

# MLaE: Whether WFH affect well-being

Boyu Chen, Yihuai Chang

2023-05-29

# Research Motivation

- ▶ During the COVID-19 pandemic, there are a lot of workers were asked to work from home because of lock-down.
- ▶ The question we want to examine is:  
During the pandemic, is the **well-being index** of home workers lower?
- ▶ Our hypothesis 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, which is cross-sectional individual-level survey data.

- ▶ We dropped all the observations do not answer whether they are working from home or not, as well as excluded those observations with no job and no income.
- ▶  $Y_i$  represents an outcome of interest:
  - ▶ Well-being index: 0~10, 0 for the worst life, 10 for the best life
- ▶  $D_i$  represents a binary treatment of interest:
  - ▶ Distance working or not : 1 for distance worker, 0 for commuter
- ▶  $X_i$  represents the a set of control variables:
  - ▶ Occupation, marital status, age, race, sex, have child or not, full-time or part-time, earning per week, State
- ▶ # of observations: 3,281, # of variables: 94

# 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

# Model

## Poisson regression

Because our  $Y$  is a count data, we use poisson regression with double machine learning to specify our treatment effect.

## Recall: Poisson

- ▶ poisson pdf:
  - ▶ If  $Y \sim \text{poisson}(\lambda)$ , then  $f(y) = \frac{\lambda^y e^{-\lambda}}{y!}$
- ▶ poisson regression: let  $\lambda_i = E(y_i|X_i) = \exp(X_i' \beta)$  ,
  - ▶ The conditional pdf is  $f(y_i|X_i) = \frac{\exp(X_i' \beta)^{y_i} e^{-\exp(X_i' \beta)}}{y_i!}$
  - ▶ The log-likelihood is
$$\ell(\beta|y_i, X_i) = y_i(X_i' \beta) - \exp(-X_i' \beta) - \ln(y_i!)$$
  - ▶ The poisson regression LASSO criterion is

$$\min_{\beta, \gamma} Q(\beta, \gamma|X, Y) = -n^{-1} \sum_{i=1}^n \ell(\beta|y_i, X_i) + \gamma \sum_{j=1}^p |\beta_j|$$

# XPOPOSSION

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

$$E(y_i|D, X) = \exp(D_i\alpha + X_i'\beta)$$

where

- ▶  $y_i$  is the dep. variable.
- ▶  $D_i$  is treatment, which is a scalar.
- ▶  $X_i$  is the control variable vector, which is a  $p \times 1$  vector.
- ▶  $\beta$  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  $k = 1, \dots, K$

1. Using all  $i \in IC_k$ , run poisson lasso for the following model

$$y_i = \exp(D_i\alpha_k + X_i'\beta_k)$$

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

2. Using all  $i \in IC_k$ , run poisson regression for the following model

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

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,  $w_i$ .

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

Denote the selected controls by  $\tilde{X}_k$ .

## XPOLPR algorithm

5. Using the observations  $i \in IC_k$ , fit a linear regression of  $D$  on  $\tilde{X}_k$ , and denote the coefficient estimates by  $\hat{\gamma}_k$ .
6. For each observation  $i \in I_k$ , fill in the instrument

$$z_i = D_i - \tilde{X}_{k,i} \hat{\gamma}'_k$$

# XPOLPR algorithm

## Step 4

Compute the point estimates  $\hat{\alpha}$  by solving the following sample-moment equations.

$$\frac{1}{n} \sum_{i=1}^n \{y_i - \exp(D_i \alpha' + \tilde{\epsilon}_i)\} z_i = 0$$

