Research Institute for the Evaluation of Public Policies

Combining process tracing and synthetic control method: Bridging two ways for making causal inference in case studies

Federico Podestà

February 2021

FBK-IRVAPP Working Paper No. 2021-01

Combining process tracing and synthetic control method: Bridging two ways for making causal inference in case studies

Federico Podestà
FBK-IRVAPP
podesta@irvapp.it

FBK-IRVAPP Working Paper No. 2021-01

February 2021

Research Institute for the Evaluation of Public Policies
Bruno Kessler Foundation
Vicolo dalla Piccola 2, 38122 Trento (Italy)

Phone: (+39) 0461.314209
Fax: (+39) 0461.314240

E-mail: info@irvapp.it
Website: http://irvapp.fbk.eu

The purpose of the IRVAPP Working Papers series is to promote the circulation of working papers prepared within the Institute or presented in IRVAPP seminars by outside researchers with the aim of stimulating comments and suggestions. Updated reviews of the papers are available in the Reprint Series, if published, or directly at the IRVAPP.

The views expressed in the articles are those of the authors and do not involve the responsibility of the Institute.

Combining process tracing and synthetic control method: Bridging two ways for making causal inference in case studies

Federico Podestà

Abstract

This paper discusses potential methods of triangulation between two leading methods for case-study research: synthetic control (SC) method and process tracing (PT) method. Both are designed to examine certain events that occur in given cases but view these events from different causal perspectives. Seeing an event as a cause, SC estimates its impact on one or more outcomes. Conversely, seeing an event as an outcome, PT discloses the causes which generated it. Hence, one can start from the causal explanation reached via one of the two methods and then proceed to examine that explanation through the other method. Once the causes of an event are traced via a PT analysis, that account can be validated by estimating the effects of those causes via SC. Equally, once the impact of a certain event is estimated through SC, causal mechanisms traceable via PT can be exploited to refine that impact evaluation.

Keywords: Synthetic control method; Process tracing; Causal inference; Case study method.

Acknowledgments: For helpful comments on previous versions, I thank Davide Azzolini, Martina Bazzoli, Sergiu Constantin Burlacu, Daniela Sicurelli. Furthermore, I would like to thank Nicola Bazoli, Joan Madia, Giovanni Mellace and Alessandra Pasquini for their very helpful discussions on the various aspects of this article. Further thanks go to Alessandra Pasquini for her precious technical support.

1. Introduction

In recent years, a growing literature has examined the combination of different methods for comparative research. Many works have proposed combining quantitative and qualitative methods (for an overview, see e.g. Kuehn and Rohlfing 2016). This article also moves in that direction. However, in contrast to most efforts, the present attempt does not concern the large-N/small-N issue but focuses instead on the sole case-study framework. The potential combination in this case regards two leading methods designed to examine a single or a few cases: the synthetic control (SC) method and the process tracing (PT) method.

SC is a quantitative method ascribable to the impact-evaluation perspective based on the counterfactual model of causality. It was initially introduced by Abadie and Gardeazabal (2003) and subsequently developed by Abadie et al. (2010; 2015) and by numerous econometricians (for an overview, see Abadie 2019). On the other hand, PT is a qualitative method designed to craft suitable explanations of specific outcomes. Referring to the historian's methodology of explanation, George (1979) first proposed using PT in the social sciences[1]. PT was then extensively developed by George and several other scholars (e.g. Bennett and George 2001; George and Bennett 2005; Bennett 2008a; 2010; Collier 2011; Mahoney 2012).

Although these two methods derive from and belong to different methodological traditions, their respective developers claim that both aim at a convergence between quantitative and qualitative approaches. For instance, Abadie et al. (2015: 495) discuss the use of SC as a way to bridge the quantitative/qualitative divide in comparative politics. Conversely, Bennett and Checkel (2015: 9) argue that PT allows for systematic and rigorous qualitative analysis that can complement the correlational approach.

On the basis of these similar ambitions, assessing the possibility of using SC and PT in conjunction is a natural objective. Nevertheless, two contrasting characteristics of the two methods and, more generally, of the quantitative/qualitative divide must be considered as cornerstones: the research question and causal explanation (Mahoney and Goertz 2006).

SC is inherently designed to answer the question: "What is the impact of the event[2] X that occurred in case Z?". For instance, SC's inventors employed it to answer the following research questions: (1)

[1] PT differs from historical scholarship because it is theoretically guided in explaining a given outcome and can be used to test and/or formulate hypotheses (Beach and Pedersen 2013; Mahoney 2012; see also section 3).

[2] Throughout the paper, the term 'event' is preferred to 'process' or 'intervention' in order to be consistent with the language adopted by PT users and counterfactual analysts. Mahoney and Barrenechea (2019) assert that an event is a group of possible worlds that are members of a distinct set denoting a phenomenon. They include general phenomena such as revolutions and wars that have multiple actual world instances. Events can also be specific occurrences such as the French Revolution or the Second World War, which have one actual world instance and multiple possible world instances.

What is the economic impact of the terrorist conflict in the Basque Country (Abadie and Gardeazabal (2003)? (2) What was the effect of Proposition 99, a large-scale tobacco control program implemented by California in 1988 (Abadie et al 2010)? (3) What was the economic impact of German reunification in 1990 (Abadie et al 2015)? Accordingly, SC is an effects-of-causes-oriented method. It makes it possible to estimate the effects of certain events (e.g. German reunification, natural disasters, terrorist conflicts, etc.) that occurred in a single or a few cases[3] (such as cities, regions, or countries) on particular aggregate outcomes (e.g. GDP, private consumption, average income, crime rate, etc).

On the other hand, PT serves to answer the question: "Which causes determined the occurrence of the event Y in case Z?". Therefore, it was employed to answer the following research questions: (1) Which conditions triggered social revolutions in France, China and Russia (Skocpol 1979)? (2) What led to the 2003 Iraq War (Harvey 2012)? (3) Which factors explain the status quo of UK and US welfare states in the 1980s (Pierson 1994)? Consequently, PT is used to trace causal processes and discern intervening mechanisms underlying the occurrence of specific events which occur in a single case or in multiple cases. In other terms, PT is a causes-of-effects oriented method.

These differences may raise some doubts about how fruitful a dialogue between SC and PT could be. If PT is a causes-of-effects oriented method, how can it contribute to answering questions such as "What is the impact of that event (X)?". Analogously, if SC is an effects-of-causes-oriented method, how can it help us understand 'why did this event (Y) happen?'.

This paper intends to dispel these doubts and show that a fertile triangulation between SC and PT can be achieved, starting from the causal explanation one of the two methods provides and then evaluating how, and in which circumstances, the other can, from its standpoint, contribute to that explanation.

Basically, if PT starts from the event Y to conclude that it is caused by the event X through the mediation of the mechanism M, SC can be then adopted to counterfactually test the effect exercised by X and/or M on Y. More precisely, if SC is applicable to estimate the impact of X and/or M, the occurrence of Y can be evaluated in relation to the presence/absence of X and/or M.

On the other hand, if SC starts from the event X to estimate its impact on the outcome Y, PT can be employed to unfold the mechanism M through which Y arises. In other words, PT can complement SC by opening up the black box of the causal relation. This might appear to be not so dissimilar to the traditional role of PT in supplementing large-N (quantitative) comparison. However, in this case, the contribution does not concern the correlational logic pertinent to numerous cases, but causal relations pertinent to individual cases. This has specific implications. First, the black-box unpacking can be, in certain case, enhanced by performing PT analysis not only for the case exposed to the event

[3] Although SC was originally designed for estimating the effect of an event occurring in a single unit, it was subsequently modified to accommodate multiple units as well (for an overview, see Abadie 2019).

of interest, but also for the comparison units used by SC to construct the counterfactual case. Second, the identification of intervening mechanisms can be utilized to perform further SC analyses precisely aimed at estimating the effect of those same intervening mechanisms.

