Pre-event Trends in the Panel Event-study Design

Simon Freyaldenhoven, *Brown University*

Christian Hansen, *University of Chicago*

Jesse M. Shapiro, *Brown University and NBER*[*]

This version: April 24, 2018

Online Appendix

List of Tables

List of Figures

---

[*]Emails: simon_freyaldenhoven@brown.edu, chansen1@chicagobooth.edu, jesse_shapiro_1@brown.edu.

Estimator	Effect of newspaper entry	First-stage F-test of excluded instruments
Proposed 2SLS estimator (one lead)	0.0034 (0.0013)	54.62
Proposed 2SLS estimator (five leads; BIC)	0.0044 (0.0011)	12.01

Online Appendix Table 1: Estimates $\hat{\beta}$ of the effect of newspapers on voter turnout from $\Delta y_{it} = \beta \Delta z_{it} + \Delta \omega_{st} + \gamma \Delta \eta_{it} + \Delta \varepsilon_{it}$, using our proposed 2SLS estimator. In the first row we use the closest lead of Δz_{it} as the excluded instrument. In the second row we use the BIC to choose the number of leads, allowing up to eight leads. Standard errors in parentheses are clustered at the county level.

| Estimator	Effect of newspaper entry	Coefficient of lead in first-stage
No control	0.0034 (0.0009)	
Controlling for x_{it}	0.0041 (0.0010)	
Proposed 2SLS estimator (one lead)	0.0029 (0.0014)	0.0079 (0.0013)

