European Commission

Case Studies in Microeconomic Evaluation
Data and methods for learning what works

Case Study 2

Competence Centre on Microeconomic Evaluation (CC-ME)

Joint Research Centre,

23rd November 2022

Joint Research Centre

Outline

- A brief **Introduction** to PSM
- **Case Study**: Estimating the causal impact of labour market policies on mental-health responses to employment shocks
- **APPENDIX**: Implementing the Matching with STATA

A Brief Introduction to Matching

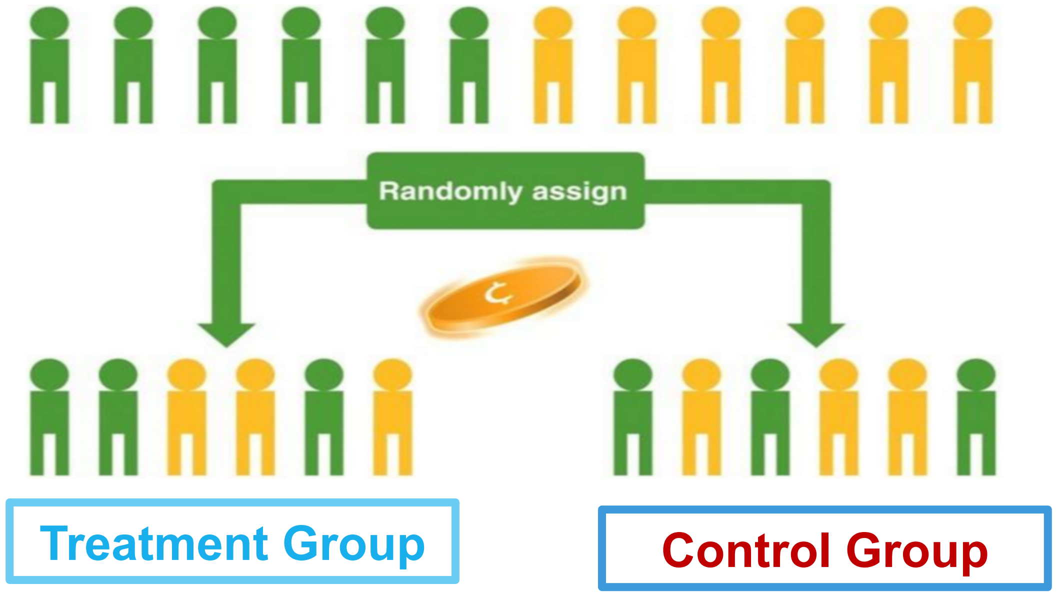
The propensity score matching (PSM) is a **quasi-experimental method.**

- The researcher uses statistical techniques to construct an "artificial" control group by **matching** each treated unit with a non-treated unit of similar characteristics to estimate the impact of an intervention.
- Matching is a useful method in data analysis for estimating the impact of a program or event for which it is not feasible to **randomize**.

A Brief Introduction to Matching

Suppose that you want to evaluate the effect of a scholarship program on students' final academic performance. You have data on every student, including age, gender, family income, parents' education, test scores etc.