These ways of triangulation are illustrated through the paper without the pretension of developing a systematic guideline for combining the two methods. Instead, some leading studies of SC and PT scholarship are simply considered to highlight how these methods can (or cannot) be pursued.

The rest of the paper is structured as follows. In the next two sections, the basic elements of the two methods under examination are outlined in order to familiarize the reader with them. The fourth and fifth sections discuss how PT and SC can complement each other respectively. Section sixth concludes.

2. The SC method in a nutshell[4]

SC is, essentially, used in a between-case framework. In order to answer causal questions about the effects of a certain event, it prescribes comparing the case exposed to the event of interest, usually called the "treated unit", with an appropriately constructed comparison unit, called the "synthetic control unit"[5]. The construction of that unit is done in accordance with Mill's method of difference and by improving on traditional procedures used in comparative research for selecting control units. As is well-known, Mill's method of difference implies selecting a control unit with the same characteristics of the case of interest with the exception of the event under scrutiny. When these conditions are satisfied, the comparison is "controlled" and provides the functional equivalent of an experiment. Nevertheless, selecting one real case which meets these conditions is often a heroic task for comparativists. Consequently, students who follow traditional comparative strategies are frequently criticised for the arbitrariness of their selection. Hence, SC addresses these shortcomings by constructing the synthetic control unit through a data-driven control-group selection procedure (Abadie et al. 2015: 496-7).

That procedure is *de facto* based on the assumption that the characteristics of the treated unit can generally be much more accurately approximated by a combination of comparison units than by any single control unit. Accordingly, the synthetic control unit corresponds to a weighted average of the available control units. Specifically, the weights assigned to each control unit are chosen so that (1) the distance between the values for the treated unit and the synthetic one prior to the event under scrutiny are minimized with respect to a set of outcome (observed) predictors and (2) the outcome

[4] In consideration of the objectives of this paper, SC is introduced ignoring its econometric features (for a detailed discussion of these, see Abadie et al. (2010; 2015)).

[5] In accordance with the impact-evaluation literature, Abadie et al (2010; 2015) use the terms "treated unit" and "untreated" or "control unit" to refer to the case exposed and not exposed to the event under examination, respectively.

trajectories for the actual and the synthetic unit best resemble each other in the pre-treatment period (i.e. prior to the occurrence of the event of interest)[6]. Accordingly, the resulting synthetic control unit can best reproduce the characteristics of the actual case in the pre-treatment period.

In this way, SC not only facilitates comparative case studies based on Mill's method of difference, but also permits the reproduction of suitable counterfactuals: the evolution of the outcome variable for the synthetic control unit in the post-treatment period coincides with what would have occurred in the absence of the event under examination. In view of this, the estimation of the effect of that event is computed as the difference between the values for the outcome variable of the treated unit and the values of the variable outcome for the synthetic control unit in the treatment period, i.e. since the occurrence of the event under scrutiny (Abadie et al. 2010; 2015)[7].

The logic behind SCM can be better clarified by reference to its original application by Abadie and Gardeazabal (2003) in estimating the economic impact of the terrorist conflict in the Basque Country. In that study, the treated unit was the Basque Country since 1968, when ETA claimed its first victim. The outcome was measured via real per capita GDP. Accordingly, Y1 refers to the values of per capita GDP in the presence the terrorist conflict, and Y0 denotes the values of per capita GDP in the absence of that treatment. Hence, the yearly gap between Y1 and Y0 corresponds to the impact of the terrorism on per capita GDP. Nevertheless, since the yearly trajectory of Y0 is not obviously observable, it had to be reproduced.

To this end, an annual regional-level panel data for the period 1955-1997 was employed so that the pre-treatment period lasted from 1955 to 1967, while the treatment period lasted from 1968 to 1997. The dataset included the Basque country plus the other Spanish regions that constituted the set of potential comparison units. For the pre-terrorism characteristics, a standard set of economic growth predictors was employed. It included sectoral share, human capital, investment ratio, and population density.

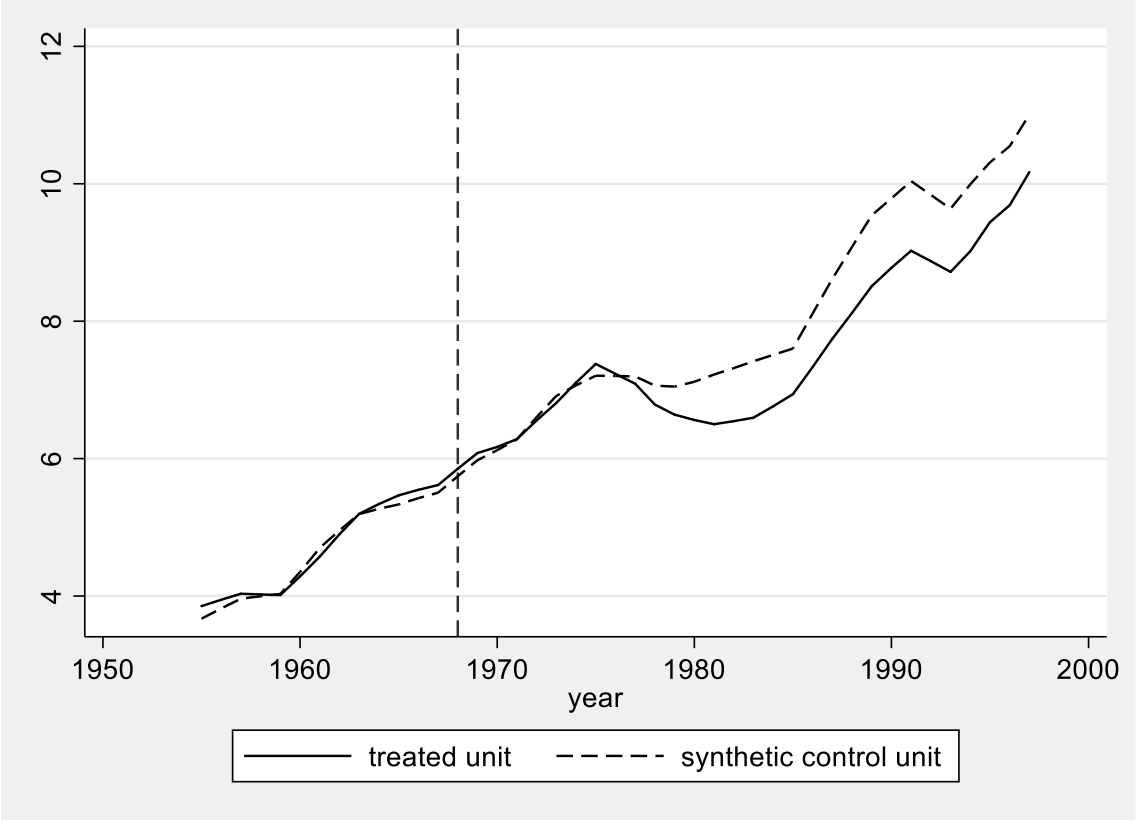
Using the data-driven procedure mentioned above, positive weights resulted for two regions, Catalonia and Madrid, with the values of 0.85 and 0.15 respectively, and the value of zero for the other potential controls. Accordingly, a synthetic Basque Country was constructed as a weighted average of Catalonia and Madrid, which best reproduced values of economic-growth predictors and

[6] The results of this second condition are more relevant. In fact, if the outcome trajectories for the actual and the synthetic unit are overlapped in the pre-intervention period, unobserved confounders are controlled for in addition to observed predictors. This is because only units that are alike in both observed and unobserved determinants of the outcome variable, as well as in the effect of those determinants on the outcome variable, should produce similar trajectories of the outcome variable over extended periods of time (Abadie et al. 2015).