Online Appendix Table 2: Estimates of the effect of newspapers on voter turnout, including covariates. Table depicts estimates $\hat{\beta}$ of the effect of number of newspapers on voter turnout from $\Delta y_{it} = \beta \Delta z_{it} + \Delta \omega_{st} + \gamma \Delta \eta_{it} + \Delta q'_{it} \theta + \Delta \varepsilon_{it}$. The model differs from the model in the main paper through the inclusion of the vector of covariates q_{it}. In line with the alternative specification in Gentzkow et al. (2011), this vector includes the share of the population that is white, the share of the white population that is foreign-born, the share of the population living in cities with 25,000+ residents, the share of the population living in towns with 2,500+ residents, the population employed in manufacturing as a share of males over 21 years old, and the log of manufacturing output per capita (as proxy for income). See Table 2 for definitions of the estimators. Standard errors in parentheses are clustered at the county level.
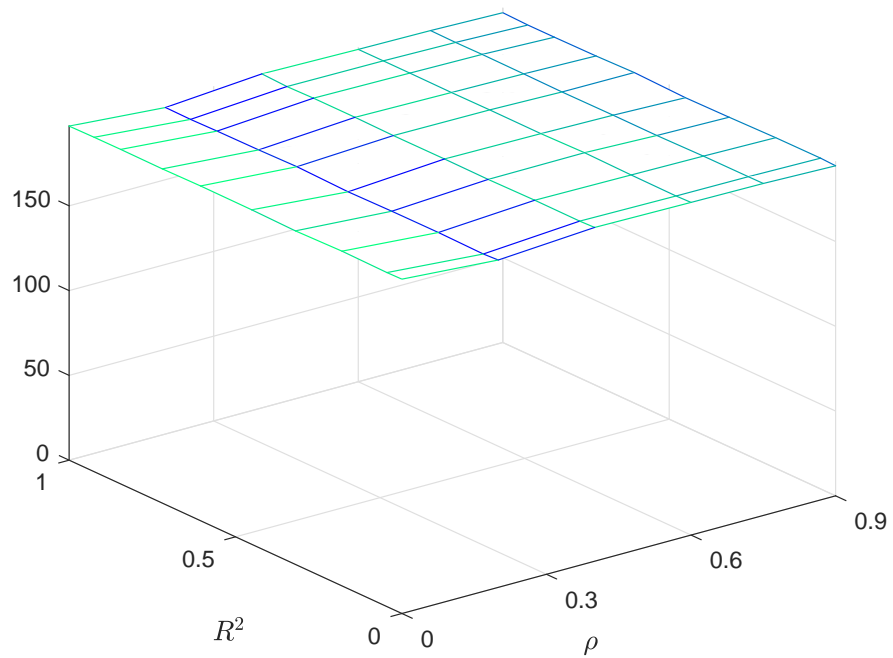
Online Appendix Figure 1: Average number of cross-sectional observations in which an event occurs across the design space considered in the simulations for Figures 4 - 6. Within each set of simulation parameters the number of units with an event generally lies between 160 and 240.
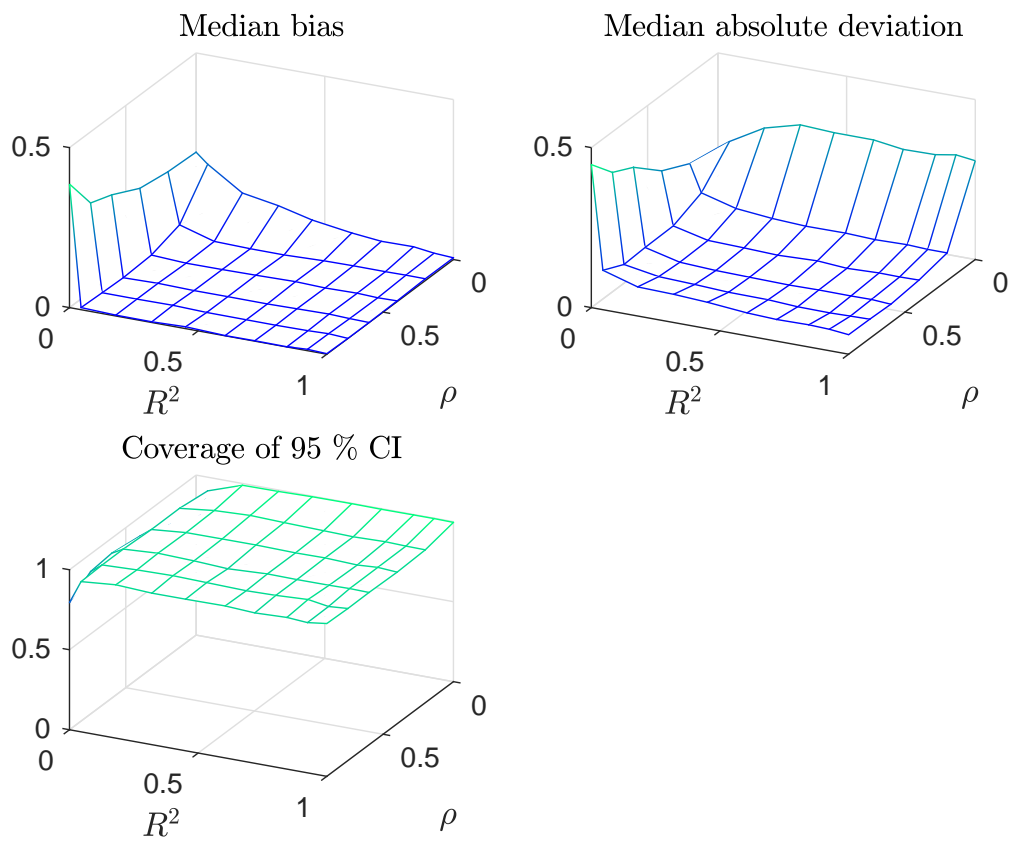
Online Appendix Figure 2: Additional panel for Figures 4 - 6 using the BIC in the first stage to choose the number of leads, between 1 and 5, to be used as excluded instruments.
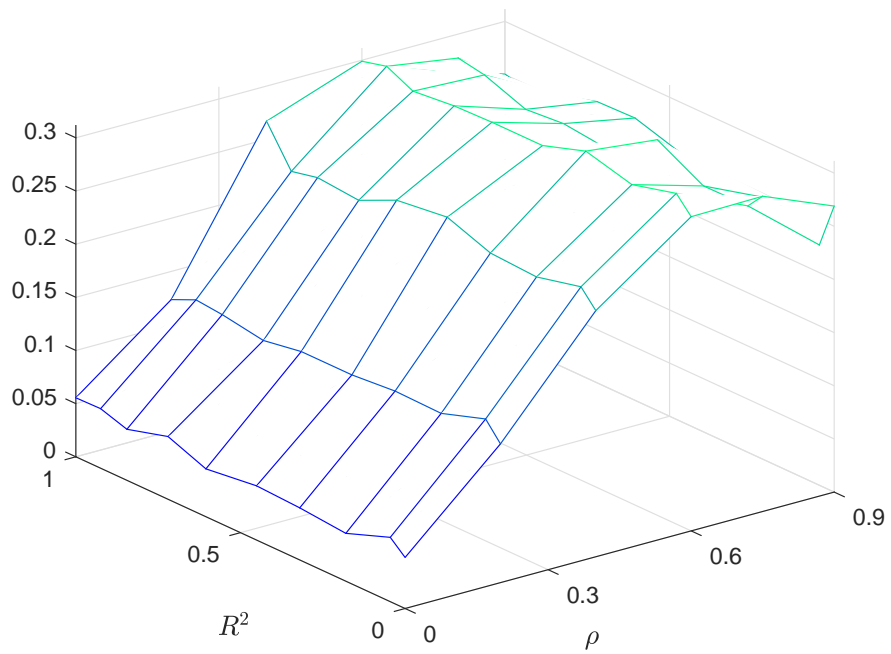
Online Appendix Figure 3: Probability of rejecting the null hypothesis of no pre-trend for the bottom right panel of Figures 4 - 6. With the event plot normalized such that the coefficient on $z_{i,t+1}$ is equal to zero, our pre-test tests that the coefficient on $z_{i,t+2}$ is equal to 0 at the 5 % level.
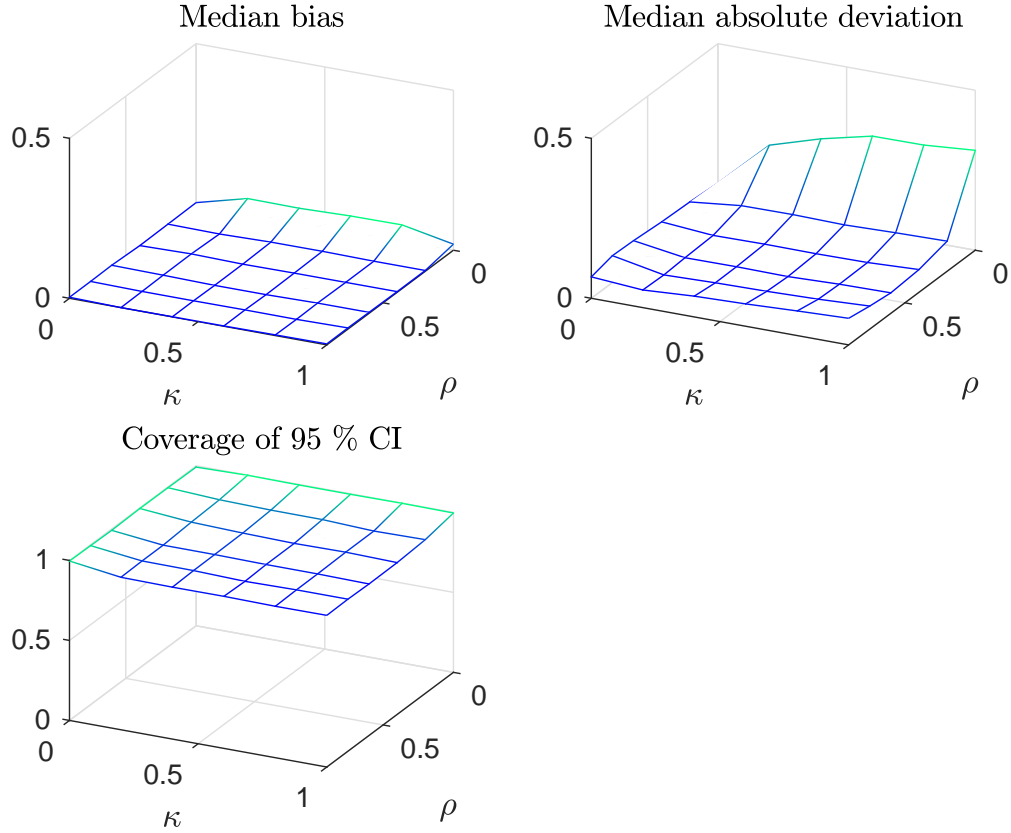
