




Lab-like findings from online experiments

Irene Maria Buso¹ · Daniela Di Cagno² · Lorenzo Ferrari² · Vittorio LaroCCA³ · Luisa Lorè⁴ · Francesca Marazzi⁵ · Luca Panaccione⁶ · Lorenzo Spadoni² 

Received: 23 December 2020 / Revised: 29 November 2021 / Accepted: 29 November 2021 /

Published online: 18 December 2021

© The Author(s), under exclusive licence to Economic Science Association 2021

Abstract

Laboratory experiments have been often replaced by online experiments in the last decade. This trend has been reinforced when academic and research work based on physical interaction had to be suspended due to restrictions imposed to limit the spread of Covid-19. Therefore, data quality and results from web experiments have become an issue which is currently investigated. Are there significant differences between lab experiments and online findings? We contribute to this debate via an experiment aimed at comparing results from a novel online protocol with traditional laboratory settings, using the same pool of participants. We find that participants in our experiment behave in a similar way across settings and that there are at best weakly significant and quantitatively small differences in behavior observed using our online protocol and physical laboratory setting.

Keywords Methodology · Experiments · Lab-like data · Covid-19

JEL Classification C81 · C90

1 Introduction

The spread of Covid-19 has temporarily prevented experimental subjects from physically entering labs. Still, the experimental approach remains a crucial tool to understand individual and group behavior. To overcome the problems raised by physical distancing, researchers have turned to online experiments and surveys employing different platforms. The validity of these protocols has been demonstrated by successfully replicating a series of classic experiments (Crump et al. 2013; Amir et al.

We thank the Editors, Maria Bigoni and Dirk Engelmann, and two anonymous referees for useful comments. We also thank Sofia De Caprariis for her assistance during this project.

✉ Lorenzo Spadoni
lspadoni@luiss.it

Extended author information available on the last page of the article

2012; Horton et al. 2013). Moreover, a recent strand of research focuses on the comparison of quality of data and reliability of results from different platforms and pools of subjects (see, e.g., Gupta et al. 2021; Peer et al. 2021; Litman et al. 2021).

Online experiments feature differences from physical ones that limit the benefits of fundamental aspects of the traditional experimental methods. A first issue concerns subjects dropping out during the experiment: dropouts are problematic both because they may result in (expensive) losses due to discarding observations and because they might be endogenous (Arechar et al. 2018). A second issue concerns participants' limited attention, which could hinder the understanding of instructions, due to limited control: Chandler et al. (2014) show that subjects could engage in other activities while participating in an online experiment (e.g. watching TV, listening to music, chatting, etc.). A third issue concerns the difficulty to control the recruiting process.

During the pandemic, we developed a novel online protocol which replicates the main features of physical experiments and therefore addresses the most relevant problems mentioned above (see Buso et al. 2020). In particular, it ensures: (i) isolated and monitored subjects, (ii) interactions mediated by computers, (iii) anonymity of participants, (iv) immediate monetary reward, and (v) the same recruiting process as in the physical lab, which allows for a better control and ensures that participants are drawn from the standard sample.

To contribute to the current debate comparing web experimental datasets and those collected in the traditional physical lab, in October 2021 we collected data on three standard games (Ultimatum, Dictator and Public Good Game) in traditional physical lab sessions and in two types of online sessions, with and without video monitoring of participants. The different settings in data collection identify our three treatments, which hereinafter are referred to as Physical Lab; Online, monitoring; Online, no monitoring. We find that participants in our experiment behave in a similar way across settings and that there are at best weakly significant and quantitatively small differences in choice data between sessions online and in the physical lab. Therefore, we confirm the validity of our protocol for online experiments and its capability to overcome the aforementioned issues.

The paper is organized as follows: in Sect. 2, we present our online protocol; in Sect. 3, we describe the experimental design; in Sect. 4, we present the results, comparing online and physical lab evidence; we conclude in Sect. 5. In the supplementary online materials, we report the translated instructions (online Appendix A), the instructions in the original language (online Appendix B) and post-experimental questionnaire (online Appendix C), together with additional material regarding our protocol (online Appendix D and online Appendix E).