[7] SC prescribes evaluating the uncertainty of the estimated effect via placebo techniques based on the principle of permutation inference (for more details, see Abadie et al. (2010, 2015)). For sake of simplicity, this issue is ignored through the paper.

per-capita GDP of the Basque Country (Y0). In particular, the per-capita GDP trajectory for the synthetic unit almost overlaps with that of the actual unit in the 1955-1967 period (see Figure 1). Consequently, the synthetic Basque Country provided reasonable approximation of the per-capita GDP that would have been experienced by that region over the treatment-period in the absence of terrorism. Hence, the Y1 and Y0 trajectories were compared over the 1968-1997 period to estimate the effect of terrorism. In doing this, Abadie and Gardeazabal (2003) found a pronounced negative effect. As Figure 1 shows, per-capita GDP of the two units trended similarly until 1975. However, from that year, when ETA's terrorist activity became a large-scale phenomenon, the two outcome lines diverged progressively and the actual one increasingly descended below the treated one ((Y1<Y0) denoting precisely the negative impact of terrorism.

Figure 1 – Per capita GDP for the Basque country and the synthetic Basque country (Abadie and Gardeazabal 2003).



As anticipated, the SC method proposed by Abadie and Gardeazabal (2003) and Abadie et al. (2010; 2015) was progressively developed by numerous econometricians in order to overcome its drawbacks. In line with the aim of this article, one of these drawbacks and the corresponding remedy must be mentioned. Standard SC suffers from a classical shortcoming often attributed to the impact

evaluation perspective: it provides a simple link between the treatment and the outcome variable, leaving the black box unopened. In other words, the standard version of the SC method provides an estimation of the total effect exercised by the event of interest, but does not make it possible to perform any causal mediation analysis. In the impact evaluation perspective, mediation analysis is a statistical framework which precisely permits us to take the fact that the impact exercised by the treatment variable on the outcome variable is mediated by an intervening variable, called the mediator (M), into account (e.g. Keele et al. 2015).

For this reason, Mellace and Pasquini (2019) recently proposed the mediation analysis synthetic control (MASC) method. It is *de facto* a generalization of the standard SC method and makes it possible to break down the total effect of a certain event into its direct and indirect components referring to a certain mediator. Specifically, through the MASC method the total effect can be broken down into the indirect effect, which goes through a mediator, and the residual effect. To compute these effects, three distinct outcome variables must be considered: Y00 representing the values of the outcome in the absence of both the treatment and the mediator, Y11 denoting the values of the outcome in the presence of both the treatment and the mediator, and Y01 representing the values of the outcome in the absence of the mediator but in the presence of the treatment. Hence, the total effect corresponds to the difference between Y11 and Y00, the residual effect to the difference between Y11 and Y01, and the indirect effect coincides with the difference between the total and the residual effect.

To illustrate the MASC method, Mellace and Pasquini (2019) estimated the effect of the adoption of the Euro (X) on labour productivity (Y) in several European countries and investigated the role of the level of specialization of exporting firms as a possible mediator (M)[8]. Through this exercise, they show that a reduction in the level of specialization helped mitigate the negative short term effects of adopting the new currency in some countries or boosted the positive effects in others (for another empirical application, see Section 4).

3. The PT method in a nutshell

In identifying the causes of a certain event, PT emphasizes the necessity of discerning the causal mechanism. In fact, PT is built around a scientific-realist understanding of explanation that highlights causal mechanisms. Causal mechanisms are defined in this literature as the causal processes and intervening steps between an initial cause and a final outcome. Specifically, since it is assumed that an outcome is generated by a sequential chain which includes an intricate blend of causes, any causal

[8] Since several European countries are affected by the Euro introduction, the analysis is performed for multiple-treated units.

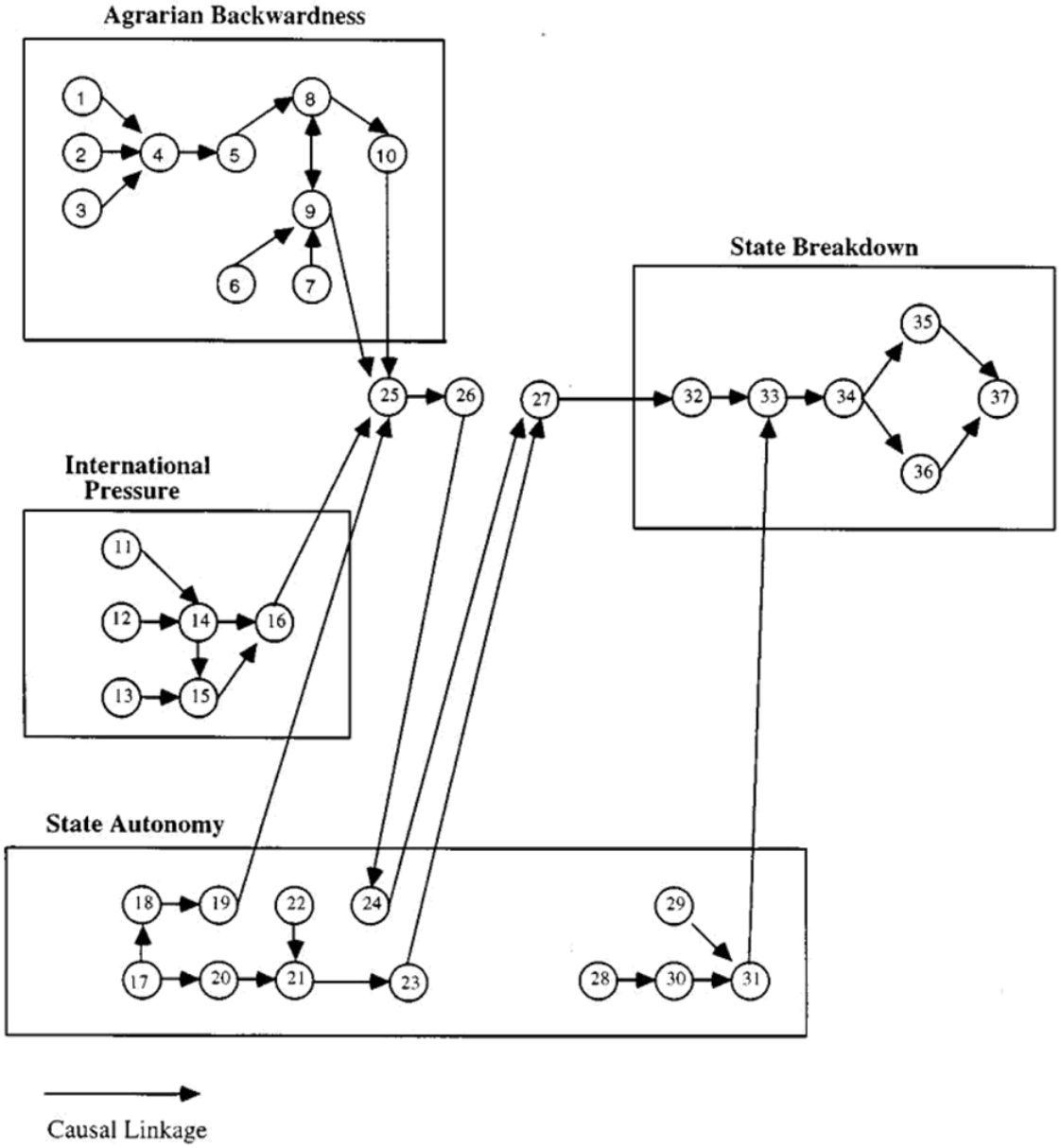
explanation is reached by identifying the mechanisms through which the outcome of interest is generated (Bennett 2008a; George and Bennett 2005; Hall 2006; Bennett and George 2001). In other words, given that the notion of causal chain and underlying process imply that there is not too much proximity between the initial cause and the outcome of interest, one may rely on the intermediating mechanisms to make causal inference (Mayntz 2004).