Online Appendix Figure 4: Performance of our proposed estimator ("2SLS - one lead") under thresholding with noise. We deviate from the the DGPs in Definition 1 and Figures 4 - 6 in assuming that $z_{it} = \mathbf{1}(\{\exists t^* \leq t : \eta_{it^*}^{\dagger} > \eta^*\})$, where $\eta_{it}^{\dagger} = \sqrt{\kappa}\eta_{it} + \sqrt{1-\kappa}\tau_{it}$ and $\tau_{it} \sim N(0,1)$ independently of the other variables. To vary the importance of η_{it} in determining z_{it}, we vary κ from zero to one. We fix the population R^2 from the infeasible regression of x_{it} on η_{it} in (11) to 0.45. Each figure is based on 2000 simulation replications. The horizontal axes in each panel correspond to the different values of ρ and κ.
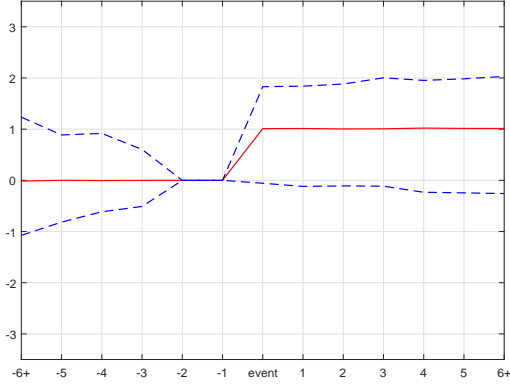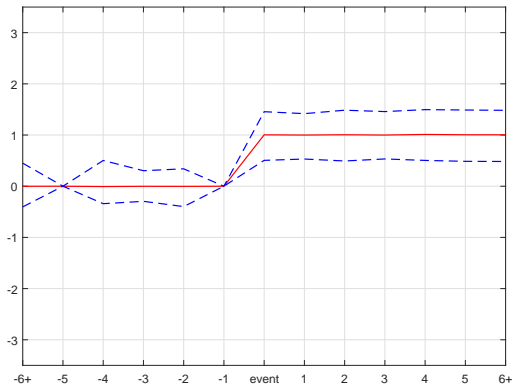
(a) Proposed 2SLS estimator, with closest lead of z_{it} as excluded instrument. Normalized such that $\delta_{-1} = \delta_{-2} = 0$.
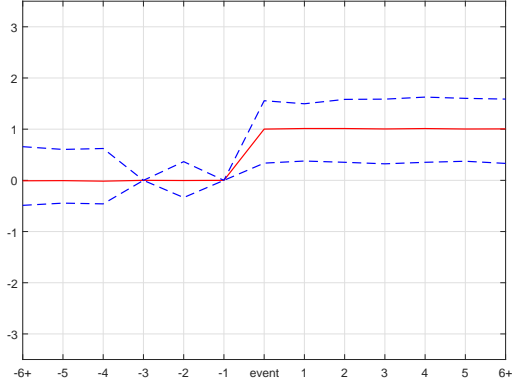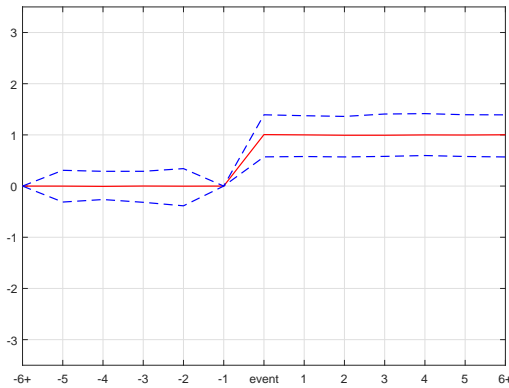