2 Experimental protocol

The online visually monitored sessions are organized as follows: we adopt an architecture of connected platforms, specifically ORSEE for recruitment (Greiner 2015), Cisco WebEx for (visual) monitoring, oTree (Chen et al. 2016) for running the experiment, and PayPal for payments. In the invitation (see online Appendix D), we

remind participants that a PayPal account is necessary to participate and receive the final payment. For privacy reasons, participants are informed that during the experiment they will be connected, but not recorded, via audio/video with the experimenter during the whole session and, therefore, that they need a suitable device. Participants are also informed that supernumerary subjects will be paid only the show-up fee. Before the beginning of the session, the experimenter randomly allocates registered participants to individual virtual cubicles created using Cisco WebEx, sending them the corresponding link. During the experiment, participants are monitored via webcam and can communicate, via chat and microphone, privately with the experimenter. They cannot see nor talk to each other while the experimenter can talk publicly to all participants. A picture of the experimenter's screen is provided in online Appendix E. As participants log in to Cisco WebEx, the experimenter checks that their webcam and microphone work properly, as well as the overall quality of their internet connection. After completing these checks, the experimenters communicate to participants the access procedure and send them individual and anonymous oTree links. After log-in, participants input their PayPal account in oTree, which will be used for payments.¹ As soon as all participants are ready, the experimenter plays a prerecorded audio file with instructions read aloud for all participants, which preserves common awareness and reduces session effects. Written instructions are also displayed on participants' screens while the audio recording is playing and remain available, by clicking a dedicated button, during the whole experiment. At the end of the session, participants answer a final questionnaire. Once participants complete the questionnaire, they are shown a receipt with their payment data and leave their virtual cubicle.

The non-monitored sessions follow the same protocol, excluding video connection, but preserving the possibility for participants and experimenters to communicate via audio or the chat. The physical lab sessions follow the traditional protocol for experiments, for example as described by Weimann and Brosig-Koch (2019).

As mentioned in the introduction, we believe our protocol addresses the most common issues of online experiments: (i) reducing involuntary dropouts (since oTree links allow participants to re-join the session and continue the experiment) and voluntary ones (by constantly monitoring participants, and communicating with them through webcam and microphone); (ii) mitigating limited attention by experimenters reading instructions aloud simultaneously before the experiment begins and by reducing the participants' engagement in other activities; (iii) controlling for participants' characteristics via recruiting on ORSEE.²

¹ The software stores PayPal accounts in a file separate from that of the experimental decisions, so as to preserve anonymity.

² Similar platforms can be easily adapted to the characteristics of the lab (e.g., IT resources and administrative constraints). For instance, Prolific could be used for payments and subject recruitment, while alternative software such as LIONESS Lab (see Giamattei et al. 2020), Veconlab (see <http://veconlab.econ.virginia.edu/>), zTree Unleashed (see Duch et al. 2020), could be used for the experiment.

3 Experimental design

The experiment features three sequences with different order (between sessions) of one-shot dictator game (DG), ultimatum game (UG) and public good game (PGG), all without feedback. In DG and UG, the proposers' endowment is 20 tokens. Subjects play using the strategy method and role reversal, indicating their offer as proposer in DG and UG, and the minimum amount they accept as receivers (i.e., the rejection threshold) in UG. In PGG, the participants' endowment is 10 tokens and the MPCR is equal to 0.5. They contribute in groups of four and report own contribution to the public good.

Subjects are informed that, at the end of the experiment, one game is randomly selected for payment (1 token = 1 euro). In DG and in UG, each participant is randomly matched with another subject and randomly assigned to a role (either proposer or receiver). In PGG, each participant is randomly assigned to a group with other 3 subjects. This matching procedure, which is performed after participants face the three games, together with the absence of feedback information, guarantees the independence of individual choices. The experiment was programmed in oTree (Chen et al. 2016) and the subjects were paid in cash after the experiment in the physical lab and via PayPal after the online sessions.