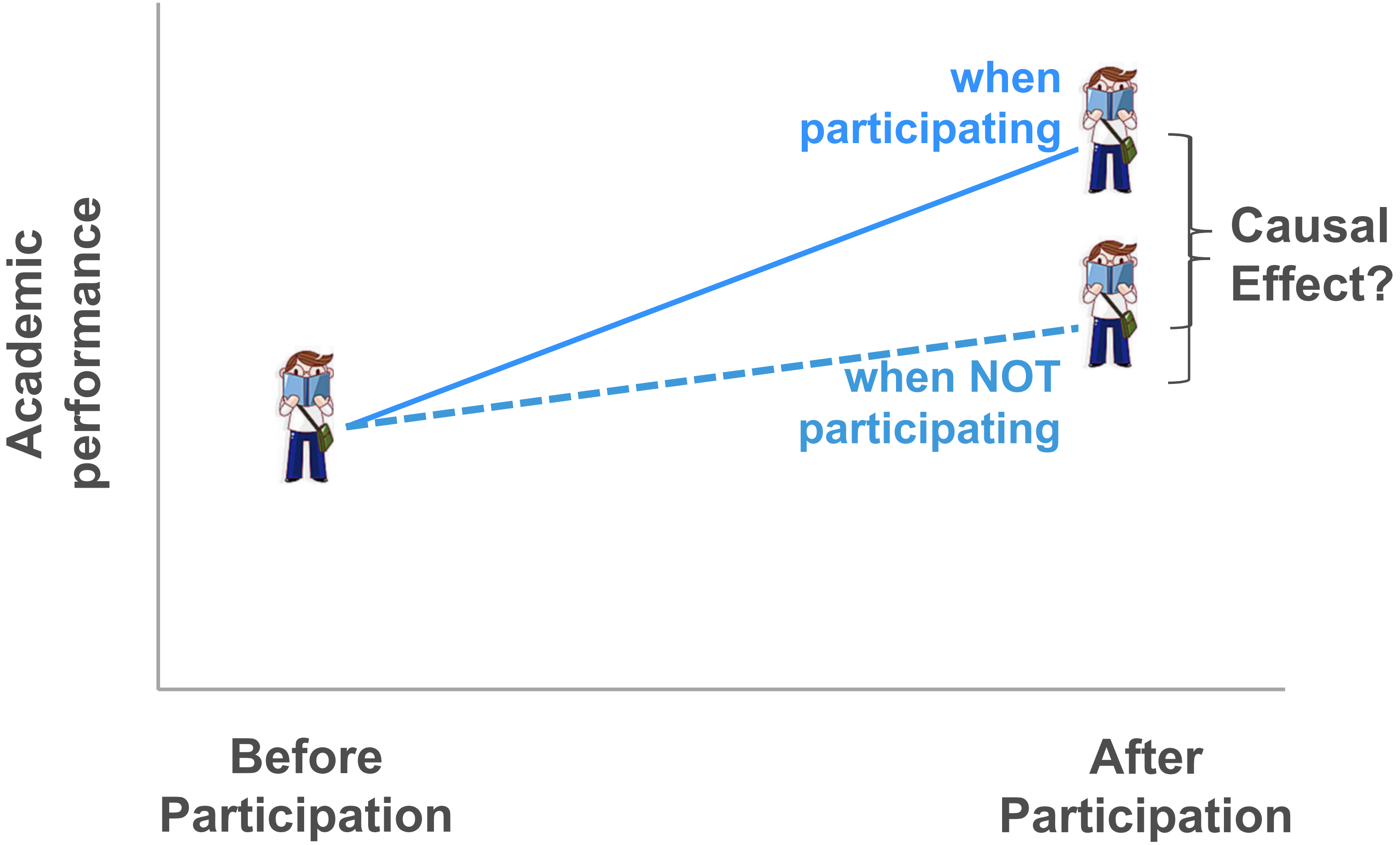
If scholarships were assigned completely at random, we could just compare treatment and control group (randomized experiment).



Let us suppose that this is not the case…

We are interested in causal effects, not association or correlation.

Casual effects describe how an outcome changes (e.g., academic performance) as a direct result of some treatment (e.g., participation in a scholarship program).



European Commission

A Brief Introduction to Matching

In principle, we could just run the following regression:

$$\textit{academic performance}_i = \beta_0 + \beta_1\, \boldsymbol{scholarship_i} + \beta_2 X_i + \varepsilon_i$$

where *scholarship* is a dummy variable for receiving the scholarship and X_i are all the variables that we think may affect students' academic performance (test scores, family income, age, gender, …), ε_i is the error term.

However, by including these controls in a regression specification, and at the same time including the dummy variable for receiving the scholarship, a problem of **endogeneity** may arise (predictor variables correlate with the error term).

Indeed, students may be selected into the "scholarship" because of better test scores or because of a lower family income etc. Estimation based on **the propensity score** can deal with these kinds of situations.

A Brief Introduction to Matching

Estimation of the propensity score:

1. Estimate the **probability of receiving treatment** using the covariates that determine selection

$$\textit{treatment}_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots (+\gamma_1 Xi_1^2 + \gamma_2 X_{i2}^2 \ldots) + \varepsilon_i$$

2. The propensity score is the **predicted value of the *treatment*** (a scalar which summarizes the observed covariates)

$$e(x)=P(T=1|x)$$

- **Important:** the model for treatment should only include variables that are ***unaffected*** by participation in the treatment.
    - **Time invariant** characteristics
    - Variables that are **measured before participation** in the treatment and that are **not affected** by anticipation of participation.

The balancing property of the propensity score

It may be difficult to find a treated and control unit that are closely matched for every one of the many covariates in x, but it is easy to match on one variable, the propensity score, e(x), and doing that will create treated and control groups that have similar distributions for all the covariates.

