CERN IdeaSquare Journal of Experimental Innovation, 2023; 7(3):5-7
DOI: https://doi.org/10.23726/cij.2023.1492

METHODOLOGICAL NOTE

Quasi-experiments to estimate the economic impact of research infrastructures

Massimo Florio[1], Paolo Castelnovo[2]

[1]Università degli Studi di Milano, Italy
[2]Università degli Studi dell'Insubria, Italy
Corresponding author: paolo.castelnovo@uninsubria.it

INTRODUCTION

Social scientists try to act as natural scientists, relying on empirical evidence from observation and experimentation. However, in social sciences, experiments (field or lab) are difficult. For this reason, in the last decades social scientists have increasingly relied on quasi-experimental designs (see Meyer, 1995), which differ from observational studies (like e.g., cohort studies, case-control studies, and cross-sectional studies) in which observations are made without the implementation of any intervention.

The purpose of this methodological note is to explain and discuss the quasi-experiment (QE) methodology. We will show that the QE approach can be useful to evaluate the impact of any policy measure. Specifically, we will provide an example of its application to the evaluation of the socio-economic impacts of large-scale Research Infrastructures (RIs). RIs require costly investment, funded by public money, and can generate several benefits to society, that occur because their construction and operation involve solving new problems. This 'learning by doing' process (Arrow 1962) has spillover effects on firms providing technologies through procurement contracts, early career researchers acquiring skill through hands-on activities, developers of open and free software (Florio 2019). In the rest of this paper, the QE approach for the empirical analysis of technological procurement spillovers is discussed.

WHAT IS A QUASI-EXPERIMENT DESIGN?

A QE is a research design which mimics experimental design but lacks the random assignment to the treatment or control group. Indeed, while in a real experiment units have the same chance to be assigned to a given treatment condition, in a quasi-experimental design the assignment is based on something other than random (Nichols and Edlund, 2023). This may happen when the researcher has control over assignment to the treatment condition but uses some criteria other than randomness (e.g., a cutoff score) to determine which unit receives the treatment. Alternatively, QEs are used when researchers cannot control or manipulate the treatment condition assignment due to ethical concerns, costs, feasibility, practical constraints, or other reasons. Instead, they take advantage of naturally occurring groups or conditions and then compare the outcomes of those groups.

As such, QEs cannot ensure that treatment and control groups are ex-ante identical (despite not even randomization itself guarantee that groups will be equivalent at baseline). For this reason, QEs are subject to concerns regarding internal validity, that is their effectiveness to estimate the causal impact of an intervention on a population outcome. This is particularly true if there are confounding variables that cannot be controlled for. To address this limitation, researchers have developed statistical techniques to account for unobserved factors and establish a causal relationship between the "quasi-independent" and the outcome variable. However, the strength of the causal claim in QEs is typically considered weaker than in true experiments.

TYPES OF QE DESIGNS

Different QE designs can be used by social scientists to investigate causal relationships when random assignment of participants is not feasible. Among them, the most used include:

Pre-Post Design: researchers observe the dependent variable before and after the intervention within a single group. Changes in the dependent variable are attributed to the intervention (see e.g., Florio et al. 2018 and Castelnovo et al. 2023b).

Time-Series/Panel Design: it is commonly used in the context of policy or program evaluation. Data are collected over a series of time points both before and after the implementation of a policy or program. The introduction of a new law, serves as the quasi-experimental manipulation, allowing researchers to analyze changes in the dependent variable over time to assess the policy impact. In this setting, time and individual fixed effects can be useful to control for unobserved heterogeneity (see e.g. Castelnovo et al., 2018).

Non-Equivalent Control Group Design: researchers compare a group that receives the treatment (intervention) with a similar group that does not. Participants in the groups are not randomly assigned, but researchers try to match them based on relevant characteristics. To this aim, they can use algorithms like propensity score matching which matches participants in treatment and control groups based on their likelihood of receiving the treatment. This procedure helps researchers to control for confounding variables and make the groups more comparable. Post-treatment differences in outcomes between the two groups will be attributed to the intervention. This design can be enriched with a pre-intervention measurement to control for differences between groups at baseline (see e.g. Castelnovo et al. 2023a).