The experiment is composed of 9 sessions run between October 15 and October 22, 2021 with a total of 183 participants, students from LUISS Guido Carli University recruited via ORSEE (Greiner 2015). In particular, we ran three sessions in the physical lab, three online sessions with visually monitored subjects, and three online sessions without visual monitoring. Moreover, for each setting we varied the order of the three games across sessions, so that in each treatment we have (i) one PGG-DG-UG session, (ii) one DG-UG-PGG session, and (iii) one UG-PGG-DG session. Sessions in the physical lab were run at CESARE Lab with 60 participants, online sessions involved 63 participants visually monitored³ and 60 non-monitored participants.⁴

Table 1 shows that the composition of the sample in the different treatments is balanced by demographic characteristics. The dummy `Economics` equals 1 when the participant is a student of Economics. `Self-reported RA` is the self-reported willingness to take risks in general on a {0,1, ...,10} scale, where 0, respectively 10 identifies a risk averse, respectively loving subject.⁵ The dummy `Resident` equals

³ One participant was excluded because of technical issues.

⁴ Subjects received the same invitation for both monitored and non-monitored online sessions. In both sessions they were informed that a webcam would be needed for the initial identification phase and during the experiment, and that registration implied the acceptance to be monitored. Although in non-monitored sessions subjects were asked to turn off the webcam after identification, the same invitation with the monitoring acceptance statement was sent for both sessions to avoid self-selection into monitored and non-monitored sessions.

⁵ Specifically, subjects answered the following question (see Dohmen et al. 2011): "Are you a person generally willing to face risks or do you prefer to avoid facing them? Please express one preference on a 0–10 scale, where 0 means 'I do not want to take any risk' and 10 means 'I am very willing to take risks'."

Table 1 Balance table of participants' characteristics by treatment

Variable	(1)	(2)	(3)	<i>t</i> test		
	Physical lab	Online, monitoring	Online, no monitoring	Difference		
	Mean/SE	Mean/SE	Mean/SE	(1)–(2)	(1)–(3)	(2)–(3)
Female	0.417 (0.064)	0.540 (0.063)	0.583 (0.064)	–0.123	–0.167*	–0.044
Age	23.200 (0.567)	23.651 (0.360)	24.200 (0.485)	–0.451	–1.000	–0.549
Economics	0.533 (0.065)	0.556 (0.063)	0.650 (0.062)	–0.022	–0.117	–0.094
Self-reported RA	6.183 (0.280)	6.460 (0.242)	5.967 (0.210)	–0.277	0.217	0.494
Resident	0.417 (0.064)	0.413 (0.063)	0.500 (0.065)	0.004	–0.083	–0.087
Center	0.483 (0.065)	0.476 (0.063)	0.550 (0.065)	0.007	–0.067	–0.074
Easy	0.950 (0.028)	0.968 (0.022)	0.933 (0.032)	–0.018	0.017	0.035
Number of participants	60	63	60			

t tests on differences in means across the groups. The value displayed for *t* tests are the differences in the means across the groups. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level

1 when the participant comes from the Italian region where LUISS Guido Carli University is located.⁶ The dummy *Center* equals 1 when the participant is from the Center of Italy versus other areas. The dummy *Easy* equals 1 when the participant declared (s)he found the experiment easy.⁷

4 Results

In this section, we discuss the experimental results with evidence from the three games. We first present some descriptive results, and then the econometric analysis.

Figure 1 reports average choices for each of the three games by treatment with confidence intervals and shows that, overall, between-treatment differences are negligible. In DG average demand amounts to 73.2% (14.64 tokens) of the pie size (20 tokens): 71.3%, i.e., 14.267 tokens, in the physical lab; 77.5%, i.e., 15.508, in online with visual monitoring and 70.5%, i.e., 14.117, in online without visual monitoring. In UG, average proposer demand amounts to 61.35% (12.27 tokens) of the pie size (20 tokens), and in particular is equal to 59.8% in the physical lab, 60.6% in online with visual monitoring and 63.75% in online without visual monitoring. The average responder rejection threshold amounts to 37.5% (7.5 tokens) of the pie size: 35.8% in the physical lab, 38.25% in online with visual monitoring and 38.4% in online without visual monitoring. In PGG, average contribution amounts to 32% (3.18

⁶ With this variable we aim at distinguishing the students who likely live with their families from those living with other students or workers.

⁷ The translated post-experimental questionnaire is reported in the supplementary online materials (online Appendix C).