After the matching, treated (T=1) and control (T=0) subjects with the same propensity score e(x) should have the same distribution of the observed covariates, x:

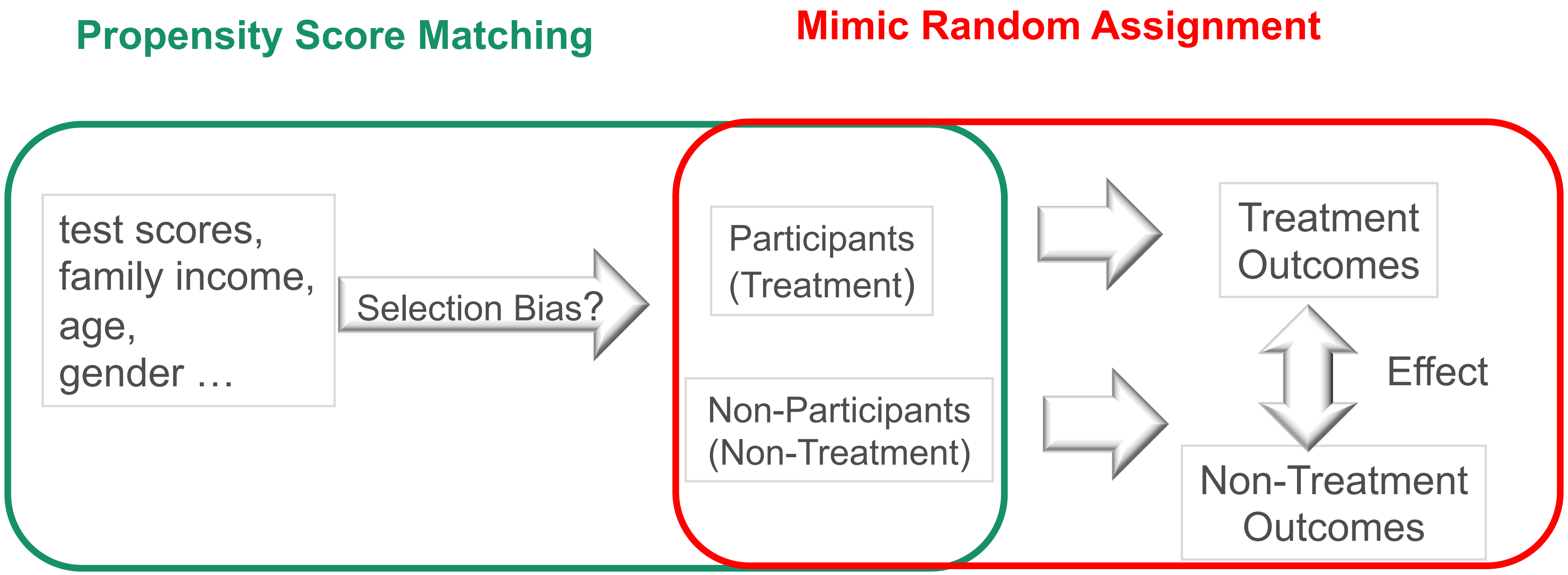
$$P(x|T=1,e(x))=P(x|T=0,e(x))$$

or

$$T \perp x | e(x).$$

conditional on propensity score, the baseline covariates are expected to be balanced between treated and untreated groups i.e. the matching should be effective in removing differences in observable characteristics between treated and controls.

A Brief Introduction to Matching



Once matches are made, we can calculate impact by comparing the means of outcomes across participants and their matched pairs

Case Study

Does labour protection influence mental-health responses to employment shocks? Evidence on older workers in Europe

European Commission

Background

At the beginning of the COVID-19 pandemic crises, when no medicines or vaccines were available, governments worldwide implemented different forms of non-pharmaceutical interventions (NPIs) such as lockdown measures that have impacted many sectors of the economy and workforce…

Despite the COVID-19 pandemic having taken a heavy toll on workers' distress, there is still limited evidence on robust quantification/measurement of this issue in a cross-country perspective.

Aims

This study aims to fill this gap and extend prior research on symptoms of depression related to the COVID-19 crisis by analysing the mental health consequences of job disruption across different countries taking advantage of the heterogeneity of employment policies at country-level.

To this aim, European countries were clustered into three macro-regions characterised by high, intermediate and low employment regulatory protection, using the Employment Protection Legislation (EPL) aggregate score (OECD, Employment Outlook 2020).

Previous literature has found that EPL strictness influences workers' perception of job insecurity that, in turn, may influence their psychological well-being, see e.g. Caroli & Godard, (2016); Clark & Postel-Vinay, (2005); Sverke et al., (2002).

Why Older Workers?

This study focused on older workers aged between **50 and over**, according to the Mutual Information System on Social Protection (MISSOC) tables of each country.

The COVID-19 pandemic has been extremely challenging for older workers who were exposed to **a twin threat**: a higher risk of adverse effects from the COVID-19 and the reduced labour demand as a consequence of the shutdown policies.

Even though older workers are less likely to be made unemployed compared to younger ones during economic downturns, unemployment shocks may have persistent effects on the employment of older workers who are highly vulnerable to **long term unemployment** or **permanent labour market exits** (Kirsten & Heywood, 2007; Crawford & Karjalainen, 2020; Goda et al., 2021).