(b) Proposed 2SLS estimator, with closest lead of z_{it} as excluded instrument. Normalized such that $\delta_{-1} = \delta_{-3} = 0$.
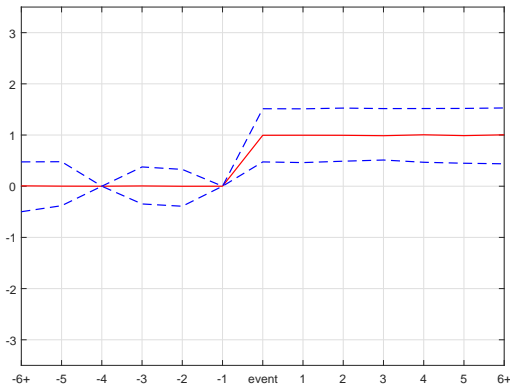
(c) Proposed 2SLS estimator, with closest lead of z_{it} as excluded instrument. Normalized such that $\delta_{-1} = \delta_{-4} = 0$.


(d) Proposed 2SLS estimator, with closest lead of z_{it} as excluded instrument. Normalized such that $\delta_{-1} = \delta_{-5} = 0$.
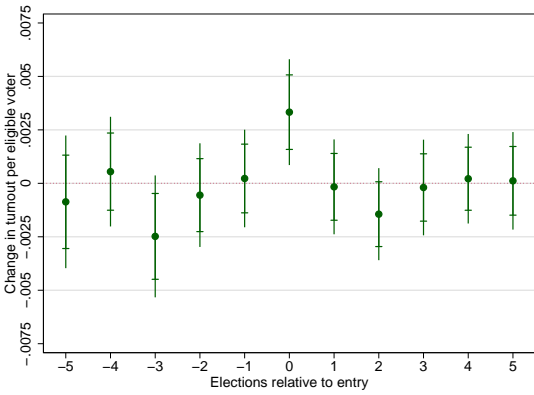