Regression Discontinuity Design: participants are assigned to different groups based on a cutoff score or threshold. The idea is that participants on either side of the cutoff are very similar in terms of their characteristics, except for the treatment eligibility, allowing for a causal inference (see Hahn et al. 1999).

In all the previous designs, it is important to address potential confounding variables, i.e., variables that could affect the outcome and that are not controlled through random assignment. Including observable control variables and using suitable statistical techniques is crucial to allow for a causal interpretation of estimation results.

AN APPLICATION: HOW TO USE QEs TO ASSESS TECHNOLOGICAL PROCUREMENT SPILLOVERS?

In the evaluation of the socio-economic impacts of large-scale RIs on firms, the units of observation are not individuals but companies; instead of health conditions, we have economic performance (the health of the company) which can be measured according to different indicators, as the number of filed/granted patents, R&D expenditure, revenues, profits, etc. The treatment is becoming a RI supplier (or more generally, being the recipient of any innovation policy). If we know in which year the contract was signed, we can identify cohorts. If we consider companies that would have been eligible for a contract with a RI but did not get it (for example, because of preferential policy in favor of Member States), we may build a control group.

It is worth noticing that firms receiving an order from a RI are not a random draw from the universe of companies, as they need to pass certain selection criteria for procurement, including legal, technological, and financial capacity. Also, the control group may be a non-random sample, for example if it is made up of companies that self-selected themselves as potential suppliers registering in a record (see, e.g., the ESA List of Potential Suppliers).

In the empirical analysis of technological procurement spillovers, QEs point to the following research question:

"Does becoming a supplier of technology for a RI have a causal impact on the company performance, after considering any confounding variables?"

An illustrative conceptual model, which can be applied to the evaluation of any policy measure, is represented by this generic equation:

$$Y_{tn} = f\,(P_{tn}, X_{tn}, E_t, Z_n, e_{tn})$$

where Y is a performance variable in year t=1 …T, for the company $n= 1,\ldots, N$; P_{tn} is the treatment event (such as receiving a procurement order) in year t for each company; X is a vector of control variables describing the company characteristics (e.g., industrial sectors, size by assets or by employees), E is a vector of time-variant exogenous effects (dummies for the year of observation), Z is a vector of time-invariant effects (e.g. a company identifier); and e_{tn} is a stochastic error component.

Ideally, we would like that the sample includes many treated and untreated units. If P_{tn} is a dichotomous variable, the value 0 means that the n-company is untreated at time t, hence it is assigned to the control group, while if $P_{tn} = 1$ the company is assigned to the treated group. In QEs, heterogeneous treatment timing is frequent: the treatment, i.e., the procurement order, is administered to different companies in different years. This means that some companies shift their status from 0 to 1 in a year t>1.

What we aim to detect is a significant statistical difference between the outcome of treated and untreated companies, after controlling for firm-level characteristics and fixed effects.

If companies are randomly assigned to each sub-sample, that would be an experiment, like a randomized control trial in medicine. As this is impossible in our context, because companies are selected as suppliers based on observable pre-treatment characteristics (e.g., technological skills), a QE design needs creating an artificial control group where the companies are as similar as possible to the treated group according to the X vector. This implies using matching algorithms like, e.g., propensity score matching, nearest neighbor matching, exact matching, or coarsened matching (see e.g., Augurzky and Kluve, 2007).

If the research design involves heterogeneous treatment timing because companies become suppliers in different years, each company may act as a control unit before shifting from 0 to 1 state. In situations where eventually all the companies in the sample are treated by the end of the considered period, one can exploit such dynamics using "not-yet treated" units as controls, instead of recurring to matching algorithms to create a

control group of "never treated" (see Callaway and Sant'Anna, 2020). An advantage of this approach is that comparing treated units receiving the treatment at different moments in time does not require the change in the firms' status to be an exogenous random event, but only requires its timing to be random.

In the next paragraph, we will present an empirical application of the QE methodology.