European Commission

Data

This study makes use of individual-level data drawn from the 8th wave of the Survey of SHARE and the SHARE Corona Survey. The 8th wave of SHARE is a regular wave collecting information on the health, demographic and socio-economic status of individuals who are 50 years old or over. The interviews took place between October 2019 and March 2020.

A sub-sample of SHARE panel respondents was interviewed from June to August 2020, via a Computer Assisted Telephone Interview (CATI), partly to collect a set of basic information as in the regular SHARE questionnaire, and partly to elicit information on life circumstances in the presence of COVID-19.

The data collected with the latter questionnaire provide a detailed picture of how older adults were coping with the health-related and socio-economic impact of COVID-19 (Scherpenzeel et al., 2020).

Sample

The empirical strategy used the employment protection legislation index (EPL) which measures the strictness of employment protection for permanent and temporary contracts and relies on three components as measured by the OECD:

- rules affecting the **individual dismissal** of workers with **regular employment contracts** (**EPR**);
- rules affecting the **collective dismissal** of workers with **regular employment contracts** (**EPC**);
- rules affecting **temporary employment** (**EPT**).

Hence, individuals who were employed (permanent and temporary workers) before the COVID-19 outbreak are included while **self-employed individuals are excluded**.

Sample

Since the OECD measure of EPL is not available for non-OECD members, the sample was further restricted excluding respondents from

- ✓ Bulgaria,
- ✓ Cyprus,
- ✓ Malta and
- ✓ Romania.

Respondents from **Croatia** were also excluded, since the most recent EPL score for this country dates back to 2015. Finally, respondents from the **Netherlands** were excluded from the sample, because information on occupations was not collected after the 6^{th} wave of SHARE, and similarly for **Hungary** and **Israel**, because of limited within-country variation in the variables of interest.

European Commission

Sample

Once conditioning on having no missing value on any dependent variable and/or covariate, the final sample consisted of **3.625 observations** across 19 European countries, namely:

- ✓ Austria,
- ✓ Belgium,
- ✓ Czech Republic,
- ✓ Denmark,
- ✓ Estonia,
- ✓ Finland,
- ✓ France,
- ✓ Germany,
- ✓ Greece,
- ✓ Italy,
- ✓ Latvia,
- ✓ Lithuania,
- ✓ Luxembourg,
- ✓ Poland,
- ✓ Slovenia,
- ✓ Slovakia,
- ✓ Spain,
- ✓ Sweden and
- ✓ Switzerland.

EPL sub-samples

Following Boeri and van Ours (2021), an overall indicator of labour market rigidity of a country was constructed, using simultaneously the strictness of permanent contract (EPR), regulation on temporary contract (EPT) and the strictness of collective dismissal (EPC); the weighted average of these three indicators provides the EPL overall index.

The overall EPL index that was used to stratify the sample into three clusters, namely: low employment regulatory protection countries with an EPL score lower than 2; intermediate employment regulatory protection countries with an EPL score ranging between 2 and 2.5; high regulatory protection countries with an EPL score higher than 2.5 (OECD, Employment Outlook 2020).

Countries by Strictness Employment Protection Legislation (EPL)

Country	*EPL*	*EPR*	*EPC*	*EPT*
Switzerland	1.58	1.61	1.69	1.5
Denmark	1.99	1.94	2.18	1.96
Austria	2.00	1.8	2.14	2.17
Lithuania	2.11	2.24	2.24	1.92
Germany	2.21	2.33	2.61	1.92
Sweden	2.21	2.54	2.72	1.67
Finland	2.22	2.48	2.75	1.75
Slovenia	2.30	2.32	2.68	2.13
Poland	2.31	2.39	2.36	2.21
Latvia	2.36	2.71	2.89	1.79
Estonia	2.41	1.93	2.04	3.04
Belgium	2.48	2.71	2.68	2.17
Slovak Republic	2.53	2.33	2.46	2.75
Czech Republic	2.66	3.03	3.05	2.13
Greece	2.70	2.54	2.55	2.92
Spain	2.71	2.43	2.43	3.1
France	2.96	2.68	3.25	3.13
Luxembourg	3.09	2.54	2.66	3.83
Italy	3.24	2.86	3.19	3.63
Average	2.42	2.39	2.55	2.41

Source: OECD, Employment Outlook 2020 and authors own elaboration. Scores are rounded to two decimals.

Empirical Strategy

From a methological point of view it should be noted that workers who experienced a job disruption may differ substantially from a worker who did not.

- poor-health workers are more vulnerable to COVID-19 adverse effects …
- workers who delivered essential services continued to do their jobs…

This potential endogeneity problem can be corrected by matching each worker who experienced job disruption (the “exposed/treated”) with a worker who did not (the “control/untreated”) on each characteristic known to be associated with job disruption and mental health conditions (Caliendo & Kopeinig, 2008).

