Political Geography 41 (2014) 1–10

Contents lists available at ScienceDirect

Political Geography

journal homepage: www.elsevier.com/locate/polgeo

Matched wake analysis: Finding causal relationships in spatiotemporal event data

Sebastian Schutte [a,*], Karsten Donnay [b]

[a] Center for Comparative and International Studies, ETH Zurich, 8092 Zurich, Switzerland
[b] Sociology, Modeling and Simulation, ETH Zurich, 8092 Zurich, Switzerland

ARTICLE INFO

Article history:
Available online 12 April 2014

Keywords:
Quantitative methods
GIS analysis
Civil war
Indiscriminate violence
Causal inference

ABSTRACT

This paper introduces a new method for finding causal relationships in spatiotemporal event data with potential applications in conflict research, criminology, and epidemiology. The method analyzes how different types of interventions affect subsequent levels of reactive events. Sliding spatiotemporal windows and statistical matching are used for robust and clean causal inference. Thereby, two well-described empirical problems in establishing causal relationships in event data analysis are resolved: the modifiable areal unit problem and selection bias. The paper presents the method formally and demonstrates its effectiveness in Monte Carlo simulations and an empirical example by showing how instances of civilian assistance to US forces changed in response to indiscriminate insurgent violence in Iraq.

Introduction

The study of political violence has benefited in recent years from a rapid increase in the availability of conflict event data sets (Raleigh, Linke, & Hegre, 2010; Sundberg, Lindgren, & Padskocimaite, 2010). In these data, single instances of violence are coded together with their geographic coordinates and the date they occurred on. Several recent publications have successfully shed light on some of the micro-dynamics of civil conflict by analyzing such data (for example Buhaug, 2010; Hegre, Østby, & Raleigh, 2009; O'Loughlin & Witmer, 2011; Raleigh & Hegre, 2009). However, while progress has been made in relating conflict intensity to geographic conditions, more complex endogenous mechanisms that drive conflict at the micro-level remain largely elusive to quantitative analysis, despite their theoretical prominence (e.g. Kalyvas, 2006).

To fill this gap, we introduce a novel approach to causal inference in disaggregated event data that combines two techniques for ensuring robust and clean causal inference: sliding spatio-temporal windows (Braithwaite & Johnson, 2012; Kulldorff, 1997) and statistical matching (Iacus, King, & Porro, 2012; LaLonde, 1986; Rubin, 1973). The presented approach clears the path for answering a whole class of high-profile research questions regarding the causal effects of specific types of events on future events. To demonstrate this approach and its capabilities, we show that the experience of indiscriminate insurgent violence in Iraq has led civilians to collaborate with the US military.

While presented in the context of conflict research, this method could be equally applied in other quantitative fields of research that rely on georeferenced event data: Criminologists might investigate the effects of law enforcement activities on subsequent levels of crime. Epidemiologists could analyze the spread of infectious disease as a function of specific types of interaction between individuals.

This paper proceeds as follows: After discussing the existing research and its shortcomings in the next section, we introduce our methodological contribution in detail and use a series of Monte Carlo simulations to test its capabilities and limitations. After that, we demonstrate the method in an empirical example by analyzing the effects of indiscriminate insurgent violence on civilian collaboration with US troops in Iraq.

Abilities and limitations of existing approaches

The theoretical prominence of endogenous conflict dynamics (Kalyvas, 2006) has motivated a number of empirical studies in recent years. In order to understand how past conflict events shape future levels of violence, a rapidly growing number of studies rely on newly available event data (see: Leetaru & Schrodt, 2013; Raleigh et al. 2010; SIGACT, 2010; Sundberg et al. 2010).

* Corresponding author. Tel.: +41 44 632 67 87.
E-mail addresses: schuttes@ethz.ch (S. Schutte), kdonnay@ethz.ch (K. Donnay).

http://dx.doi.org/10.1016/j.polgeo.2014.03.001
0962-6298/© 2014 Published by Elsevier Ltd.

In principle, event data reflect changes in the trajectory of conflicts brought about by specific incidents. Along these lines, research into the causes and effects of violence against civilians in civil war (Kalyvas, 2006; Lyall, 2009) and escalation dynamics (Haushofer, Biletzki, & Kanwisher, 2010; Jaeager & Paserman, 2008; Linke, Wittmer, & O'Loughlin, 2012) has drawn on conflict event data. Several studies have used village-level counts of violent events to investigate whether indiscriminate incumbent violence has a deterrent or escalating effect on subsequent insurgent activity. Especially Lyall (2009) and Kocher, Pepinsky, and Kalyvas (2011) pioneered this type of analysis with innovative matching designs and villages as units of analysis.

However, in many situations such natural spatial units of analysis are missing. Some studies have circumvented this problem by relying on artificial units of analysis, such as grid-cell months, and aggregated event counts and covariates accordingly. While introducing these artificial units conveniently clears the way for econometric analysis, it also leads to two problems widely described in the methodological literature. First, if cells of arbitrary sizes are the units of analysis, the number of available observations directly scales with the chosen cell size: the smaller the cells, the more observations. Of course, regular null hypothesis tests crucially depend on the number of available observations. As *N* increases, the standard errors tend to decrease and even the smallest empirical signals becomes statistically "significant". A second problem extensively described in the geographic literature is the "modifiable areal unit problem" (MAUP), i.e. the fact that the selection of artificial cell sizes drives spatial inference (Cressie, 1996; Dark & Bram, 2007; Openshaw, 1984).