In order to trace processes, researchers usually rely on particular information, called causal process observations (CPOs). A CPO can be defined as ''[a]n insight or piece of data that provides information about context, process or mechanism, and that contributes distinctive leverage in causal inference" (Collier, Brady, and Seawright 2010). CPOs can thus be viewed as piece of evidence resulting from some combination or accumulation of empirical observations and other contextual information about certain processes occurred within a single case. In other words, CPOs reflect in-depth knowledge of sequential processes rather than data collected as part of a systematized array of variables. CPOs may be either qualitative or quantitative in nature. However, each piece of evidence is qualitatively different so that they remain reciprocally non-comparable (Gerring 2007).

Furthermore, causal inference is made via within-case analysis. In other words, PT is a within-case method irrespective of the number of cases considered. If a single case shows a strong positive result with respect to the outcome of interest (e.g. the occurrence of a revolution) this is selected and then individually examined. This is precisely because, with PT, causation is not established through cross-case comparison, but through uncovering traces which connect events over time within a context of a single case. Hence, even if more cases are contrasted to assess the relevance of certain factors in determining similar or different outcomes, PT implies that their processes are traced singularly and separately to explain their specific case-outcomes (Bennett and Elman 2006; Tarrow 2010).

A widely discussed example of this empirical strategy is the Skocpol's book *State and Social Revolutions* (1979). As Mahoney (1999: 1164) observed, in explaining the causes of China, Russia and France revolution, Skocpol not only adopted cross-case comparison, but also analysed those national outcomes as the product of unique, temporally ordered, and sequentially unfolding events that occur within cases. Therefore, she *de facto* adopted the PT approach. By reference to detailed historical evidence, complex causal chains for each revolutionary transformation in the three countries examined were elucidated in the book. In the case of France, Skocpol detected three main causal factors underlying the breakdown of state authority in the eighteenth century: agrarian backwardness, international pressure, and state autonomy. These factors, in turn, were disaggregated into thirty-seven discrete steps that connect structural causes to the outcome of interest. The corresponding overall causal process was accurately reproduced in an apposite diagram by Mahoney (1999) (see Figure 2).

Figure 2 – Causal chain showing conditions for state breakdown in France (Mahoney 1999).



Legend: 1. Property relations prevent introduction of new agricultural techniques; 2. Tax system discourages agricultural innovation; 3. Sustained growth discourages agricultural innovation; 4. Backwardness of French agriculture; 5. Weak domestic market for industrial goods; 6. Internal transportation problems; 7. Population growth; 8. Failure to achieve industrial breakthrough; 9. Failure to sustain economic growth; 10. Inability to successfully compete with England; 11. Initial military successes under Louis XIV; 12. Expansionist ambitions of state; 13. French geographical location vis-a`-vis England; 14. Sustained warfare; 15. State needs to devote resources to both army and navy; 16. Repeated defeats in war; 17. Creation of absolutist monarchy; 18. Dominant class often exempted from taxes; 19. State faces obstacles generating loans; 20. Socially cohesive dominant class based in proprietary wealth; 21. Dominant class possesses legal right to delay royal legislation; 22. Dominant class exercises firm control over offices; 23. Dominant class is capable of blocking state reforms; 24. Dominant class resists financial reforms; 25. Major financial problems of state; 26. State attempts tax/financial reforms; 27. Financial reforms fail; 28. Recruitment of military officers from privileged classes; 29. Military officers hold grievances against the crown; 30. Military officers identify with the dominant class; 31. Military is unwilling to repress dominant class resistance; 32. Financial crisis deepens; 33. Pressures for creation of the Estates-General; 34. King summons the Estates-General; 35. Popular protests spread; 36. Conflict among dominant class members in the Estates-General ; 37. Municipal revolution; the old state collapses.

In the diagram, causal linkages between each events of the causal chain are represented by arrows. It is superfluous to present a full summary of Mahoney's diagram here. However, the combination of property rights relations that prevented agricultural innovations (event 1), a tax system that

discouraged agricultural innovation (event 2), and sustained growth that discouraged agricultural innovation (event 3) conjointly caused agricultural backwardness (event 4). This latter was the cause of weak domestic market for industrial goods (event 5) which, in turn, produced the failure to achieve industrial breakthrough (event 8). And so on.

Like Skocpol (1979), numerous scholars aim to present an explanation of the entire causal chain, addressing and shedding light on each individual step. This is especially the case when PT is simply used to explain the occurrence of a certain outcome in a single case and not to test or build a hypothesis. On the other hand, when it is applied to a single case to test a more general hypothesis the focus is not on an articulated causal chain, but on only a few events. More precisely, when the existing literature indicates an association between event X (e.g., democracy) and event Y (e.g., peace), but there is uncertainty about the existence of a causal mechanism (e.g., political transparency) linking the two events, PT can be adopted to evaluate whether that mechanism actually links X and Y in a particular case (George and Bennett 2005; Blatter and Haverland 2012; Bennett and Checkel 2015; Beach and Pedersen 2013).

In this vein, Mahoney (2012) proposed a more structured procedure to use PT to test a causal hypothesis. Basically, he argued that the analyst should first combine pre-existing generalizations with specific observations (i.e. CPOs) from a single case and then evaluate the validity of the hypothesis via two tests commonly used in PT scholarship: the hoop test and the smoking gun test. A hoop test prescribes that a certain CPO must be present for a hypothesis to be valid. Failing a hoop test rejects a hypothesis but passing a hoop test does not corroborate a hypothesis. Conversely, smoking gun tests prescribes that if a certain CPO is present, then the hypothesis must be valid. Passing a smoking gun test lends decisive support in favour of a hypothesis, though failing a smoking gun test does not eliminate a hypothesis (Van Evera 1997; Bennett 2008a; Bennett 2010). Since a mechanism M is normally more closely situated to X and/or Y, Mahoney (2012) recommended using these two tests referring to CPOs about the relation between X and M or M and Y, rather than pieces of evidence about the direct link between X and Y.

4. Complementing SC analysis via PT method

This section aims to elucidate how PT can be used for complementing an impact estimation obtained via SC method. To this end, two applications of this latter method are contrasted. They are the study on the economic impact of German re-unification (Abadie et al. 2015) and the already discussed analysis of the economic impact of the terrorist conflict in the Basque Country (see section 2). They are chosen because they make clear a remarkable difference regarding the supplementary use of

causes-of-effects explanation and, thus, they precisely demonstrate how the adoption of that explanation can complement a SC study.