This matching was performed by using a propensity score matching, as formalized by Rosenbaum & Rubin (1983).

The Propensity Score Model

To begin with, a probit model was estimated to get the individual propensity score. The dependent variable is a binary variable that takes a value of 1 for respondents who experienced job disruption and 0 otherwise.

"Due to the Corona crisis have you become unemployed, were you laid off or have you had to close your business?" With yes and no as the available answer options.

Control variables: age, sex, macro area of residence (or country of residence – see sensitivity check) marital status, workers education, family size, workers self-reported ability to meet work and family commitments, occupation (essential workers, part-time workers, employed in the public sector vs private sector, worker with multiple job), computer skills, ability to make ends meet, health condition before the COVID-19 outbreak (SAH and chronic conditions).

Finally, we control for COVID-19 specific variables:

COVID-19 spread	1 if anyone close had suffered from the Coronavirus, and/or was hospitalized due to the infection, and/or died after being affected by the Coronavirus; 0 otherwise	SHARE Corona Survey
COVID-19 Government Response Stringency Index (SI) relative change	Relative change in the SI between 12 March 2020 and the month of the interview date	Oxford Coronavirus Government Response Tracker (OxCGRT)

The Propensity Score Model

Once the propensity score was calculated, statistical matching was performed so as to form twin data that differ in terms of the job disruption status alone and not in terms of any of the other observed characteristics.

Since the sample consists of comparatively few workers who experience job disruption in relation to many untreated ones, Kernel and Radius (with caliper 0.05) matching were chosen as the matching algorithms.

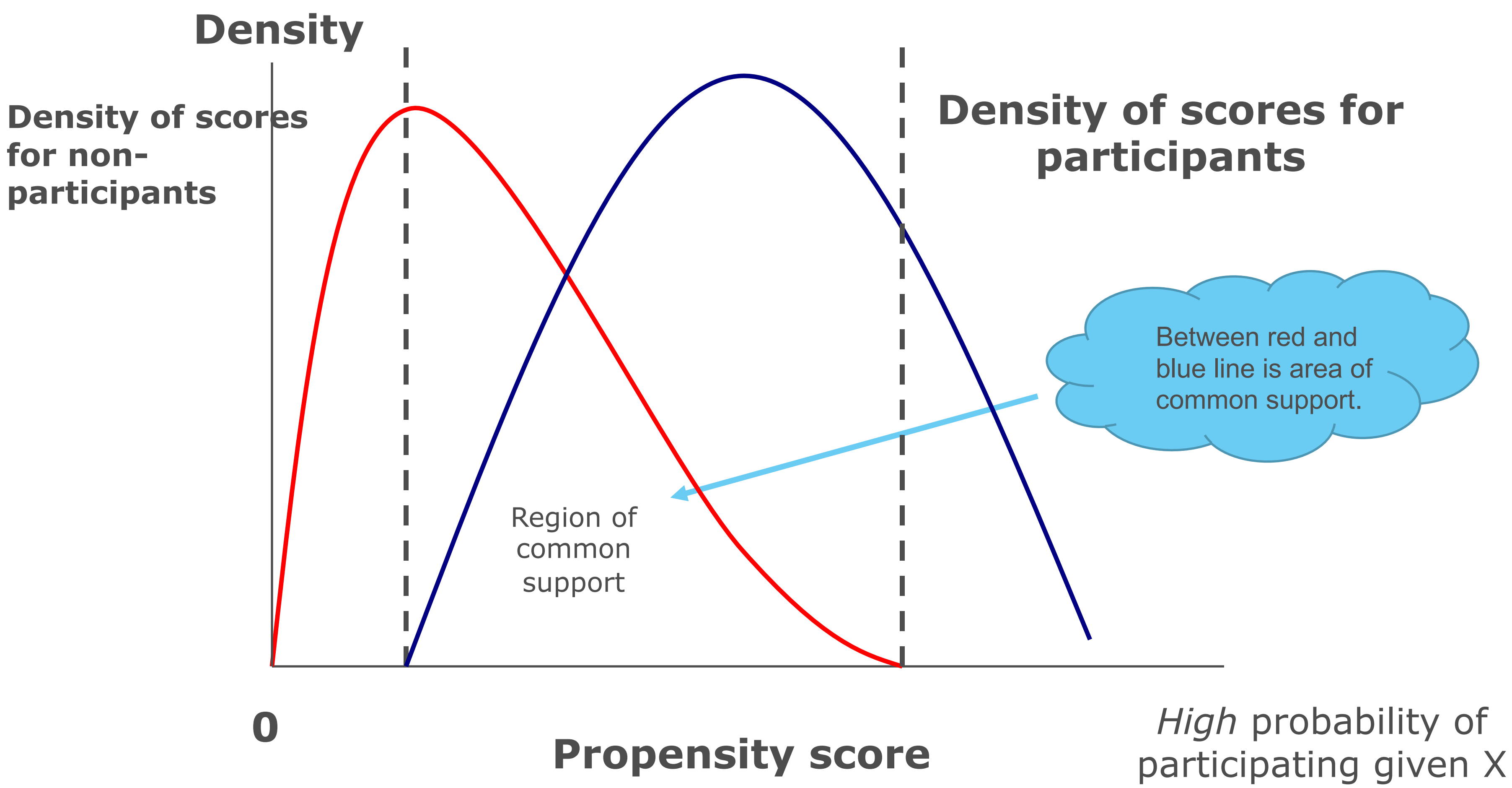
These techniques use the maximum amount of data and, in the case of Radius matching, the imposition of a tolerance threshold avoids the risk of bad matches (Caliendo and Kopeinig 2008; Imbens and Wooldridge, 2009).