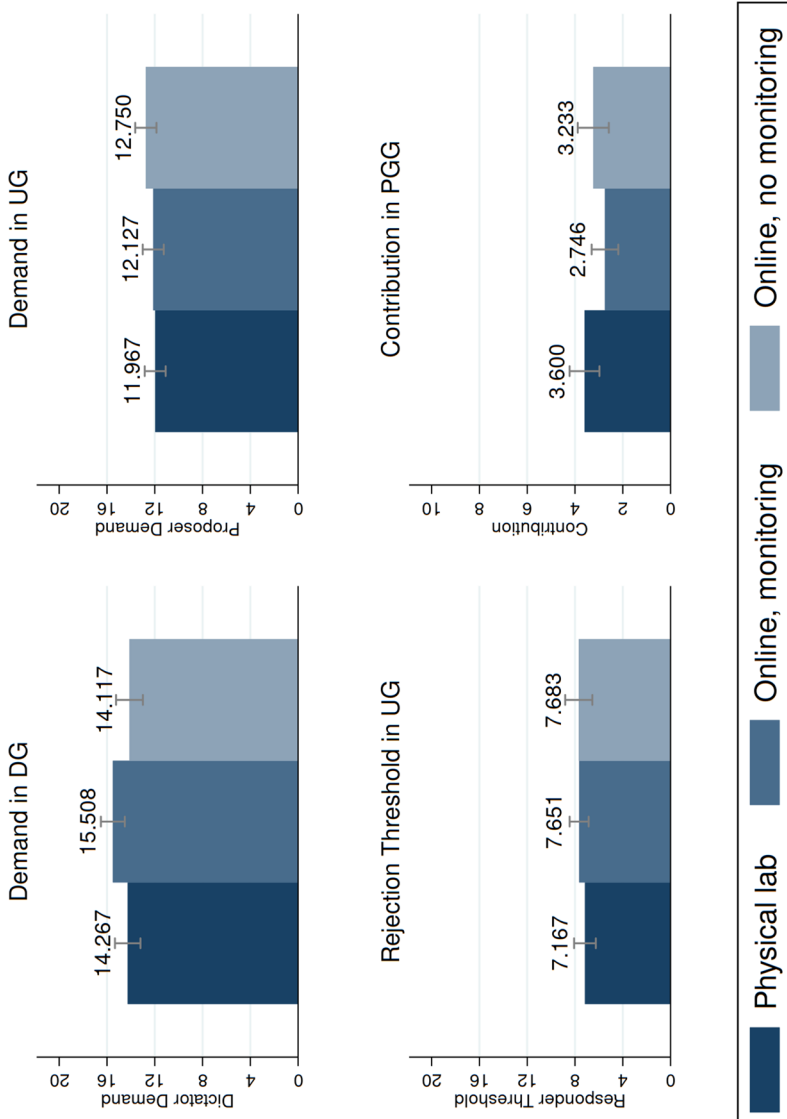


Fig. 1 Average choices by games and treatments

Table 2 Results of OLS regression with individual choices as dependent variables

	DG demand $\beta/(SE)$	UG demand $\beta/(SE)$	UG threshold $\beta/(SE)$	PGG contrib. $\beta/(SE)$
Physical lab	-1.241 (0.758)	-0.160 (0.630)	-0.484 (0.681)	0.854* (0.436)
Online, no monitoring	-1.391* (0.758)	0.623 (0.630)	0.033 (0.681)	0.487 (0.436)
Constant	15.508*** (0.530)	12.127*** (0.440)	7.651*** (0.476)	2.746*** (0.305)

Coefficients from OLS regression with individual choices in each game as dependent variables. Standard errors in parentheses, * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

tokens) of the endowment (10 tokens), and in particular it equals 36% in the physical lab, 27.5% in online with visual monitoring and 32.3% in online without visual monitoring.

To verify whether there are significant differences in choice data between online and physical lab sessions, we run two sets of regressions. In both analyses, Physical Lab and Online, no monitoring are compared with the baseline, i.e. Online, monitoring.