(e) Proposed 2SLS estimator, with closest lead of z_{it} as excluded instrument. Normalized such that $\delta_{-1} = \delta_{-6+} = 0$.

Online Appendix Figure 5: Distribution of event plots (in levels) under the presence of a confounding factor using different normalizations. Each plot shows estimates of the coefficients δ_k from (13) under simulated data from benchmark DGP. The red line in the center represents the average estimate across 10,000 realizations, while the blue dotted lines depict the uniform 95% confidence band: 95% of the estimated sets of coefficients lie within this band. Panel 5a is identical to Figure 3d in the paper.
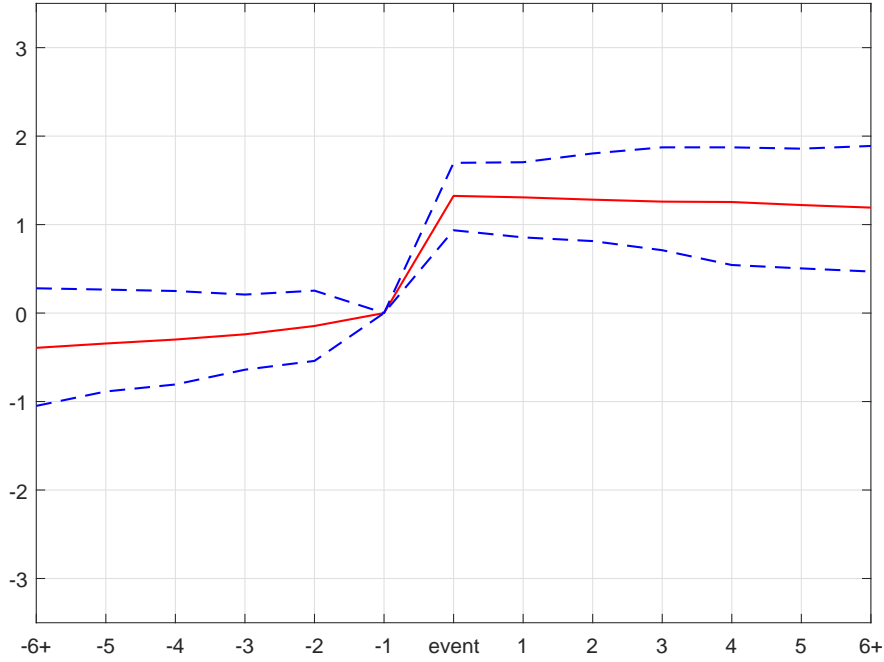
Online Appendix Figure 6: Distribution of event plots (in levels) under the presence of a confounding factor using simulated data from the benchmark DGP defined in Section 3.1 with true causal effect $\beta = 1$, adding unit specific linear time trends to the estimating equation. Specifically, the plot depicts estimates $\hat{\delta}_k$ of coefficients δ_k from $y_{it} = \delta_{-6+}(1 - z_{i,t+5}) + \delta_{6+}z_{i,t-6} + \sum_{k=-5}^{5} \delta_{-k}\Delta z_{i,t+k} + \omega_t + \alpha_i + \xi_i t + \eta_{it} + \varepsilon_{it}$. The red solid line in the center represents the median estimate across 10,000 realizations, while the blue dotted lines depict the uniform 95% confidence band: 95% of the estimated sets of coefficients lie within this band.
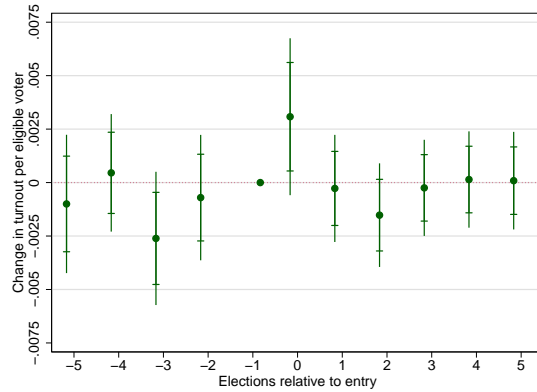
(a) Not controlling for market profitability
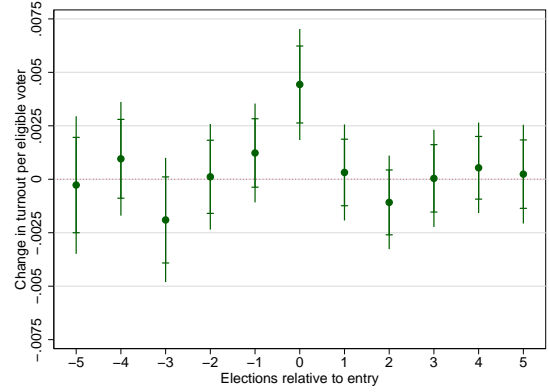
(b) Using the log of voting-eligible population as a proxy for market profitability

(c) Proposed 2SLS estimator, with closest lead of z_{it} as excluded instrument

Online Appendix Figure 7: Estimated effects on presidential turnout at election years around newspaper entries/exits including demographic controls. Figure depicts estimates of the coefficients δ_k from three specifications of the equation $\Delta y_{it} = \sum_{k=-5}^{5} \delta_k \Delta z_{i,t-k} + \Delta \omega_{st} + \gamma \Delta \eta_{it} + \Delta q'_{it} \theta + \Delta \varepsilon_{it}$. The model differs from (18) through the inclusion of the vector of covariates q_{it}. In line with the alternative specification in Gentzkow et al. (2011), this vector includes the share of the population that is white, the share of the white population that is foreign-born, the share of the population living in cities with 25,000+ residents, the share of the population living in towns with 2,500+ residents, the population employed in manufacturing as a share of males over 21 years old, and the log of manufacturing output per capita (as proxy for income). Inner confidence sets as indicated by the dashes correspond to 95% pointwise confidence intervals, while outer confidence sets are the uniform 95% sup-t bands (with critical values obtained via simulation). Confidence sets are clustered at the county level.