The Balancing Test

Out of the region of **“common support”**.

	n. of treated	n. of untreated	n. of treated off	Probit Pseudo R^2 before matching	after matching	p > X2 before matching	after matching	Median Bias before matching	after matching	% reduction in median bias
Kernel Matching										
Full sample	581	3043	1	0.077	0.007	0.000	0.940	9.6	2.7	72%
EPL=1	70	497	1	0.205	0.012	0.000	1.000	16.6	4.2	75%
EPL=2	257	1740	0	0.062	0.012	0.000	0.989	6.4	2.3	64%
EPL=3	254	806	0	0.078	0.004	0.000	1.000	12.6	3.3	74%
Radius Matching										
Full sample	581	3043	1	0.077	0.001	0.000	1.000	9.6	1.2	88%
EPL=1	70	497	1	0.205	0.009	0.000	1.000	16.6	2.7	84%
EPL=2	257	1740	0	0.062	0.002	0.000	1.000	6.4	1.1	83%
EPL=3	254	806	0	0.078	0.001	0.000	1.000	12.6	1.4	89%

The Common Support



The area of common support occurs where the densities of the estimated propensity scores for participants and for non-participants overlap.

Imposing the common support condition in the estimation of the propensity score may improve the quality of the matches used to estimate the ATT.

Individuals' Mental Health

In order to measure the deterioration of workers' mental health related to the pandemic itself, four self-reported psychological distress symptoms were considered, based on the SHARE Corona Survey:

- worsened depressed mood;
- worsened anxiety symptoms;
- worsened sleep problems;
- worsened loneliness.

a synthetic continuous indicator of psychological distress is constructed by extracting the first principal component from the indicators described above. Since all the indicators of psychological health are discrete the polychoric correlation matrix is used as the starting point of the principal component analysis.

Finally, the index is standardized to lie on a continuous scale between 0 (absence of symptoms or worsened symptoms of psychological distress) to 1 (presence of all symptoms of psychological distress that worsened during the COVID-19 outbreak) to aid in interpretation of the results.

Results

Average Treatment Effect on Treated (ATT) - psychological distress index without countries' fixed effect

	Kernel matching			Radius Matching		N.Obs
	ATT	SE		ATT	SE	
Full Sample	0.0526***	0.012		0.0487***	0.013	3,625
Analysis by Cluster						
EPL=1	0.0206	0.032		0.0143	0.046	568
EPL=2	0.0184	0.015		0.0167	0.016	1,997
EPL=3	0.0842***	0.023		0.0780***	0 .021	1,060

Starting from the full sample, the present results show that experiencing a job disruption had a positive and significant impact on worsened symptoms of psychological distress.

These findings also reveal the presence of an EPL gradient: in the group of countries characterized by stronger employment regulation job disruption significantly affected individuals' mental health conditions; specifically, the ATT is significant at 1% level and positive.

Results

Average Treatment Effect on Treated (ATT) - psychological distress index with countries fixed effect

	Kernel matching			Radius Matching		N.Obs
	ATT	SE		ATT	SE	
Full Sample	0.044***	0.014		0.0402***	0.012	3,625
Analysis by Cluster						
EPL=1	0.0178	0.036		0 .0166	0.043	568
EPL=2	0.0183	0.016		0.018	0.015	1,997
EPL=3	0.0749***	0.024		0.0714***	0.023	1,060

Results

Average Treatment Effect on Treated (ATT) - psychological distress index for job disruption for more than 8 weeks

