IMPERIAL

Evaluation of the causal effect of disruption on delay as a benchmarking tool for public transport network performance

Guillaume Amann,
Supervisor: Prof. Daniel Graham,
Collaborator: Dr Nan Zhang,

Introduction

The growth of significant transport infrastructure projects has highlighted the increasing complexity and size of urban rapid transit networks. These expansions make it more challenging to operate and benchmark public transport network performances. It is crucial to comprehend both the networks' structures and the causal effect of disruptions on delay to develop relevant strategies to improve their resilience and maintain reliable and efficient service.

Aim and objectives

This paper proposes to apply Synthetic Control (SC) to generate a synthetic benchmark to which can be compared the outcome of a specific treated unit. This can be used to facilitate and scale up the quantification of system resilience from the operator point of view, by addressing the causal effects of service disruptions on delay in arrival time. Indeed, disruptions are non-randomly events occurring on the transport network. This confounding effect (see Figure 1) can be addressed to obtain clear and accurate measurements of the true causal inference of disruptions.

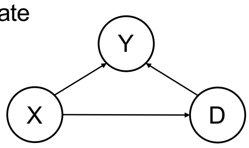

Figure 1: Graph of non-randomised disruption occurrence

Synthetic Control

This study applies Synthetic Control to generate a synthetic control day of operation on the Orange line of the Washington Metro network. For the SC method to provide valid estimates of causal effects, it relies on two simple assumptions:

- synthetic control units serve as a reasonable approximation of what the disrupted unit's outcomes would have been.
- the parallel trends assumption, the disrupted and control units would have followed the same pattern in the absence of the disruption

This study applies Synthetic Control to generate a synthetic control day of operation that mirrors the pre-treatment trends. The leveraged control units are the other non-disrupted days observed over a 2-month period. The weights are then estimated to determine how much each control unit should contribute to the synthetic control unit.

Finally, the disrupted day is plotted against the synthetic control unit. Both units follow the same upside-down U shape. This similarity lasts up until the point of disruption occurrence when the recorded hourly aggregated delay of the treated unit starts following a different path, thus highlighting the relevancy of SC method.

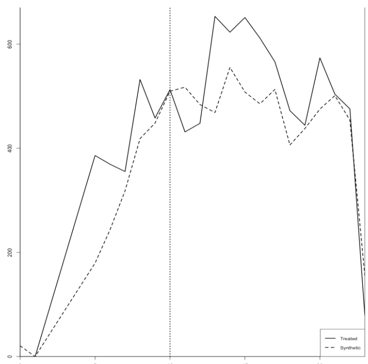

Figure 2: Hourly aggregated delay along the Orange line of the Washington metro network on a disrupted day (20th March 2024, 10am EST), against a Synthetic Control day of operation based on other non-disrupted days

Robust Key Performance Indicator

Once the SC group has been derived, it is possible to compare its evolution with respect to the treated group during the whole period. To proceed further analysis and slightly drift away from causal inference, this gap between the disrupted unit and its synthetic twin can be used to establish a time-to-recovery metric, denoted as τ, to reach a certain arbitrary threshold ξ, as shown by the inequality *infra*.

$$Y_{1,x}(\tau) - Y_{0,x} \cdot W^{*}(V) < \xi_t$$

This method can be either timely or spatially adapted. A conceptual visualisation of the time propagation of delay is shown in Figure 3.

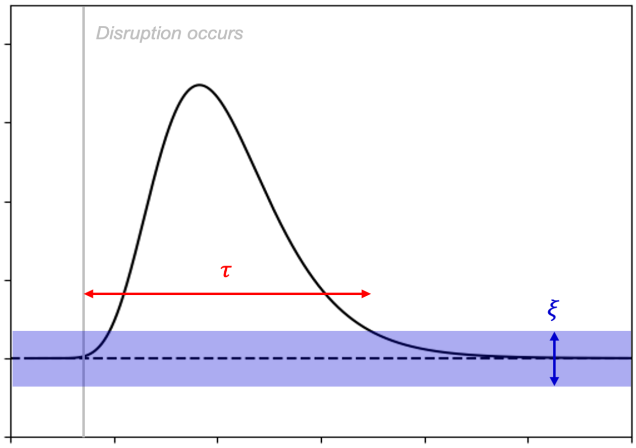

Figure 3: Schematic view of the time evolution a disrupted delay (plain black) against a benchmark non-disrupted delay generated via Synthetic Control (dashed line) (not to scale)

Conclusion

The research employs the Synthetic Control method, which is a robust approach to estimate the causal impacts of interventions. This is particularly useful in scenarios where randomised trials are not feasible, like in transport. In adapting Synthetic Control to use time as the treatment unit and control group, this research proposes a novel approach where the traditional distinction between pre-treatment and post-treatment periods is redefined. Here, the pre-treatment period encompasses all time points leading up to the disruption. This approach creates a synthetic benchmark day of operation, allowing for a more accurate measurement of the system's resilience to delay.