# XPOLPR algorithm

## Step 5

Variance estimation is estimated by

$$\hat{Var}(\hat{\alpha}) = n^{-1} \hat{J}_0^{-1} \hat{\Psi} (\hat{J}_0^{-1})'$$

where

$$\hat{\Psi} = K^{-1} \sum_{k=1}^K \hat{\Psi}_k$$

$$\hat{\Psi}_k = n_k^{-1} \sum_{i \in I_k} \hat{\psi}_i \hat{\psi}_i'$$

$$\hat{\psi}_i = \{y_i - \exp(d\hat{\alpha} + \hat{s}_i)\} z_i$$

$$\hat{J}_0 = K^{-1} \sum_{k=1}^K (n_k^{-1} \sum_{i \in I_k} \hat{\psi}_i^{\alpha})$$

$$\hat{\psi}_i^{\alpha} = \frac{\partial \hat{\psi}_i}{\partial \hat{\alpha}}$$

analysis

# Descriptive Statistics

	mean	standard deviation
well being	7.286	1.780
WFH	0.253	0.435
age	44.265	13.552
female	0.486	0.500
have child	0.437	0.496
married	0.527	0.499
earnings per week	1277.225	793.090
fulltime job	1.134	0.341
observations	3281	



# Main Result

Cross-fit fold 10 of 10 ...  
Estimating lasso for wbladder using plugin  
Estimating lasso for distance\_work using plugin

Cross-fit partialing-out	Number of obs	=	3,281
Poisson model	Number of controls	=	94
	Number of selected controls	=	20
	Number of folds in cross-fit	=	10
	Number of resamples	=	1
	Wald chi2(1)	=	3.80
	Prob > chi2	=	0.0513

wbladder	Robust				
	IRR	Std. Err.	z	P> z	[95% Conf. Interval]
distance_work	.9794901	.0104156	-1.95	0.051	.9592871 1.000119

Note: Chi-squared test is a Wald test of the coefficients of the variables of interest jointly equal to zero. Lassos [select controls](#) for model estimation. Type [lassoinfo](#) to see number of selected variables in each lasso.

## Subgroup: gender

male:

Cross-fit fold 10 of 10 ...

Estimating lasso for wbladder using plugin

note: female dropped because it is constant

Estimating lasso for distance\_work using plugin

note: female dropped because it is constant

Cross-fit partialing-out	Number of obs	=	1,686
Poisson model	Number of controls	=	94
	Number of selected controls	=	21
	Number of folds in cross-fit	=	10
	Number of resamples	=	1
	Wald chi2(1)	=	4.33
	Prob > chi2	=	0.0375

wbladder	Robust					
	IRR	Std. Err.	z	P> z	[95% Conf. Interval]	
distance_work	.9677906	.0152292	-2.08	0.037	.9383974	.9981044

Note: Chi-squared test is a Wald test of the coefficients of the variables of interest jointly equal to zero. Lassos [select controls](#) for model estimation. Type [lassoinfo](#) to see number of selected variables in each lasso.

## Subgroup: gender

female:

Cross-fit fold 10 of 10 ...

Estimating lasso for wbladder using plugin

note: female dropped because it is constant

Estimating lasso for distance\_work using plugin

note: female dropped because it is constant

Cross-fit partialing-out	Number of obs	=	1,595
Poisson model	Number of controls	=	94
	Number of selected controls	=	17
	Number of folds in cross-fit	=	10
	Number of resamples	=	1
	Wald chi2(1)	=	0.35
	Prob > chi2	=	0.5548

wbladder	IRR	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
distance_work	.9914048	.0144921	-0.59	0.555	.9634039	1.02022

Note: Chi-squared test is a Wald test of the coefficients of the variables of interest jointly equal to zero. Lassos [select controls](#) for model estimation. Type [lassoinfo](#) to see number of selected variables in each lasso.

## Subgroup: have children or not

have children:

Cross-fit fold 10 of 10 ...