The first set of regressions aims at checking whether the between-treatment differences observed in Fig. 1 are statistically significant. Individual choices in each game are analysed separately via OLS regressions using treatment dummies as covariates. The results, reported in Table 2, confirm the absence of treatment effects for both UG choices. For DG demand and PGG contributions, we find weakly significant effects between the two online treatment, and the physical lab and the baseline, respectively.⁸

In the second set of OLS regressions we expand the set of covariates with dummies indicating the sequence according to which the games were played (with the sequence DG-UG-PGG used as baseline) and individuals' characteristics. The latter include participants' gender and age, whether they reside in the Center of Italy,⁹ and self-reported risk attitude. Results are reported in Table 3 and confirm the same treatment effects observed in Table 2. Furthermore, we find a significant positive (negative) effect on the contribution when the PGG is played first (by students of Economics).

5 Conclusion

We compare lab-data collected in physical and online labs using participants from the same pool of subjects to validate our lab-like methodology that ensures (visual) monitoring, common reading of instructions, and isolation of participants.

⁸ Two sample t tests reveal no treatment effect for all choices when comparing the physical lab setting and the online setting without monitoring.

⁹ Results are not affected if we substitute this variable with `Resident`.

Table 3 Results of OLS regression with individual choices as dependent variables with sequence and demographic controls

	DG demand $\beta/(SE)$	UG demand $\beta/(SE)$	UG threshold $\beta/(SE)$	PGG contrib. $\beta/(SE)$
Physical lab	-1.276 (0.773)	-0.023 (0.637)	-0.432 (0.696)	0.818* (0.438)
Online, no monitoring	-1.414* (0.775)	0.619 (0.638)	0.017 (0.698)	0.589 (0.439)
PGG-DG-UG	0.474 (0.775)	-0.670 (0.639)	0.051 (0.699)	0.967** (0.439)
UG-PGG-DG	0.336 (0.760)	-0.396 (0.626)	0.851 (0.685)	0.376 (0.430)
Female	-0.359 (0.651)	1.004* (0.536)	0.470 (0.587)	-0.410 (0.369)
Age	-0.059 (0.089)	-0.062 (0.073)	0.068 (0.080)	0.004 (0.050)
Center	0.467 (0.629)	0.707 (0.518)	0.453 (0.567)	0.240 (0.356)
Economics	0.911 (0.672)	0.130 (0.553)	0.197 (0.605)	-0.808** (0.380)
Self-reported RA	0.082 (0.168)	0.162 (0.138)	0.106 (0.151)	0.049 (0.095)
Constant	15.591*** (2.707)	11.927*** (2.231)	4.479* (2.440)	2.454 (1.533)

Coefficients from OLS regression with individual choices in each game as dependent variables. Standard errors in parentheses, * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Results from UG, DG, and PGG show that there is only one weakly significant difference in one of the choices between the physical lab and the online setting with visual monitoring, but no significant differences between the physical lab and the online setting without visual monitoring. Therefore, data generated on the web with our protocol is comparable with that in the physical lab. Furthermore, we find only one weakly significant difference in one of the choices between the online setting with and without visual monitoring.¹⁰ Overall, we have proposed a validation of our protocol which reduces the debated side-effects arising from online experiments.

Moreover, since our protocol is based on recruitment from a pool of registered students, for example, via ORSEE, it allows to control for their characteristics, such as experience, gender, or field of study. More importantly, it allows to run a new strand of experiments based on the interaction of participants located in different geographical areas, and therefore embedded in their own cultural environment, as if they were simultaneously in the same physical lab.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s40881-021-00114-8>.

References


Amir, O., Rand, D., & Kobi, Y. (2012). Economic games on the internet: The effect of \$1 stakes. *PLoS One*. <https://doi.org/10.1371/journal.pone.0031461>.

¹⁰ This result is not directly comparable to those of Gupta et al. (2021), since they compare different populations commonly employed in economic experiments.