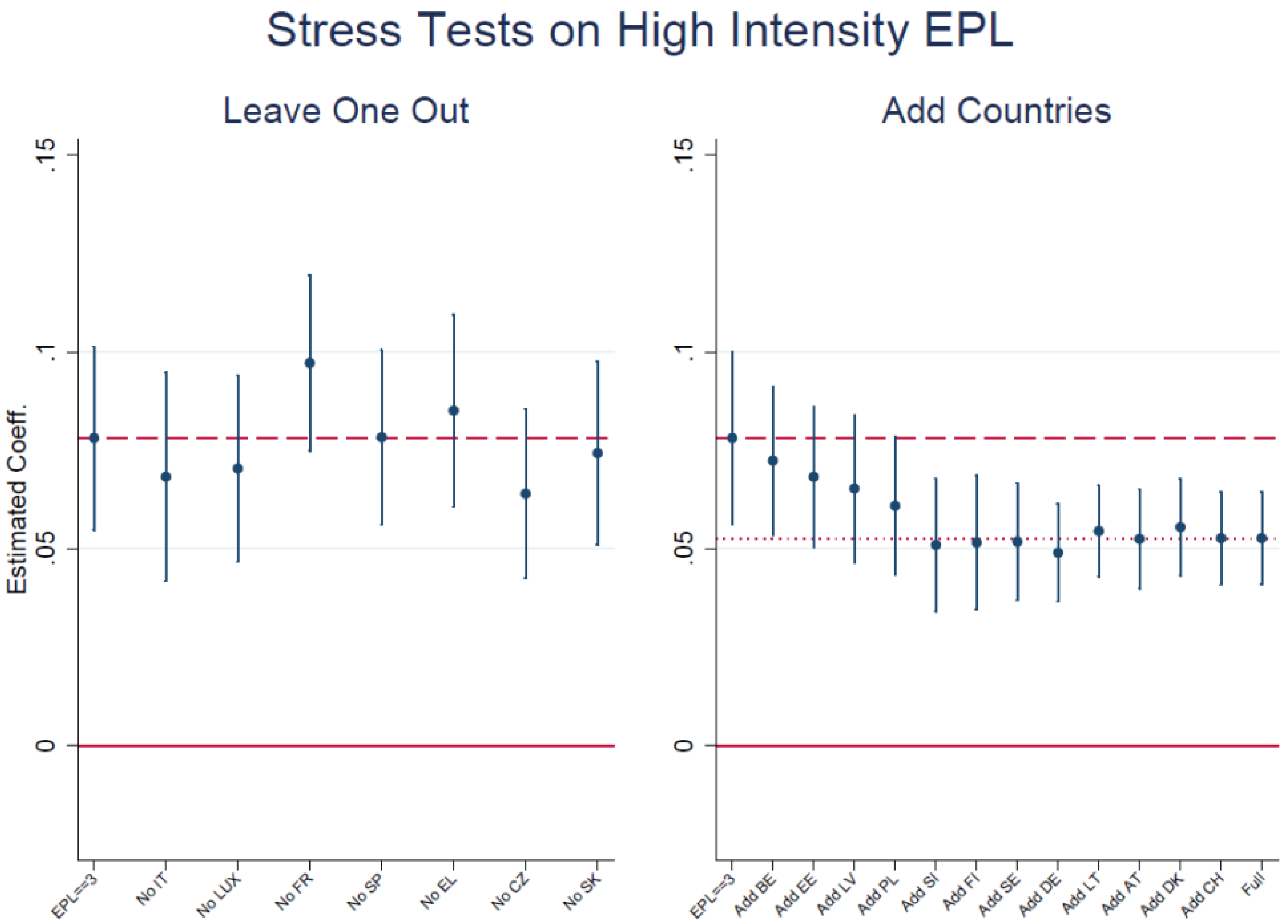
	Kernel matching			Radius Matching		N.Obs
	ATT	SE		ATT	SE	
Full Sample	0.0673***	0.018		0.0587***	0.018	3,287
Analysis by Cluster						
EPL=1	-0.0168	0.054		-0.0159	0.058	520
EPL=2	0.0255	0.026		0.0196	0.026	1,835
EPL=3	0.107***	0.027		0.102***	0.036	918

Results

Average Treatment Effect on Treated (ATT) – Stringency Index as by March, 12 2020

	Kernel matching			Radius Matching		N.Obs
	ATT	SE		ATT	SE	
Full Sample	0.044***	0.014		0.0402***	0.012	3,625
Analysis by Cluster						
EPL=1	0.0178	0.036		0 .0166	0.043	568
EPL=2	0.0183	0.016		0.018	0.015	1,997
EPL=3	0.0749***	0.024		0.0714***	0.023	1,060

Results


Stress Tests on High Intensity EPL
Leave One Out
Add Countries
Estimated Coeff.
.15
.1
.05
0
EPL==3
No IT
No LUX
No FR
No SP
No EL
No CZ
No SK
Add BE
Add EE
Add LV
Add PL
Add SI
Add FI
Add SE
Add DE
Add LT
Add AT
Add DK
Add CH
Full

