(Visualizing) Plausible Treatment Effect Paths

Simon Freyaldenhoven
Federal Reserve Bank of Philadelphia

Christian Hansen
University of Chicago[*]

November 12, 2024

Online Appendix

List of Tables

List of Figures

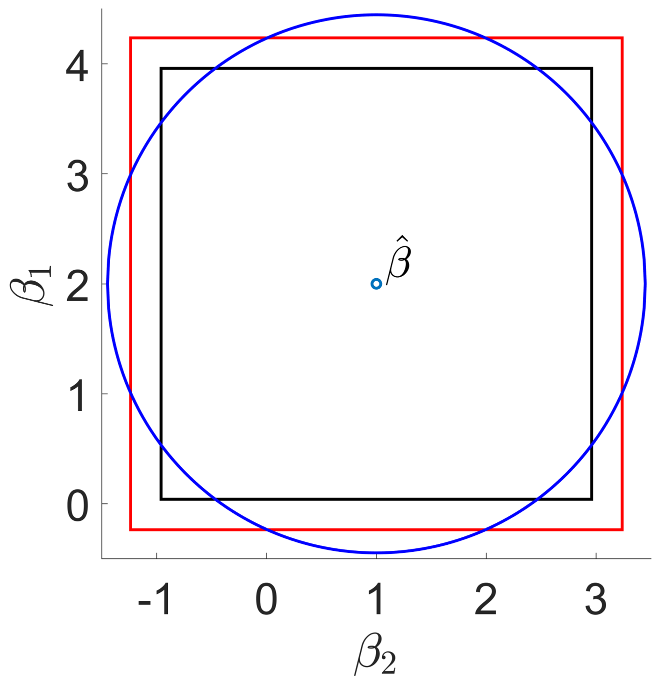
[*]Emails: simon.freyaldenhoven@phil.frb.org, chansen1@chicagobooth.edu

A A primer on confidence regions

If β is a scalar, the standard approach in economics to quantify and visualize the uncertainty associated with an estimate for β is to construct a confidence interval. For a chosen size α, such a confidence interval covers the true value β in $100 * (1 - \alpha)\%$ of all realizations of the data: $\mathbb{P}(\ell(X) < \beta < u(X)) = 1 - \alpha$ where $\ell(X)$ and $u(X)$ denote the lower and upper bounds of the confidence interval and observed data are a realization from random variable X. Intuitively, these intervals visualize to the reader what values of β are "plausible" based on the observed data. The idea being that values inside this confidence interval appear "plausible," while values outside of the interval do not. More formally, values outside this interval are rejected by a standard t-test at level α, while values inside the interval are not rejected.

Since in this paper a dynamic treatment path is the object of interest, $\beta = \{\beta_h\}_{h=1}^{H}$ is an ordered vector instead. We start with a diagram in Online Appendix Figure 1 that illustrates standard methods in the case of a two dimensional parameter $\beta = (\beta_1, \beta_2)$ where estimates are $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2) = (2, 1)$ and $V_\beta = I_2$ is the 2×2 identity matrix.

The predominant practice today is to include pointwise confidence intervals in depictions of estimated treatment effect paths. $100 * (1 - \alpha)\%$ pointwise intervals for a specific β_h simply correspond to choosing $(\ell_h(X), u_h(X)) = (\hat{\beta}_h - z_{1-\alpha/2}\sqrt{V_\beta[h,h]}, \hat{\beta}_h + z_{1-\alpha/2}\sqrt{V_\beta[h,h]})$ where $V_\beta[h,h]$ is the variance of $\hat{\beta}_h$ and $z_{1-\alpha/2}$ is the $1 - \alpha/2$ quantile of a standard normal distribution. For example, the pointwise 95% confidence intervals in the case of the example in Online Appendix Figure 1 for β_1 and β_2 are respectively 2 ± 1.96 and 1 ± 1.96. Correspondingly, the black square depicts the Cartesian product of these pointwise confidence intervals for β_1 and β_2. Denote the region provided by the black square as CR^{pw}. Treated as a confidence region for (β_1, β_2), CR^{pw} *a*) ignores any information in the off-diagonal entries in the covariance matrix of $\hat{\beta}$ and *b*) is only valid for testing pre-specified hypotheses involving single coefficients. Thus, it does not achieve correct coverage for the vector $\beta = (\beta_1, \beta_2)$: For a chosen significance level α, it will generally be true that $\mathbb{P}(\beta \in CR^{pw}) = \mathbb{P}(\ell_h(X) < \beta_h < u_h(X) \ \forall \ h) < (1 - \alpha)$, such that the black square will generally cover the true parameter β in less than $100 * (1 - \alpha)\%$ of realizations of the data. For example, if $Cov(\hat{\beta}_1, \hat{\beta}_2) = 0$, the probability that the pointwise confidence region covers