The study of the economic impact of German re-unification was included in a methodological paper essentially aimed at discussing the use of the SC method in comparative research. Consequently, the authors introduced that SC analysis as a mere application of the more general SC method. The estimation of impact exercised by German re-unification on the selected outcome, i.e. per capita GDP for West Germany was thus illustrated simply to describe the steps one must take to implement the SC method for a comparative case study. No information about the mechanisms through which the event of interest exercised its impact was provided. Only some previous studies about the economic costs of the re-unification for West Germany were mentioned. As a result, the black box of the causal relation of interest was left closed.

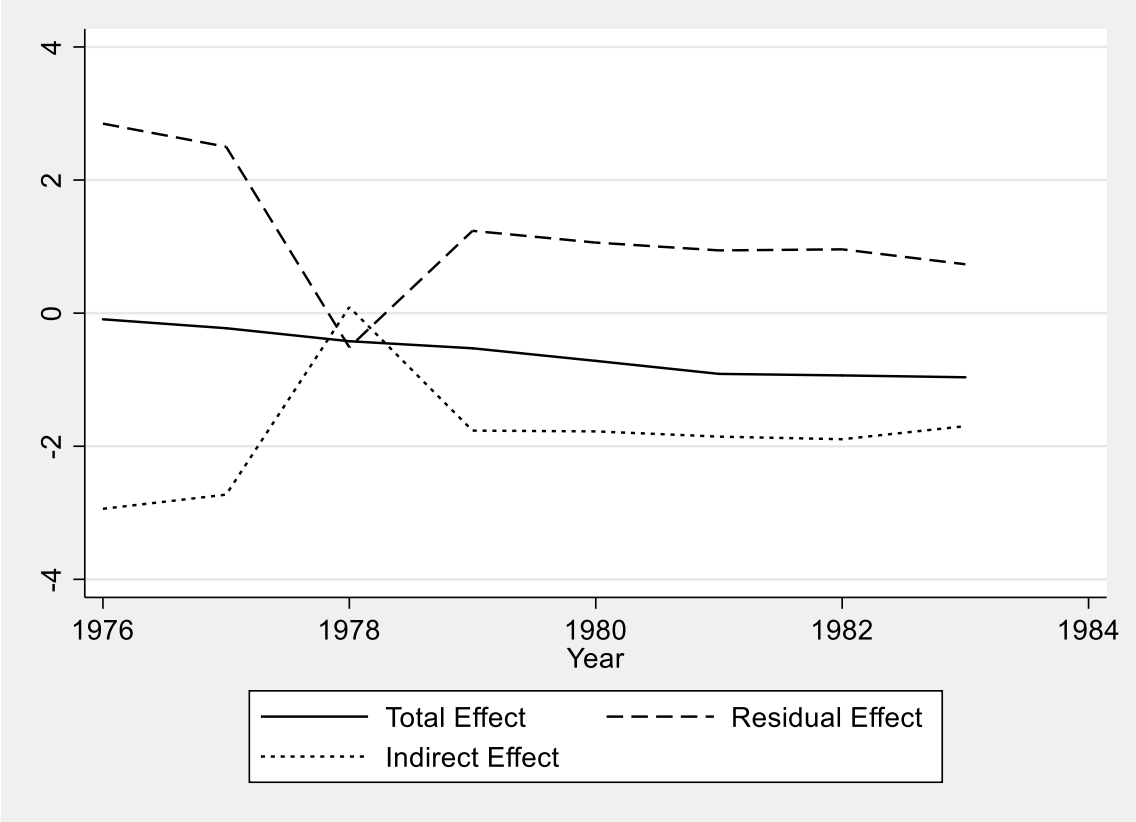
On the other hand, the study of the economic impact of terrorism in Basque Country was included in a substantive article. Although SC was applied for the first time, Abadie and Gardeazabal (2003) remained mainly engaged with their research topic. In performing this case study, these authors, quite surprisingly, made several steps markedly consistent with the recommendations regarding using PT as a method for testing hypotheses (see section 3). First, they argued that the terrorism in the Basque Country was used as a case study to investigate the more general relationship between political conflict and economic growth. Hence, Abadie and Gardeazabal combined pre-existing generalizations with specific observations of the case of interest. Concerning pre-existing generalizations, they referred to the literature about the association between political stability/conflict and economic performances. Second, and more importantly, in addition to the SC estimation of the effects of the terrorism activity, they provided a qualitative causes-of-effects account, i.e., a brief history of the Basque Country over the treatment period of 1968-1997 (see Section 2). In providing that account, these authors took a further important step in line with the PT scholarship: making causal inference providing evidence about intervening mechanisms regarding the relationship between an initial cause and a subsequent outcome. Taking into account that, in this case study, terrorism represents the initial cause X and economic growth the subsequent outcome Y, Abadie and Gardeazabal presented some relevant pieces of evidence about the intermediating mechanism M. Specifically, they noted that, in order to finance its operations, ETA used kidnappings-for-ransom, extortion, and, less frequently, robberies. The main targets of such money-raising activities were Basque entrepreneurs, who consequently started to abandon the Basque Country. In addition, these authors cited some literature as evidence that the terrorist conflict deterred domestic and foreign direct investment in the Basque Country. Therefore, the economic downturn caused by terrorist conflict

would be mediated by a progressive reduction of investment, in turn, due to pressures (directly or indirectly) suffered by local, national or foreign entrepreneurs.

Comparing these two studies, what can one conclude about the use of PT to supplement a SC analysis? If one refers essentially to the reliability of SC results, a complementary causes-of-effect explanation will appear superfluous. The total effect exercised by German re-unification is *per se* reliable. In this sense, the SC method can be regarded as self-sufficient. Therefore, one might wonder what contribution the causes-of-effect account about the Basque Country may offer to the corresponding SC analysis. The causal inference reached via that account may appear as the standard contribute of PT to a more general association between two events X and Y (see section 3). The causal-mechanism unfolded for the Basque case serves to demonstrate how political conflict and economic development are linked in a single case. Nonetheless, the SC analysis performed on this region plays quite a similar role with respect to the general association between political conflict and economic development. It provided evidence of the negative economic impact of the terrorist conflict. In other words, as case studies, both SC and PT shed light on a more general association. PT contributes with an effects-of-causes account, while SC contributes a causes-of-effects account. However, one should, at this point, wonder how the two methods can speak to each other to further develop the case-study of interest. The causal mechanism unfolded by Abadie and Gardeazabal (2003) *de facto* opens the door to further SC analysis. In particular, since terrorist tensions would have progressively reduced (local and foreign) investment (see above), a MASC analysis (see section 2) can be performed to estimate to what extent the impact exercised by terrorism on per capita GDP was mediated by that intervening variable.

Figure 3 precisely shows that the total and negative impact arisen since the mid-1970s (see section 2) was essentially due to the more pronounced negative indirect effect which went through investment. Nevertheless, the figure also reveals quite a surprising positive residual effect which in part compensates the detrimental impact attributable to the indirect effect. Clearly, this residual effect could work through other, possibly unobserved, causal pathways. Their identification goes beyond the aim of this paper but indicates that a further causes-of-effects analysis might be performed to precisely discern those causal pathways.

Figure 3 – Total, indirect and residual effect of terrorism in the Basque Country.



