is a $k \times 1$ vector, β is a $k \times 1$ vector.

Run DML

Given

$$\xi_0(X_i) = E(Y_i|X_i)$$
$$m_0(X_i) = E(D_i|X_i) = Pr(D_i = 1|X_i)$$

We consider the following model:

$$Y_i - \xi_0(X_i) = \tau_{dml}(D_i - m_0(X_i)) + u_i$$

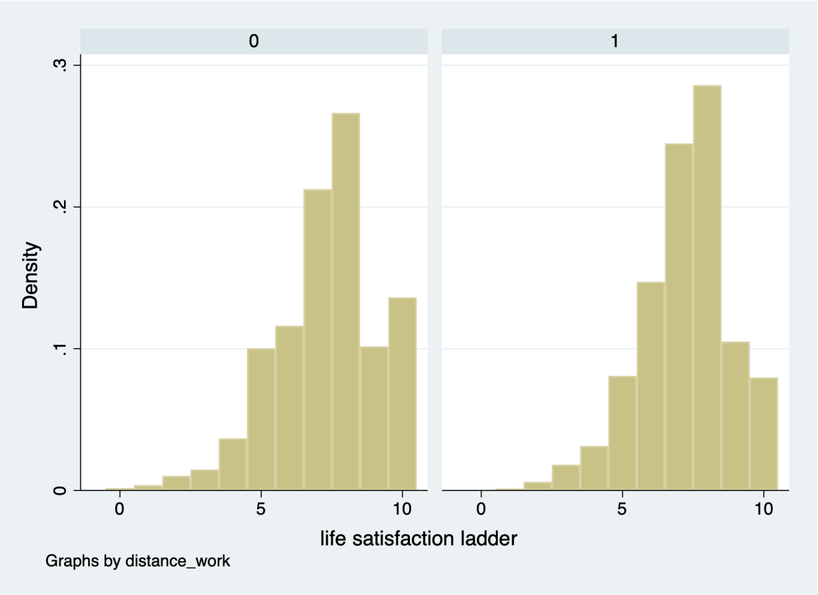
- First, we run Lasso (Y on X) to get $\hat{\xi}_0(X_i)$ and run Lasso (D on X) to get $\hat{m}_0(X_i)$
- Second, we regress $Y_i - \hat{\xi}_0(X_i)$ on $(D_i - \hat{m}_0(X_i))$ to get $\hat{\tau}_{dml}$
- $\hat{\tau}_{dml}$ is the treatment effect we want.

Preliminary analysis

Descriptive Statistics

variables	Remote Working	No Remote Working	difference
age	43.589	44.493	-0.905*
	(12.253)	(13.958)	[0.544]
female	0.490	0.485	0.005
	(0.500)	(0.500)	[0.020]
household_numkids	0.772	0.817	-0.045
	(0.971)	(1.168)	[0.045]
earnweek	1,727.669	1,124.934	602.736***
	(801.828)	(729.871)	[30.079]
workinghours	42.501	40.878	1.623***
	(8.812)	(11.633)	[0.441]
wbladder	7.222	7.308	-0.086
	(1.616)	(1.832)	[0.072]
N	829	2,452	

Graph


0
1
Density
0
.1
.2
.3
0
5
10
0
5
10
life satisfaction ladder
Graphs by distance_work

Estimation Result: OLS

```
. reg wbladder $D $X, r
```

Linear regression

Number of obs	=	3,281
F(120, 3157)	=	.
Prob > F	=	.
R-squared	=	0.0914
Root MSE	=	1.7293

wbladder	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
distance_work	-.1469358	.0764536	-1.92	0.055	-.2968396	.002968
statefip						
alaska	.9637804	.4090905	2.36	0.019	.1616702	1.765891
arizona	-.2536854	.3155871	-0.80	0.422	-.872462	.3650911
arkansas	-.3382687	.3595688	-0.94	0.347	-1.043281	.3667435
california	-.253545	.2388851	-1.06	0.289	-.7219307	.2148407
colorado	-.1503513	.3326177	-0.45	0.651	-.8025201	.5018174
connecticut	-.448235	.3844557	-1.17	0.244	-1.202043	.3055734
delaware	.2846163	.394298	0.72	0.470	-.4884899	1.057723
district of columbia	-.4508386	.4869128	-0.93	0.355	-1.405536	.503859
florida	-.3995233	.2618223	-1.53	0.127	-.9128825	.1138358
georgia	-.3572428	.3023941	-1.18	0.238	-.9501517	.2356661
hawaii	-.8010099	.7557691	-1.06	0.289	-2.282858	.6808385

Estimation Result: DDML

```
DDML estimation results:
spec  r     Y learner     D learner         b         SE
 opt  1     Y2_rlasso        D1_reg    -0.173    ( 0.079)
 opt  2     Y2_rlasso        D1_reg    -0.154    ( 0.079)
 opt  3     Y2_rlasso        D1_reg    -0.169    ( 0.079)
 opt  4     Y2_rlasso        D1_reg    -0.181    ( 0.080)
 opt  5     Y2_rlasso        D1_reg    -0.179    ( 0.079)
opt = minimum MSE specification for that resample.

Mean/med.   Y learner     D learner         b         SE
 mse mn     [min-mse]         [mse]    -0.171    ( 0.080)
 mse md     [min-mse]         [mse]    -0.173    ( 0.079)

Median over min-mse specifications
y-E[y|X]   = Y2_rlasso                          Number of obs     =      3281
D-E[D|X,Z]= D1_reg
```

wbladder	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
distance_work	-.1729803	.0792368	-2.18	0.029	-.3282816	-.017679

```
Summary over 5 resamples:
       D eqn      mean       min       p25       p50       p75       max
distance_work   -0.1712   -0.1811   -0.1789   -0.1730   -0.1691   -0.1538
```

Conclusion

- It seems that working from home will reduce the distance workers' well-being.
- We will further examine whether if distance workers' exercise time, sleep time and social time are significantly different to control group.