Online Appendix Figure 1: Illustration of different confidence regions. Pointwise (black), Sup-t (red), and Wald (blue) 95% confidence region in two dimensions. $\hat{\beta} = (2, 1)$ with covariance matrix $V_\beta = I_2$.

the vector β will be $P(\beta \in CR^{pw}) = (1-\alpha)^2$.

One way to construct a uniformly valid confidence region is to take the Cartesian product of *sup-t* confidence intervals (depicted in red) instead. Denote this region CR^{sup-t}. Sup-t intervals are easy to construct, and simply use a slightly large critical value compared to pointwise CIs. Specifically, $100 * (1-\alpha)\%$ sup-t intervals are constructed by choosing $\big(\ell_h(X), u_h(X)\big) = \big(\hat{\beta}_h - c_\alpha\sqrt{V_\beta[h,h]}, \hat{\beta}_h + c_\alpha\sqrt{V_\beta[h,h]}\big)$ where c_α is set such that $\mathbb{P}(\ell_h(X) < \beta_h < u_h(X)\ \forall\ h) \geq (1-\alpha)$. For a chosen significance level α, CR^{sup-t} thus achieves valid coverage, since $\mathbb{P}(\beta \in CR^{sup-t}) \geq (1-\alpha)$ by construction. For example, the sup-t 95% confidence intervals in the case of the example in Online Appendix Figure 1 for β_1 and β_2 are respectively 2 ± 2.24 and 1 ± 2.24. See, e.g., Freyberger and Rai [2018] and Olea and Plagborg-Møller [2019] for details about sup-t interval construction as well as further discussion of different (rectangular) confidence regions. We focus on pointwise and sup-t confidence intervals, since these two are the predominant intervals used in practice (e.g. Callaway and Sant'Anna [2021]; Jordà [2023]; Boxell et al. [2024]).

Finally, the blue circle in Online Appendix Figure 1 corresponds to an alternative confidence region for β, namely the Wald confidence region. Denote this region CR^{Wald}. This region simply collects all parameter values (b_1, b_2) that are not rejected by a standard Wald test of the null hypothesis that $(\beta_1, \beta_2) = (b_1, b_2)$ at level α. For a chosen significance level α, this region also achieves valid coverage: $\mathbb{P}(\beta \in CR^{Wald}) = (1 - \alpha)$ by construction. We note that, in contrast to the pointwise confidence region, both sup-t and Wald confidence regions depend on the off-diagonal entries in $Var(\hat{\beta})$.

B Algorithmic Implementation

Recall that the algorithm takes as input jointly normal estimates of the treatment path $\hat{\beta} \sim N(\beta, V_\beta)$, where $V_\beta = \sigma^2 V$, $\sigma^2 = \frac{1}{H}\sum_{h=1}^{H} V_\beta(h,h)$, and V is positive-definite. We define the following object:

$$\begin{aligned}
\tilde{\beta}(\lambda_1, \lambda_2, K) &= \arg\min_b \; Q(b, \lambda_1, \lambda_2, K) \\
&= \arg\min_b \; \underbrace{(\hat{\beta} - b)' V^{-1} (\hat{\beta} - b)}_{\text{distance from } \hat{\beta}} + \lambda_1 \underbrace{b' D_1' W_1(K) D_1 b}_{\substack{\text{penalty on first difference} \\ \text{after horizon K}}} + \lambda_2 \underbrace{b' D_3' W_3 D_3 b}_{\substack{\text{penalty on} \\ \text{third difference}}},
\end{aligned}$$