Furthermore, if PT analysis can supplement the SC method in identifying causal mechanism, that identification can be strengthened by providing causes-of-effect accounts for the comparison units of the case of interest too. In other words, one may perform what Tarrow (2010) calls 'dual-process tracing' in order to trace and then contrast causal mechanisms of treated and comparison units. This would, moreover, make PT more compatible with SC logic. In fact, although processes would be traced for these units in accordance with within-case logic, a between-case analysis would be also carried out for them. Clearly, the selection of comparison units is, in this case, ensured by the transparency of the SC method – i.e. the explicit contribution of each potential comparison unit in constructing the synthetic unit. As Abadie et al (2015: 508) themselves note, by explicitly specifying the set of units that are used for comparison, the SC method does not preclude but facilitates detailed qualitative analysis and comparison between the case of interest and the set of comparison units selected by the SC method.

Accordingly, since the synthetic Basque Country was constructed as a weighted average of Catalonia and Madrid (see section 2), the socio-economic processes that occurred in these two regions during 1968-1997 period (i.e. the treatment period for Basque Country) may be traced and then contrasted with the corresponding processes for Basque Country. Specifically, given that these two regions did

not experience terrorism in that period, this could make it possible to evaluate whether they denoted opposite causal mechanisms from those observed in Basque Country. More precisely, one could evaluate whether the relationship between political stability and economic growth observed in those regions was mediated by a significant permanence of local entrepreneurs as well as relevant domestic and foreign investments – i.e. the contrary of what happened in the Basque Country. Moreover, an in-depth examination of these two regions might also help to discern the causal pathways underlying the above mentioned residual effect estimated for the Basque Country.

Nonetheless, one must note that the study on the economic impact of the terrorist conflict in the Basque Country is extremely lucky to perform this kind of triangulation between SC and PT. Only two regions were involved in constructing the synthetic Basque and one of these, Catalonia, was given a very high weight (0.85). This means that this region alone quite adequately approximates the most relevant characteristics of Basque Country, excluding terrorism.

However, exploiting potential control units to perform further PT analyses is conditioned by their number and, more importantly, by the values of their weights. In fact, in several cases more controls units are used to construct the synthetic unit and their weights are quite low, denoting limited similarity with the treated unit. For instance, in the analysis of the economic impact of German re-unification, the synthetic West Germany was given by a weighted average of Austria (0.42), US (0.22), Japan (0.16), Switzerland (0.11) and the Netherlands (0.09) – with the weights in parentheses.

5. Complementing PT analysis with SC method

This section shows how and under what conditions SC can be adopted to enhance PT analysis. Although a causal chain reconstructed through PT can be regarded as valid *per se*, some criticism may arise. One reason derives from the fact that PT makes causal inference via a mere tracing of actual process. Furthermore, even when a causal relation is assessed via the hoop test and the smoking gun test (see section 3) scepticism may persist. This is because these tests are far more subjective than standard statistical tests. They lack the precision and decisiveness of formal tests (Skarbek 2020).

Accordingly, several PT developers suggest counterfactual analyses can validate the causal inference made via PT. In fact, in counterfactual framework, the outcome is specified in case of both X and not-X (Kay and Baker 2015; Blatter and Haverland 2012; Collier 2011).

Using the SC method is clearly one way of accomplishing this[9]. It makes it possible to construct synthetic units which precisely represent what would have occurred in the absence of the event/cause of interest (see section 2). Furthermore, since SC is inherently designed to estimate the impact of an

[9] Other procedures are employed to construct counterfactual cases in case-study explanation (e.g. Fearon 1991; Levy 2015; Mahoney and Barrenechea 2019). Here, the objective is certainly not to contrast SC with those other procedures or identify their respective merits and drawbacks. Another paper would be necessary to achieve that.

event that occurred after a particular point of time, it appears suitable to the sequential logic underlying PT. Specifically, SC can be employed to evaluate the occurrence of a certain outcome in relation to the presence/absence of a putative cause identified via PT. Specifically, if SC demonstrates that that cause is irrelevant for the occurrence of the outcome, PT analysis will have to be reviewed.

However, SC is not useful for counterfactually testing any causes-of-effects account. A first reason concerns data restriction. PT analyses can be carried out for a large range of cases due to the limited constraints in recovering CPOs. On the other hand, the lack of quantitative data limits the opportunity to perform SC analysis. For instance, since some PT analyses regard social or political events that happened in the remote past, the corresponding time series necessary to perform a pertinent SC analysis may be unavailable. This is almost certainly the case for the widely cited study by Skocpol (1979). To the best of my knowledge, the existing data do not allow us to perform a SC analysis to test the causes-of-effects explanation developed for the French case (see section 3).

The complexity of the causal chain reconstructed for the French case makes it possible to highlight another constraint in using SC method for complementing PT studies. Specifically, even if one adopts MASC instead of the standard SC, a single SC analysis cannot enable us to assess an entire and articulated causal chain like that of Figure 2 (see section 3). SC results are more suitable for evaluating single causal relations of more complex causal chains or, as illustrated in section 4, to assess causal relations between a few events.

Moreover, outcome operationalization and some problems of comparability between treated and control units can involve other problems of SC applicability. To show this, Harvey's (2012) study of the 2003 Iraq War can be considered. Through PT analysis, this author provided a detailed tracing of all key decisions, choices and tactics adopted by George W. Bush during that period. The reconstructed sequence of decisions that led to war, i.e. the outcome of interest, is definitely complex. Therefore, in line with the argument developed before, SC method is assessed in testing the direct impact of George W. Bush's presidency on the invasion of Iraq. The corresponding counterfactual – also analysed by Mahoney and Barrenechea (2019) – is: 'If George W. Bush had not been elected president, the United States would not have invaded Iraq'.

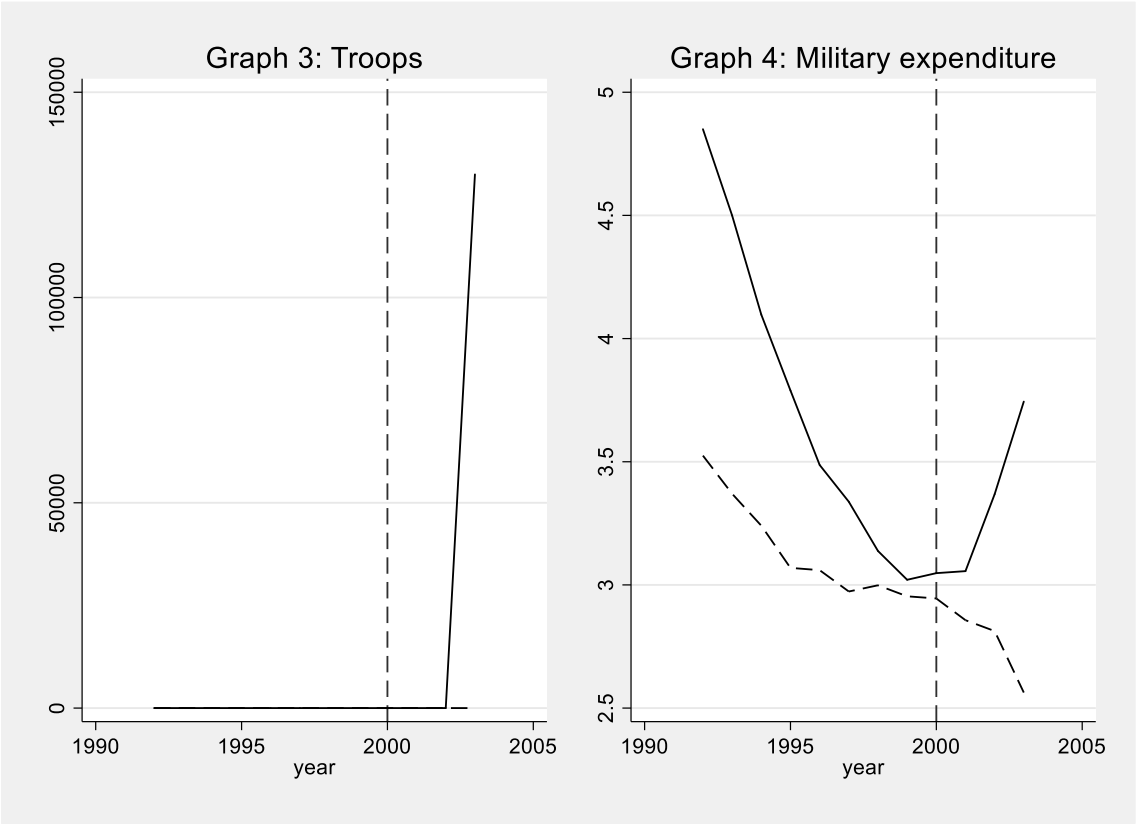
Since the SC method inherently works with continuous variables for measuring the outcome, it is impracticable when referring to a dichotomous variable denoting the occurrence vs. the not-occurrence of US invasion of Iraq. Nonetheless, proxying that event via quantitative indicators involves serious problems as well. In fact, an obvious solution would be using the number of US troops who arrived in Iraq. This would imply that the corresponding US time series would denote zero values for the pre-treatment period (e.g. from 1992 to the George W. Bush 2000 election) as well as for 2001-02 period and a value of about 130,000 for 2003. Using SC method to reproduce that

