

# Counterfactual and Synthetic Control Method: Causal Inference with Instrumented Principal Component Analysis\*

Cong Wang<sup>†</sup>

September 12, 2024

[Job Market Paper, latest version available here.](#)

## Abstract

In this paper, we propose a novel method for causal inference within the framework of counterfactual and synthetic control. Matching forward the generalized synthetic control method developed by Xu (2017), our instrumented principal component analysis method instruments factor loadings with predictive covariates rather than including them as regressors. These instrumented factor loadings exhibit time-varying dynamics, offering a better economic interpretation. Covariates are instrumented through a transformation matrix,  $\Gamma$ , when we have a large number of covariates it can be easily reduced in accordance with a small number of latent factors helping us to effectively handle high-dimensional datasets and making the model parsimonious. Finally, the novel way of handling covariates is less exposed to model misspecification and achieved better prediction accuracy. Our simulations show that this method is less biased in the presence of unobserved covariates compared to other mainstream approaches. In the empirical application, we use the proposed method to evaluate the effect of Brexit on foreign direct investment to the UK.

**JEL Codes:** G11, G12, G30

---

\*We thank Matteo Lacopini, Emanuele Bacchiocchi for helpful discussion on this paper. There is a Python Package and a Github repository for this paper, available at <https://github.com/CongWang141/JMP.git>, which contains the latest version of the paper, the code, and the data.

<sup>†</sup>Department of Economics and Law, Sapienza University of Rome.

# 1 Introduction

In this paper, we introduce a novel counterfactual imputation method for causal inference, called the Counterfactual and Synthetic Control method with Instrumented Principal Component Analysis (CSC-IPCA). This method combines the dimension reduction capabilities of Principal Component Analysis (PCA) described by [Jolliffe and Cadima \(2016\)](#) to handle high-dimensional datasets with the versatility of the factor models studied by [Bai and Perron \(2003\)](#), [Bai \(2009\)](#), among others, which accommodate a wide range of data-generating processes (DGPs). The CSC-IPCA method represents a significant advancement over the Generalized Synthetic Control (GSC) method proposed by [Xu \(2017\)](#), which utilizes the Interactive Fixed Effects (IFE) approach to model DGPs and impute missing counterfactuals for causal inference.

## References

- Jushan Bai. Panel data models with interactive fixed effects. Econometrica, 77(4):1229–1279, 2009.
- Jushan Bai and Pierre Perron. Computation and analysis of multiple structural change models. Journal of applied econometrics, 18(1):1–22, 2003.
- Ian T Jolliffe and Jorge Cadima. Principal component analysis: A review and recent developments. Philos. Trans. R. Soc. A Math. Phys. Eng. Sci, 374(2065):20150202, 2016.
- Yiqing Xu. Generalized synthetic control method: Causal inference with interactive fixed effects models. Political Analysis, 25(1):57–76, 2017.