Approaches to overcoming the MAUP have been proposed in the past and also been applied in conflict research (O'Loughlin & Witmer, 2011). A commonly used method called "SaTScan" (Kulldorff, 1997) relies on sliding spatial and temporal windows to reveal clusters of events on different levels of aggregation.[1] Applied to epidemiology, SaTScan was originally introduced as a tool for testing whether a certain region faces an elevated per capita risk of disease. The method provides a fast assessment of whether event clusters could have been brought about by chance under corresponding distributional assumptions. To establish a baseline level of clustered events, SaTScan applies a simulation technique: For each size of the spatiotemporal window under consideration, the software allocates events randomly in space and time. Repeating this process in multiple iterations generates a distribution of simulated events under baseline assumptions. Significant empirical deviations from this baseline can then be identified for different cell sizes. In other words, comparing the distribution of artificial events to the empirical record yields an estimate of how likely is it that observed clustering was brought about by chance.

In the epidemiological case of Kulldorff (1997), this baseline is well justified as it assumes a constant per capita rate of instances of non-infectious disease. In conflict settings, however, finding suitable baselines is usually much more difficult. Instances of insurgent violence, for example, are likely to result from a host of factors, including geographic exposure and reaction to previous violence. Randomly allocating events in space and time might not adequately capture plausible counterfactual scenarios: Instances of violence against civilians, for example, might be simulated to take place in uninhabited areas and a simulated baseline would not reflect the causal order of events found in the empirical record.

Relaxing the assumption of a uniform spatial distribution of events, Braithwaite and Johnson (2012) apply a permutation test within the framework of sliding spatiotemporal windows to the analysis of violent events in Iraq. In this setup, a random baseline is also simulated, but not by relocating conflict events in space and time. Instead, events remain in their original positions but event categories are randomly swapped. By holding constant the location and timing of events while changing event categories, a baseline scenario can be established in which event types are independent of one another. Comparing this simulated baseline to empirical distributions of event categories shows whether or not specific classes of events tend to occur together, i.e. in clusters that are unlikely to have been brought about by chance. However, this measure of systematic co-occurrence, as well as SaTScan's identification of event clusters, does not establish a clear causal relationship between the event types.[2] We therefore decided to introduce a new framework for inferential analysis in conflict event data.

In the following section we describe a new method called *Matched Wake Analysis* (MWA) for finding causal relationships in event data that combines the best of the two most promising techniques reviewed above: sliding spatio-temporal windows to overcome the MAUP and statistical matching to allow for clean causal inference.

Matched wake analysis

Any attempt to overcome the discussed methodological shortcomings in the analysis of causal relationships in conflict event data must start with a theoretical understanding of the data generating process. A first crucial insight is that events come into existence through a variety of different mechanisms. In conflict research, there is the widely described effect of exogenous geographic conditions that drive overall levels of violence (Hegre et al. 2009; McColl, 1969; O'Loughlin & Witmer, 2011; Raleigh & Hegre, 2009). For example, strategic locations might see higher levels of violence. Ethnic settlement patterns have been linked to conflict events in Iraq (Weidmann & Salehyan, 2013) and in Israel (Bhavnani, Donnay, Miodownik, Mor, & Helbing, 2014). For conceptual clarity one can refer to these factors as the a priori *exposure* of any location to violence. Furthermore, levels of violence generally vary over time. A negotiated ceasefire and seasonal cycles may drive the intensity of conflict across a war zone. These aspects can be referred to as the *momentum* of a conflict at any given time. Isolating the effects of exposure and momentum is a crucial prerequisite for cleanly analyzing the third mechanism driving levels of violence: *reaction* to specific events, i.e. the causal effect of specific interventions. Fig. 1 illustrates the logic of this empirical strategy.

In this conceptual sketch, three types of conflict events are depicted. The rectangular symbol in the center of the left cylinder

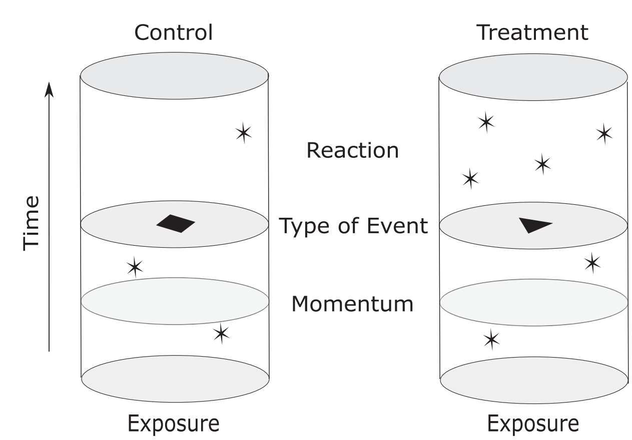

Fig. 1. Illustration of the empirical strategy. Conflict events are divided into two classes of "treatment" and "control" events. For each event, previous levels of "dependent" events and their temporal trends and subsequent levels are established in an automated GIS analysis.