trajectory would be nonsensical. This is because, if one considers a set of OECD countries as potential control units, all the corresponding time series for the outcome variable would be entirely composed by zero values. Consequently, any weighted average of the available control units to best reproduce the outcome trajectories for US would end up mathematically irrelevant precisely because it would produce the same result. However, beside any SC calculation, Graph 1 of Figure 4 reproduces the logic of this hypothetical counterfactual exercise. The counterfactual and US trajectory for the outcome variable would be overlapped until 2003. From that year, the two lines would diverge, but the effect they denote would be trivial. In the absence of George W. Bush, none of the US troops would have been arrived in Iraq, so the Bush impact would correspond to 130,000 troops (130,000 - 0).

Serious problems would also arise if military expenditure as a percentage of GDP were employed to measure the outcome of interest. In this case, the US trajectory for the outcome variable would not be synthetically reproducible because, during the pre-treatment period, the US level of military expenditure was, on average, significantly higher than that of any potential control unit. Moreover, the Bush impact on the Iraq war specifically would be indiscernible. This is because, after a prolonged period of decrease, US military expenditure had already begun to increase again since 2001 as a consequence of the Afghanistan war (see Graph 2 of Figure 4).

Using SC to test the above-mentioned counterfactual would also imply a problem of unit comparability. The US, the treated unit, experienced the Bush presidency from 2000 onwards. Consequently, in order to know what would have occurred in its absence, the synthetic US should be constructed using a reasonable set of potential controls units. Assuming that the negation of the actual antecedent, i.e. George W. Bush election is not equivalent to the election of his rival, Al Gore, a certain set of affluent democracies could be used as potential controls. Nonetheless, one could argue that a problem of comparability between treated and potential control units arises: no other affluent democracy could experience the election of George W. Bush. This is certainly true, even if a similar problem was ignored by Abadie et al (2015) in constructing the synthetic West Germany (see above). They used a sample of OECD countries as potential comparisons even if none of them could have undergone a reunification like that experienced by Germany. However, another problem of comparability emerges in our case. Using a sample of affluent democracies as a potential control group, one may reasonably claim that most Western democracies rarely initiate a war, but participate as US allies much more frequently.

Figure 4 – Iraq war



That said, let us consider a causes-of-effects study counterfactually test via SC method. It is Pierson's (1994) analysis of Reagan and Thatcher's welfare state retrenchments, counterfactually replicated by Podestà (2020). Pierson observed that at the end of the so-called conservative resurgence of 1980s, when Thatcher and Reagan left power, the UK and US welfare states ended up substantially unaltered. That outcome was attributed to dense networks of interest groups developed around numerous welfare policies. Tracing and contrasting the decision-making of several social programs, Pierson showed that retrenchments were only successful for those programs with less organized recipients. In this sense, Reagan and Thatcher's plans to dismantle the welfare state can be regarded as the initial cause X, while the mobilization of organized recipients constitutes the intermediating mechanism of the process M. However, Pierson's analysis did not provide any evidence about the extent to which the retrenchments would have been implemented in the absence of influential groups. Analogously, we do not know whether the retrenchments would have been substantially the same or even smaller in the absence of conservative governments. Stressing the difficulty in testing the first counterfactual, Podestà (2020) performed a SC analysis to test the second one. Specifically, he argued that finding developed countries without organized supporters of the welfare state to be used as potential control units is *de facto* unfeasible. On the other hand – he claimed – one may reasonably assume that, during the 1980s, the Reagan and Thatcher administrations were the only examples of the conservative

resurgence[10]. Accordingly, the US and the UK were regarded as the sole treated units, while a sample of developed countries was used as the set of potential controls.

Furthermore, since the SC method inherently works with continuous variables to measure the outcome, six quantitative indicators (i.e. social security transfers, welfare state generosity, unemployment-benefit generosity, pension generosity, total government revenue, and union density) were adopted to operationalise some of the policy areas examined by Pierson. Hence, the trajectories of those indicators, observed in the presence of a conservative administration, were contrasted with corresponding trajectories that were reconstructed in the absence of such a political circumstance.

The results of this replication broadly confirmed Pierson's substantive conclusion: the conservative revolution of the 1980s did not substantially alter the UK or US welfare state. More precisely, in many policy arenas, the two conservative governments did not provoked noticeable retrenchment. Nevertheless, the impacts estimated for specific components of the two welfare states are, in some cases, different from the conclusions stressed in Pierson's book. For instance, while Pierson concluded that unemployment-insurance benefits have undergone extensive retrenchment in both countries, Podestà's SC analysis revealed that if the UK and US had not undergone a conservative revolution, their unemployment generosity would have not been significantly different. These differing findings probably depend on the fact that (1) the causal chain was not entirely counterfactually tested and (2) the quantitative indicators used in the SC analysis differ from the qualitative information adopted by Pierson. In any case, these differences question some of Pierson's conclusions and require some reviews of his causal explanation. As argued above, if a SC analysis highlights a causal effect inconsistent with the PT explanation, this latter should be amended accordingly.

6. Conclusion

Despite increasing efforts to combine large-N (quantitative) and small-N (qualitative) comparative methods, this paper has discussed potential ways of triangulation between two methods currently central in case-study research: SC and PT. To put these two methods in dialogue with each other, the starting point has been not their affinities, but rather a substantial divergence between them and more generally between quantitative and qualitative traditions: the causes-of-effects versus effects-of-causes perspective. SC starts from an event that occurred in a certain case and – assuming it to be a cause – estimates its impact on one or more outcomes. Conversely, PT starts from an event and – considering it as an outcome – discloses the causal chain which generated it.

[10] Pierson (1994: 4) himself admitted that the Thatcher and Reagan administrations constitute a crucial test of the welfare state's status because, for the first time since before World War II, two political executives were openly critical of central features of social policy.