- Arechar, A., Gächter, S., & Molleman, L. (2018). Conducting interactive experiments online. *Experimental Economics*, 21(1), 99–131. <https://doi.org/10.1007/s10683-017-9527-2>.
- Buso, I. M., De Caprariis, S., Di Cagno, D., Ferrari, L., Larocca, V., Marazzi, F., Panaccione, L., & Spadoni, L. (2020). The effects of covid-19 lockdown on fairness and cooperation: Evidence from a lablike experiment. *Economics Letters*, 196(C), S0165176520303487. <https://EconPapers.repec.org/RePEc:eee:ecole:v:196:y:2020:i:c:s0165176520303487>
- Chandler, J., Mueller, P., & Paolacci, G. (2014). Nonnaïveté among Amazon Mechanical Turk workers: Consequences and solutions for behavioral researchers. *Behavior Research Methods*, 46, 112–130. <https://doi.org/10.3758/s13428-013-0365-7>.
- Chen, D., Schonger, M., & Wickens, C. (2016). oTree—an open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, 9(C), 8–97. <https://EconPapers.repec.org/RePEc:eee:beexfi:v:9:y:2016:i:c:p:88-97>
- Crump, M., McDonnell, J., & Gureckis, T. (2013). Evaluating Amazon’s Mechanical Turk as a tool for experimental behavioral research. *PLoS One*, 8(3), e57410. <https://doi.org/10.1371/journal.pone.0057410>.
- Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., & Wagner, G. G. (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association*, 9(3), 522–550. <https://doi.org/10.1111/j.1542-4774.2011.01015.x>.
- Duch, M. L., Grossmann, M. R. P., & Lauer, T. (2020). z-Tree unleashed: A novel client-integrating architecture for conducting z-Tree experiments over the internet. *Journal of Behavioral and Experimental Finance*, 28(3), 100400.
- Giamattei, M., Yahosseini, K. S., Gächter, S., & Molleman, L. (2020). LIONESS Lab: A free web-based platform for conducting interactive experiments online. *Journal of the Economic Science Association*, 6(1), 95–111. <https://doi.org/10.1007/s40881-020-00087->. https://ideas.repec.org/a/spr/jesaex/v6y2020i1d10.1007_s40881-020-00087-0.html
- Greiner, B. (2015). Subject pool recruitment procedures: Organizing experiments with ORSEE. *Journal of the Economic Science Association*, 1(1), 114–125. <https://doi.org/10.1007/s40881-015-0004-4>.
- Gupta, N., Rigotti, L., & Wilson, A. (2021). The experimenters’ dilemma: Inferential preferences over populations. [arXiv:2107.05064](https://arxiv.org/abs/2107.05064)
- Horton, J., Rand, D., & Zeckhauser, R. (2013). The online laboratory: Conducting experiments in a real labor market. *Experimental Economics*, 14, 399–425. <https://doi.org/10.1007/s10683-011-9273-9>.
- Litman, L., Moss, A., Rosenzweig, C., & Robinson, J. (2021). Reply to MTurk, Prolific or panels? Choosing the right audience for online research. <https://ssrn.com/abstract=3775075> or <https://doi.org/10.2139/ssrn.3775075>
- Peer, E., Rothschild, D., Gordon, A., Evernden, Z., & Damer, E. (2021). Data quality of platforms and panels for online behavioral research. *Behavior Research Methods*. <https://doi.org/10.3758/s13428-021-01694-3>.
- Weimann, J., & Brosig-Koch, J. (2019). *Methods in Experimental Economics*. Springer.

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Irene Maria Buso¹ · Daniela Di Cagno² · Lorenzo Ferrari² · Vittorio Larocca³ · Luisa Lorè⁴ · Francesca Marazzi⁵ · Luca Panaccione⁶ · Lorenzo Spadoni² 

Irene Maria Buso
irenemaria.buso@unive.it

Daniela Di Cagno
ddicagno@luiss.it

Lorenzo Ferrari
lferrari@luiss.it

Vittorio LaroCCA
vlarocca@luiss.it

Luisa Lorè
luisa.lore@uibk.ac.at

Francesca Marazzi
francesca.marazzi@uniroma2.eu

Luca Panaccione
luca.panaccione@uniroma1.it

- ¹ Department of Economics, Ca' Foscari University of Venice, Venice, Italy
- ² Department of Economics and Finance, Luiss University, Rome, Italy
- ³ Center for Experimental Studies of Internet, Entertainment and Gambling (CESIEG), Luiss University, Rome, Italy
- ⁴ Department of Economics, University of Innsbruck, Innsbruck, Austria
- ⁵ Centre for Economic and International Studies (CEIS), University of Rome Tor Vergata, Rome, Italy
- ⁶ Department of Economics and Law, Sapienza University of Rome, Rome, Italy