where

- λ_1, λ_2, K are tuning parameters that determine the surrogate M
- D_1 and D_3 are the $H \times (H-1)$ and $H \times (H-3)$ first and third difference operators
- $V_1 = D_1 V D_1'$, $V_3 = D_3 V D_3'$ are (scaled) variance matrices for first and third differences
- $V_1(K)$ is the $(H-K) \times (H-K)$ matrix consisting of the lower right entries of V_1, $V_1(K : H-1, K : H-1)$
- $\bar{V}_3 = \frac{1}{H-3}\sum_{h=1}^{H-3} V_3(h,h)$, $\bar{V}_1(K) = \frac{1}{H-K}\sum_{h=K}^{H-1} V_1(h,h)$
- $W_1(K) = \begin{pmatrix} \mathbf{0}_{(K-1)\times(K-1)} & \mathbf{0}_{(K-1)\times(H-K)} \\ \mathbf{0}_{(H-K)\times(K-1)} & diag(V_1(K))/\bar{V}_1(K) \end{pmatrix}$
- $W_3 = diag(V_3)/\bar{V}_3$

Intuitively, $W_1(K)$ and W_3 are analogs to natural scaling in standard ridge with independent columns but different variances.

To select the surrogate M from the data we choose $M = (\lambda_1, \lambda_2, K)$ that minimizes a BIC analog over $\mathcal{M}$: $\hat{M} = \arg\min_{M \in \mathcal{M}} (\hat{\beta} - \tilde{\beta}(M))' V_\beta^{-1} (\hat{\beta} - \tilde{\beta}(M)) + \log(H)\mathrm{df}(\lambda_1, \lambda_2, K)$. The universe of models considered, $\mathcal{M}$, includes

(a) A constant, linear, quadratic, and cubic treatment effect model (with one, two, three, and four degrees of freedom respectively)

(b) Surrogates of the form $P(M)\beta = \beta(\lambda_1, \lambda_2, K)$ using a grid over $(\lambda_1, \lambda_2, K)$

(c) the unrestricted estimates $\hat{\beta}$ ($df = H$)

We construct the grid for the surrogates under (b) as follows. First, we set a lower and upper bound for λ_1 and λ_2. Independent of K, these bounds are equal to $(\underline{\lambda}_1, \bar{\lambda}_1) = (e^{-10}, e^{10})$, and $(\underline{\lambda}_2, \bar{\lambda}_2) = (e^{-10}, \bar{\lambda}_2)$, where $\bar{\lambda}_2$ is defined as the λ_2 such that $\mathrm{df}(e^{-10}, \lambda_2, K) = 4$.[1] Note that, with $\lambda_1 = 0$, $\bar{\lambda}_2$ also does not depend on K. We then consider the Cartesian product of an equal spaced grid of 20 points between $(\log(\underline{\lambda}_1), \log(\bar{\lambda}_1))$ and equal spaced grid of 20 points between $(\log(\underline{\lambda}_2), \log(\bar{\lambda}_2))$, and retain those grid points with $df \in [4, H-1]$.

[1] Recall that $\mathrm{df}(\lambda_1, \lambda_2, K) = \mathrm{trace}\left(\left(V^{-1} + \lambda_1 D_1' W_1(K) D_1 + \lambda_2 D_3' W_3 D_3\right)^{-1} V^{-1}\right)$.