Hence, a basic rule in triangulating the two methods is to start from the causal explanation reached via one of the two methods and then examine that explanation using the other method, i.e. looking at the causal relation from the opposite standpoint. Accordingly, once the causes of an event are identified via a PT analysis, this latter can be complemented and validated via a SC analysis by estimating the effects of those putative causes. On the other hand, once the impact of a certain event on a given outcome is estimated through a SC analysis, supplementary information about causal mechanisms can be provided via PT and hence, the effects-of-causes analysis can be refined. These ways of triangulation mean that the analyst must not necessarily start with one method and then proceed with the other. One may start with a effects-of-causes via SC and then proceed with a causes-of-effects perspective via PT or vice versa. Therefore, if the combination of the two methods is successfully performed, the causal relation linking the events of interest can be viewed from opposite standpoints.

However, some studies considered in the previous pages reveal that the combination between SC and PT cannot be systematically applied. This essentially depends on some constraints in implementing the SC method. The assessment of Skocpol's (1979) and Harvey's (2013) studies demonstrated that data unavailability, problems of comparability between treated and control units, difficulty of outcome operationalization and the complexity of the causal-chain limit the adoption of SC to complement PT analysis. Therefore, the possible methods of triangulation sketched in the previous pages depend on the research topic under examination.

Beyond this, a more extensive exchange between SC and PT depends on the stance of the developers of each the two methods towards the other approach. SC scholars appear quite open to embracing PT. The causes-of-effects account on the Basque Country provided by Abadie and Gardeazabal (2003) constitutes a good example of this predisposition. Signals of this propensity are also traceable into the fact that researchers practicing quantitative and experimental approaches are currently looking at causal mechanisms (Gerring 2010). On the other hand, several PT students invocate counterfactual perspectives to complement their explanations. Nonetheless, they appear to be somewhat hesitant to employ SC. To the best of my knowledge, none of the leading PT developers have yet employed this kind of counterfactual strategy.

References

Abadie, A. (2019). Using synthetic controls: Feasibility, data requirements, and methodological aspects. *Journal of Economic Literature*.

Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *Journal of the American statistical Association*, *105*(490), 493-505.

Abadie, A., Diamond, A., & Hainmueller, J. (2015). Comparative politics and the synthetic control method. *American Journal of Political Science*, *59*(2), 495-510.

Abadie, A., & Gardeazabal, J. (2003). The economic costs of conflict: A case study of the Basque Country. *American economic review*, *93*(1), 113-132.

Beach, D., & Pedersen, R. B. (2019). *Process-tracing methods: Foundations and guidelines*. University of Michigan Press.

Bennett, A. (2008). Process tracing: A Bayesian perspective. in J. Box-Steffensmeier, H. E. Brady, and D. Collier (Eds.), *The Oxford handbook of political methodology.* 702-21. Oxford: Oxford University Press.

Bennett, A. (2010). Process tracing and causal inference. in H. Brady and D. Collier (Eds.), *Rethinking social inquiry,* 207-19. Lanham MD: Rowman and Littlefield.

Bennett, A. & Checkel, J.T. (2015). Process tracing: From philosophical roots to Best practices, in A. Bennett and J.T. Checkel (Eds.), *Process tracing in the social sciences*: *From metaphor to Analytic tool*, 3–37. Cambridge: Cambridge University Press.

Bennett, A., & Elman, C. (2006). Qualitative research: Recent developments in case study methods. *Annual Review of Political Science*, *9*, 455-476.

Bennett, A, & George, A.L. (2001). Case studies and process tracing in history and political science: Similar strokes for different foci. in C. Elman & M.F. Elman (Eds.), *Bridges and boundaries: Historians, political scientists, and the study of international relations,* 137–166. Cambridge, MA: MIT Press.

Blatter, J. & Haverland, M. (2012). *Designing case studies. Explanatory approaches in small-N research.* Houndmills: Palgrave Macmillan.

Collier, D. (2011). Understanding process tracing. *Political Science and Politics*, 44(4): 823–30.

Collier, D., Brady, H.E. & Seawright, J. (2010). Sources of leverage in causal inference: Toward an alternative view of methodology. in H. E. Brady, & D. Collier (Eds.). *Rethinking social inquiry*, 161-99. Lanham MD:Rowman and Littlefield.

Fearon, J. (1991). Counterfactuals and hypothesis testing in political science. *World Politics*, 43(2), 169-95.

Gerring, J. (2007). *Case study research*. Cambridge: Cambridge University Press.

Gerring, J. (2010). Causal mechanisms: Yes, but…. *Comparative Political Studies*, 43(11), 1499-1526.

George, A.L. (1979). Case studies and theory development: The method of structured, focused comparison. in P.G. Lauren (Ed.), *Diplomacy: new approaches in history, theory, and policy.* 43–68. New York: Free Press.

George, A.L. & Bennett, A. (2005). *Case studies and theory development in the social sciences*. Cambridge, MA: MIT Press.

Hall, P. A. (2006). Systematic process analysis: When and how to use it. *European Management Review*, 3, 24-31.

Harvey, F. P. (2012). Explaining the Iraq war: Counterfactual theory, logic, and evidence. Cambridge: Cambridge University Press.

Kay, A., & Baker, P. (2015). What can causal process tracing offer to policy studies? A review of the literature. Policy Studies Journal, 43(1), 1-21.

Keele, L., Tingley, D. & Yamamoto T. (2015). Identifying mechanisms behind policy interventions via causal mediation analysis. *Journal of Policy Analysis and Management*, 34(4), 937–963.

Kuehn, D., & Rohlfing, I. (2016). Are there really two cultures? A pilot study on the application of qualitative and quantitative methods in political science. European Journal of Political Research, 55(4), 885-905.

Levy, J.S. (2015). Counterfactuals, causal inference, and historical analysis. Security Studies, 24, 378–402.

Mahoney, J. (1999). Nominal, ordinal, and narrative appraisal in macro-causal analysis. *American Journal of Sociology*, 104(4), 1154–96.

Mahoney, J. (2012). The logic of process tracing tests in the social sciences. *Sociological Methods & Research*, 41(4), 570–597.

Mahoney, J., & Barrenechea, R. (2019). The logic of counterfactual analysis in case-study explanation. *The British journal of sociology*, 70(1), 306-338.

Mahoney, J., & Goertz, G. (2006). A tale of two cultures: Contrasting quantitative and qualitative research. *Political analysis*, 227-249.

Mayntz, R. (2004). Mechanisms in the analysis of social macro-phenomena. Philosophy of the Social Sciences, 34(2), 237-259.

Mellace, G. & Pasquini A. (2019). Identify more, observe less: Mediation analysis synthetic control. arXiv e-prints, arXiv:1909.12073.

Pierson, P. (1994). Dismantling the welfare state? Reagan, Thatcher, and the politics of retrenchment. Cambridge: Cambridge University Press.

Podestà, F. (2020). Did the conservative resurgence matter? A counterfactual replication of Pierson's analysis on Reagan and Thatcher's welfare state retrenchments. *Comparative European Politics*.

Skarbek, D. (2020). Qualitative research methods for institutional analysis. *Journal of Institutional Economics*, 16(4), 409-422.

Skocpol, T. (1979). States and social revolutions: A comparative analysis of France,Russia, and China. Cambridge: Cambridge University Press.

Tarrow, S. (2010). The strategy of paired comparison: Toward a theory of practice. Comparative Political Studies, 43(2): 230-259.

Van Evera, Stephen. 1997. Guide to Methods for Students of Political Science. Ithaca, NY: Cornell University Press.