Results

ATT in EPL 3 cluster by subgroups

	Kernel matching		Radius Matching		n. obs
	ATT	SE	ATT	SE	
Male	0.0747**	0.0298	0.0686**	0.0309	436
Female	0.0798**	0.0328	0.0742**	0.0344	624
High education	0.0700	0.0479	0.0657	0.0565	307
Medium education	0.0670*	0.0309	0.0637**	0.0293	542
Low education	0.1708***	0.0486	0.1836***	0.0559	211

*: significant at 10% level; **: significant at 10% level; ***: significant at 1% level, bootstrap SE

Conclusions

Results reveal a clear EPL gradient:

➢ job disruption has a positive and significant impact (about 8%) on older workers' psychological distress especially in the countries with more binding EPL that might have acted as a **"double-edged sword"**

1. increasing the job security for older workers who did not suffer from any job disruption but

2. increasing at the same time the uncertainty for those who have experienced layoffs given its potential to reduce the outflow rate from unemployment.

Appendix:Implementing the Matching with STATA

Implementing the Matching with STATA

Download the command **“psmatch2”**.

1. Install psmatch2

ssc install psmatch2, replace

2. Estimate the propensity score on the *X*’s e.g. *via* probit or logit;

psmatch2 depvar [indepvars], logit***

Ex (logit): psmatch2 scholarship age gender test_scores family_income, logit

Ex (probit): psmatch2 scholarship age gender test_scores family_income

*** if you want to estimate the propensity score via logit include “logit” instead of the default probit.

Implementing the Matching with STATA

Once the propensity score is computed, you have to choose the matching algorithm that forms statistical twins that differ in their treatment but do not differ in terms of observable.

3. Choose the matching estimator

Ex: psmatch2 scholarship age gender test_scores family_income, logit neighbor

Nearest neighbor: treated and controls are closest in terms of propensity score. Several variants of NN matching are proposed, e.g. NN matching 'with replacement' and 'without replacement'. In the former case, an untreated individual can be used more than once as a match, whereas in the latter case it is considered only once.

Implementing the Matching with STATA

Finally, we look at the effect of treatment on the outcome (in our example of getting the scholarship on academic performance), using the propensity score.

4. Evaluate the outcome

psmatch2 depvar [indepvars], logit nieghbor out (varlist)

Ex: psmatch2 scholarship age gender test_scores family_income, logit neighbor out (academic performance)

the average treatment effect on the treated (att) is displayed.

The matching technique produces two balanced groups, one of treated and one of untreated: the score substitutes a collection of confounding variables with a single covariate that is a function of all the variables.

Implementing the Matching with STATA

The matching should be effective in removing differences in observable characteristics between treated and controls.

In order to test whether the machting has removed differences in observable characteristics between individuals who have received the treatment and individuals who have not, we need to run a balancing test.

The STATA command to run the balancing test is pstest:

pstest age gender test_scores family_income t (scholarship) both

This command sorts data by the propensity score and divides it into blocks (groups) of observations with similar propensity scores. Within each block, pstest tests (using a t-test), whether the means of the covariates are equal in the treatment and control group.

If so → stop, we are done with the first stage. If a particular covariate is unbalanced, we have to modify the initial logit or probit equation by including higher order terms and/or interactions with that covariate and start again.

Implementing the Matching with STATA

- PSMATCH2: Stata module for propensity score matching, common support graphing, and covariate imbalance testing
  - psmatch2.ado
- PSCORE – same basic features
    - More user "friendly"
    - pscore.ado

SUMMARIZING THE GENERAL PROCEDURE

Run Regression:

- Dependent variable: T=1, if participate; T = 0, otherwise.
- Choose appropriate conditioning variables, *X*
- Obtain propensity score: predicted probability (p)

1-to-1 match

- estimate difference in outcomes for each pair
- Take average difference as treatment effect

1-to-n Match

- Nearest neighbor matching
- Caliper matching
- Nonparametric/kernel matching

Multivariate analysis based on new sample

References

Caliendo, M., & Kopeinig, S. (2008) “Some practical guidance for the implementation of propensity score matching” *Journal of Economic Surveys*, 22: 31-72.

Di Novi, C., Paruolo, P., & Verzillo, S. (2022). The Role of Employment Protection Legislation Regimes in Shaping the Impact of Job Disruption on Older Workers’ Mental Health in Times of COVID-19. JRC Working Papers in Economics and Finance, 2022/2.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70, 41–55.

Thank you

This presentation has been prepared for internal purposes. The information and views expressed in it do not necessarily reflect an official position of the European Commission or of the European Union.

European Commission