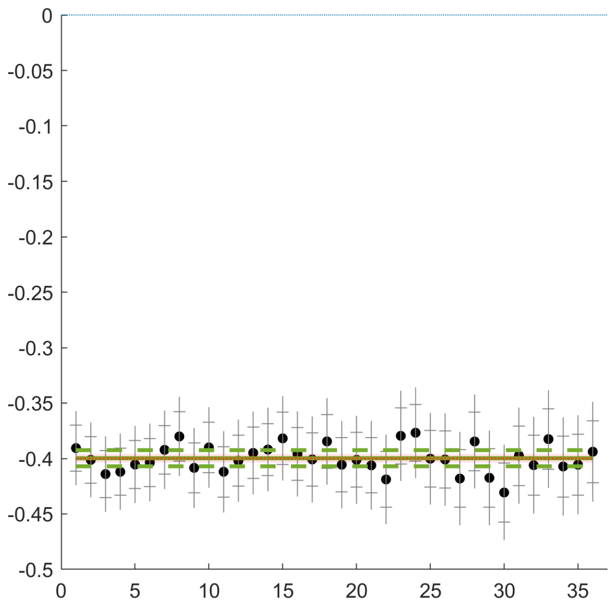
C Simulation Design

Scenario	Treatment Path β
Constant Treatment Effect	$\beta_h = -0.4 \; \forall \; h$
Smooth, eventually flat	$\beta_h = \begin{cases} -0.289 + \frac{(18-h)^2}{1000} & h \leq 17 \\ -0.289 & h \geq 18 \end{cases}$
Hump-shaped	$\beta_h = -0.4 - 0.4 \sin\left(\frac{3}{70}\pi(h-1)\right) \; \forall \; h$
Wiggly	$\beta_h = \breve{\beta}_h + \xi_h$, where $\xi_h \sim N(0, 0.1)$ iid across h and $\breve{\beta}_h = \begin{cases} -0.4 \sin\left(\frac{1}{35}\pi(h-1)\right) & h \leq 19 \\ -0.4 & h \geq 20 \end{cases}$

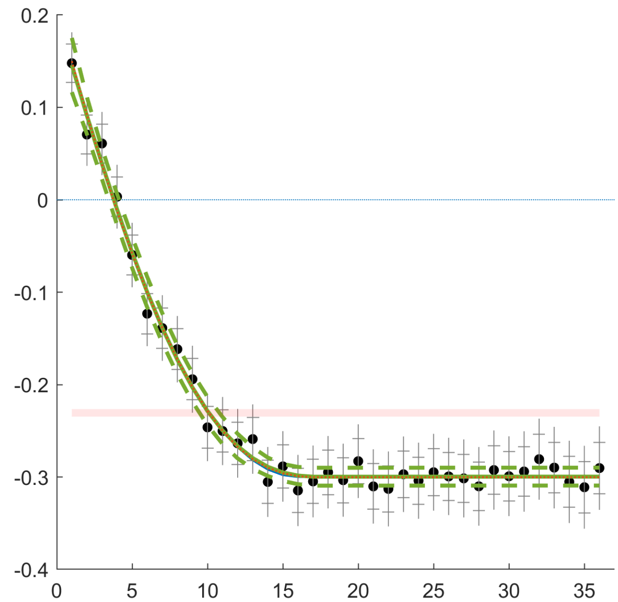
Online Appendix Table 1: Detailed description of the four different treatment paths $\beta = \{\beta\}_{h=1}^{36}$ considered in the simulations. We draw a single realization of the "Wiggly" scenario (which is depicted in Figure 3d) to use throughout our simulations.

We generate the covariance matrix of $\hat{\beta}$ as $V_\beta = \sigma^2 * diag(S)R \; diag(S)$, where $S_h = (100+h)/100$, and R is a $H \times H$ Toeplitz matrix with $R_{ij} = R_{i+1,j+1} = \rho^{i-j}$. For all results in the main text, we set $\rho = 0$ (such that R becomes the identity matrix).