CASE STUDY: THE ITALIAN SPACE AGENCY

Castelnovo et al. (2023a) used a combined approach. After having built the dataset of companies receiving a technological order from the Italian Space Agency (ASI), they selected firms' patents' stock as their *Y* (including a depreciation parameter to account for value loss over time). As a second step, an artificial control group of not-suppliers was created. Ideally, these should be 'twins' of the suppliers. To achieve this goal, they built a very large database of around two million companies, selecting companies belonging to the same industry 4-digit NACE code of ASI suppliers. Then a subsample of 250,000 companies was randomly extracted. As a last step, a propensity score matching approach is applied year by year, using pre-treatment tangible and intangible assets, operating revenues, number of employees, year of incorporation and sector as key variables in the matching algorithm. The result is a control group of companies with no statistical differences from the treated ones in the relevant variables.

Then, they used a staggered diff-in-diff approach (Fadlon and Nielsen, 2021), where the exploitation of the different treatment timing allows addressing potential endogeneity issues stemming from the non-random assignment of treatment, being the control group made by treated units as well, which nevertheless receive the treatment in a different year. This approach is particularly suited in that setting, since the treatment cannot be considered exogenous because, as already mentioned, ASI is likely to select its suppliers according to some pre-treatment characteristics.

The authors concluded that becoming an ASI supplier increased companies' patent stock by around 10% compared to the control group. This effect is more significant for hi-tech firms.

CONCLUSION

Are QEs always better than observational studies? The most important caveat is building a control group that accurately mimics the treated group. Indeed, in social sciences, as well as in medicine, we never observe all the features of a company, or of a patient, thus the matching procedure can be rather crude, as it is limited to selected observables which may or may not represent well possible confounding factors.

This issue does not prevent the design of a QE, but researchers should be aware of its limitations, and consider that in some cases a well-executed observational study, based on a lot of effort in getting and cleaning data, may lead to results that point in the right direction. This is particularly true when we are not much interested to a precise estimate of the effect on the selected outcome, but rather on its approximated magnitude. However, when the goal is claiming causality, a QE may be more defensible than an observational study.

REFERENCES

Arrow, K. J., 1962, The economic implications of learning by doing. The review of economic studies, 29(3), 155-173.

Augurzky, B. and Kluve, J., 2007, Assessing the Performance of Matching Algorithms When Selection into Treatment Is Strong. Journal of Applied Econometrics, 22(3), 533–557.

Callaway, B., Sant'Anna, P., 2020, Difference-in-Differences with multiple time periods. Journal of Econometrics, 225 (2), 200–230.

Castelnovo P., Florio M., Forte S., Rossi L. and Sirtori E., 2018, The economic impact of technological procurement for large-scale research infrastructures: Evidence from the Large Hadron Collider at CERN. Research Policy, Vol. 47(9): 1853-1867.

Castelnovo, P., Clò, S., & Florio, M., 2023a, A quasi-experimental design to assess the innovative impact of public procurement: An application to the Italian space industry. Technovation, *121*, 102683.

Castelnovo P., Catalano J., Giffoni F. and Landoni M., 2023b, The outcomes of public procurements: an empirical analysis of the Italian space industry. *The* Journal of Technology Transfer, available online at: https://link.springer.com/article/10.1007/s10961-023-10038-6

Fadlon, I., Nielsen, T.H., 2021, Family labor supply responses to severe health shocks: evidence from Danish administrative records. Am. Econ. J. Appl. Econ. 13 (3), 1–30.

Florio, M., 2019, Investing in science: Social cost-benefit analysis of research infrastructures. MIT Press.

Florio M., Giffoni F., Giunta A. and Sirtori E., 2018, Big science, learning, and innovation: evidence from CERN procurement. Industrial and Corporate Change, 27 (5), 915-936.

Hahn J., Todd P. and Van der Klaauw W., 1999, Evaluating the Effect of an Antidiscrimination Law Using a Regression-Discontinuity Design. NBER paper n. 7131.

Meyer, B. D., 1995, Natural and Quasi-Experiments in Economics. Journal of Business & Economic Statistics, 13(2), 151–161.

Nichols A.L. and Edlund J., 2023, The Cambridge Handbook of Research Methods and Statistics for the Social and Behavioral Sciences. Cambridge University Press, Cambridge, UK.