Estimating lasso for wbladder using plugin

note: hh\_child dropped because it is constant

Estimating lasso for distance\_work using plugin

note: hh\_child dropped because it is constant

Cross-fit partialing-out	Number of obs	=	1,433
Poisson model	Number of controls	=	94
	Number of selected controls	=	13
	Number of folds in cross-fit	=	10
	Number of resamples	=	1
	Wald chi2(1)	=	0.33
	Prob > chi2	=	0.5628

wbladder	Robust					[95% Conf. Interval]
	IRR	Std. Err.	z	P> z		
distance_work	.9914021	.0147933	-0.58	0.563	.9628277	1.020825

Note: Chi-squared test is a Wald test of the coefficients of the variables of interest jointly equal to zero. Lassos `select controls` for model estimation. Type `lassoinfo` to see number of selected variables in each lasso.

## Subgroup: have children or not

do not have any child:

Cross-fit fold 10 of 10 ...

Estimating lasso for wbladder using plugin

note: hh\_child dropped because it is constant

Estimating lasso for distance\_work using plugin

note: hh\_child dropped because it is constant

Cross-fit partialing-out	Number of obs	=	1,848
Poisson model	Number of controls	=	94
	Number of selected controls	=	18
	Number of folds in cross-fit	=	10
	Number of resamples	=	1
	Wald chi2(1)	=	4.24
	Prob > chi2	=	0.0396

wbladder	Robust					
	IRR	Std. Err.	z	P> z	[95% Conf. Interval]	
distance_work	<b>.9700444</b>	<b>.0143347</b>	<b>-2.06</b>	<b>0.040</b>	<b>.9423518</b>	<b>.9985507</b>

Note: Chi-squared test is a Wald test of the coefficients of the variables of interest jointly equal to zero. Lassos [select controls](#) for model estimation. Type [lassoinfo](#) to see number of selected variables in each lasso.

## Subgroup: marital status

married:

Cross-fit fold 10 of 10 ...

Estimating lasso for wbladder using plugin

Estimating lasso for distance\_work using plugin

Cross-fit partialing-out	Number of obs	=	1,728
Poisson model	Number of controls	=	94
	Number of selected controls	=	18
	Number of folds in cross-fit	=	10
	Number of resamples	=	1
	Wald chi2(1)	=	1.01
	Prob > chi2	=	0.3161

wbladder	Robust					
	IRR	Std. Err.	z	P> z	[95% Conf. Interval]	
distance_work	.9880929	.0118057	-1.00	0.316	.9652229	1.011505

Note: Chi-squared test is a Wald test of the coefficients of the variables of interest jointly equal to zero. Lassos `select controls` for model estimation. Type `lassoinfo` to see number of selected variables in each lasso.

## Subgroup: marital status

not married:

Cross-fit fold 10 of 10 ...

Estimating lasso for wbladder using plugin

Estimating lasso for distance\_work using plugin

Cross-fit partialing-out	Number of obs	=	1,553
Poisson model	Number of controls	=	94
	Number of selected controls	=	17
	Number of folds in cross-fit	=	10
	Number of resamples	=	1
	Wald chi2(1)	=	4.03
	Prob > chi2	=	0.0447

wbladder	Robust				
	IRR	Std. Err.	z	P> z	[95% Conf. Interval]
distance_work	.9644475	.0173882	-2.01	0.045	.9309622 .9991371

Note: Chi-squared test is a Wald test of the coefficients of the variables of interest jointly equal to zero. Lassos [select controls](#) for model estimation. Type [lassoinfo](#) to see number of selected variables in each lasso.

# Subgroup: single and married men

## single men:

Cross-fit fold 10 of 10 ...

Estimating lasso for wbladder using plugin

note: female dropped because it is constant

Estimating lasso for distance\_work using plugin

note: female dropped because it is constant

Cross-fit partialing-out	Number of obs	=	735
Poisson model	Number of controls	=	94
	Number of selected controls	=	18
	Number of folds in cross-fit	=	10
	Number of resamples	=	1
	Wald chi2(1)	=	8.06
	Prob > chi2	=	0.0045

wbladder	Robust					
	IRR	Std. Err.	z	P> z	[95% Conf. Interval]	
distance_work	.9219021	.0264072	-2.84	0.005	.8715709	.9751398

Note: Chi-squared test is a Wald test of the coefficients of the variables of interest jointly equal to zero. Lassos [select controls](#) for model estimation. Type [lassoinfo](#) to see number of selected variables in each lasso.