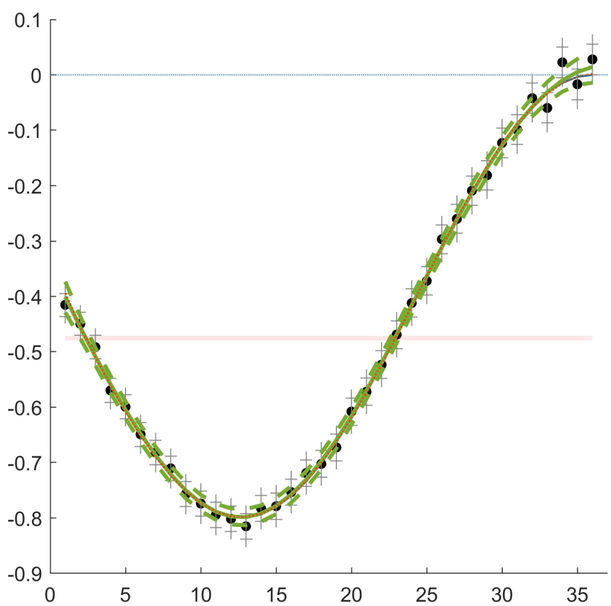
D Additional simulation results



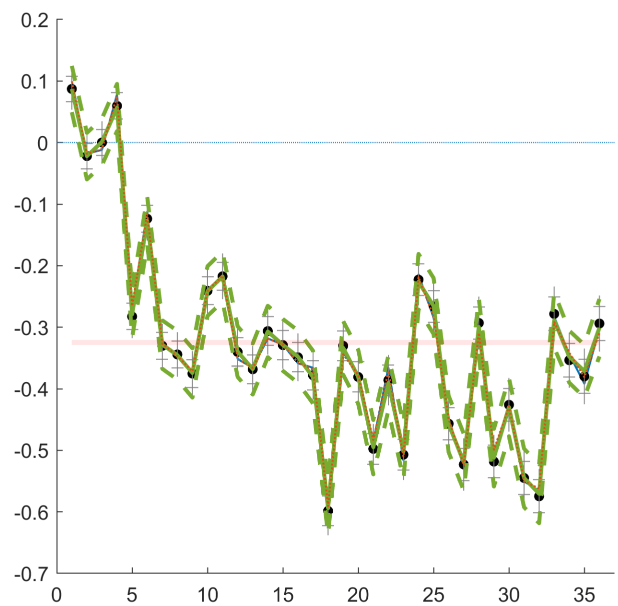
(a) treatment path constant



(b) treatment path smooth, eventually flat

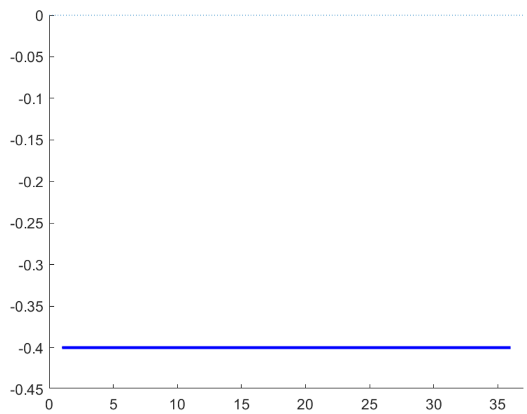


(c) treatment path hump-shaped

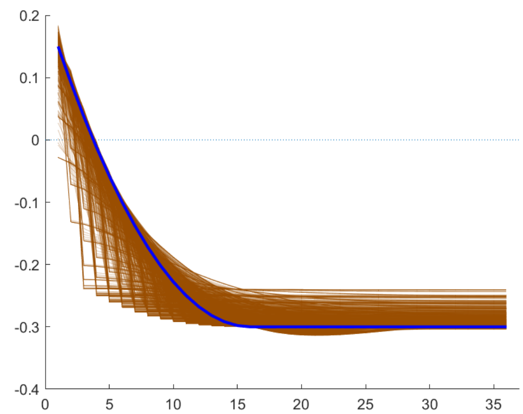


(d) treatment path wiggly

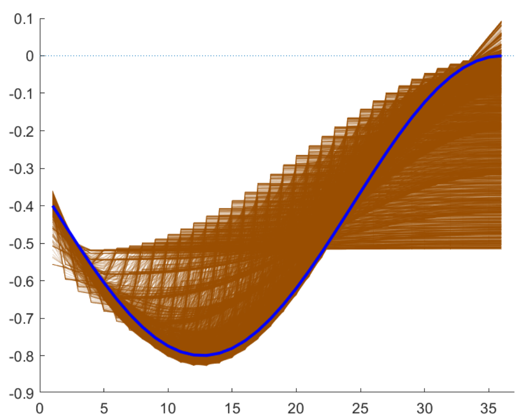
Online Appendix Figure 2: Exemplary event-study plots including our proposals with smaller noise than in Figure 2.