## Subgroup: single and married men

### married men:

Cross-fit fold 10 of 10 ...

Estimating lasso for wbladder using plugin

note: female dropped because it is constant

Estimating lasso for distance\_work using plugin

note: female dropped because it is constant

Cross-fit partialing-out	Number of obs	=	<b>951</b>
Poisson model	Number of controls	=	<b>94</b>
	Number of selected controls	=	<b>19</b>
	Number of folds in cross-fit	=	<b>10</b>
	Number of resamples	=	<b>1</b>
	Wald chi2(1)	=	<b>0.61</b>
	Prob > chi2	=	<b>0.4349</b>

wbladder	Robust				
	IRR	Std. Err.	z	P> z	[95% Conf. Interval]
distance_work	<b>.9877181</b>	<b>.0156312</b>	<b>-0.78</b>	<b>0.435</b>	<b>.9575519 1.018835</b>

Note: Chi-squared test is a Wald test of the coefficients of the variables of interest jointly equal to zero. Lassos [select controls](#) for model estimation. Type [lassoinfo](#) to see number of selected variables in each lasso.

# Robustness Check: PSM

Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
wbladder	Unmatched	<b>7.21980676</b>	<b>7.31010309</b>	<b>-.090296329</b>	<b>.071507743</b>	<b>-1.26</b>
	ATT	<b>7.21980676</b>	<b>7.40398551</b>	<b>-.184178744</b>	<b>.103629524</b>	<b>-1.78</b>

Note: S.E. does not take into account that the propensity score is estimated.

psmatch2: Treatment assignment	psmatch2: Common support On support	Total
Untreated	<b>2,425</b>	<b>2,425</b>
Treated	<b>828</b>	<b>828</b>
Total	<b>3,253</b>	<b>3,253</b>

# Robustness Check: DML

```
. qddml wbladder $D ($X), kfolds(2) model(partial) cmd(rlasso) reps(5)  
minimum Python version required is 2.7
```

DDML estimation results:

spec	r	Y learner	D learner	b	SE
opt 1	1	Y2_rlasso	D1_reg	-0.115	( 0.081)
opt 2	2	Y2_rlasso	D1_reg	-0.135	( 0.081)
opt 3	3	Y2_rlasso	D1_reg	-0.127	( 0.081)
opt 4	4	Y2_rlasso	D1_reg	-0.095	( 0.081)
opt 5	5	Y2_rlasso	D1_reg	-0.110	( 0.081)

opt = minimum MSE specification for that resample.

Mean/med.	Y learner	D learner	b	SE
mse mn	[min-mse]	[mse]	-0.117	( 0.082)
mse md	[min-mse]	[mse]	-0.115	( 0.082)

Median over min-mse specifications

$y-E[y|X] = \text{Y2\_rlasso}$

Number of obs = 3281

$D-E[D|X,Z] = \text{D1\_reg}$

wbladder	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
distance_work	-.1151211	.0822525	-1.40	0.162	-.2763332	.0460909

Summary over 5 resamples:

D eqn	mean	min	p25	p50	p75	max
distance_work	-0.1166	-0.1352	-0.1274	-0.1151	-0.1098	-0.0953

## Conclusion

- ▶ Our main finding indicates that working from home slightly decreases workers' well-being by 0.98%, but this decrease is not statistically significant at a 5% significance level. However, it is approaching statistical significance.
- ▶ Working from home has a negative impact on well-being for men, but not for women.
- ▶ Additionally, married individuals reported higher levels of well-being compared to unmarried individuals.
- ▶ Specifically, working from home decreases the well-being of single men by 0.92%.

# Limitation

- ▶ Using survey data may cause some problems. All the data in this study was self-reported.
- ▶ Our sample size is rather small.