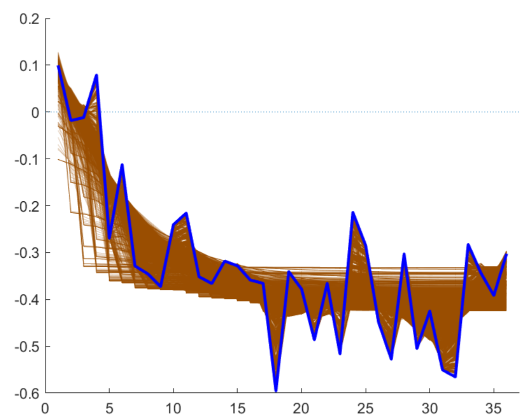
(a) Constant Treatment Effect



(b) Smooth, eventually flat

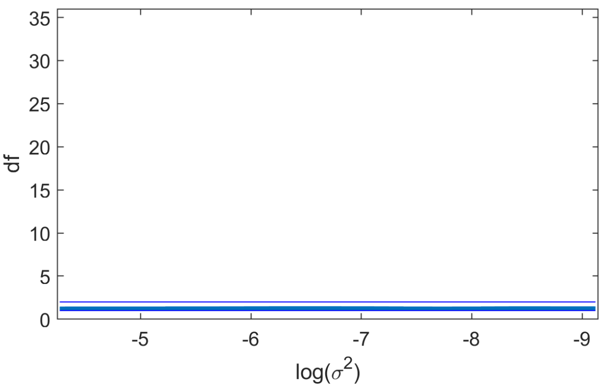


(c) Hump-shaped

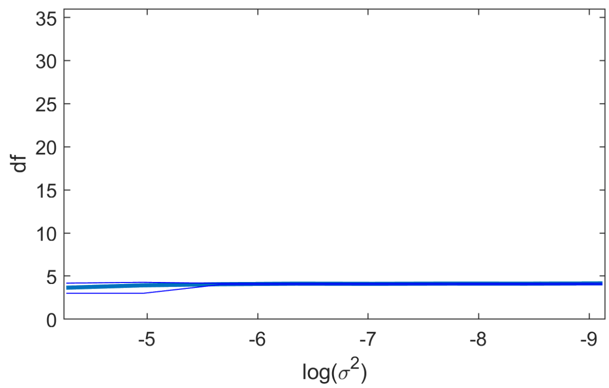


(d) Wiggly

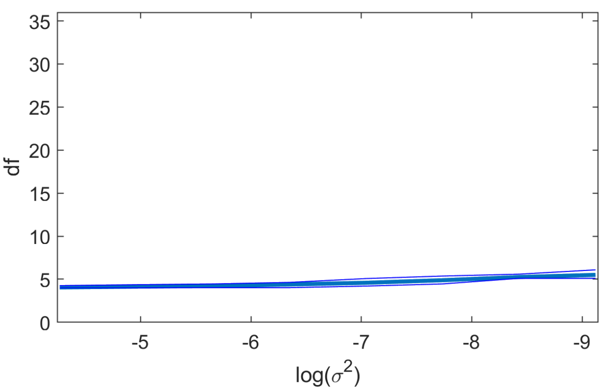
Online Appendix Figure 3: Illustration of Model universe $\mathcal{M}$. Brown lines correspond to all considered models M of the form $P(M)\beta = \beta(\lambda_1, \lambda_2, K)$ with $df \in [4, H-1]$. Blue line corresponds to true treatment effect β.



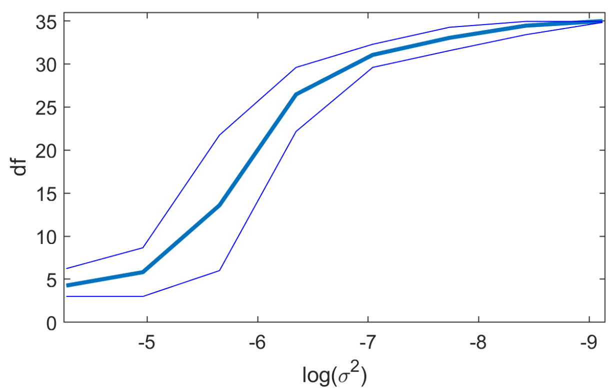
(a) Constant Treatment Effect



(b) Smooth, eventually flat

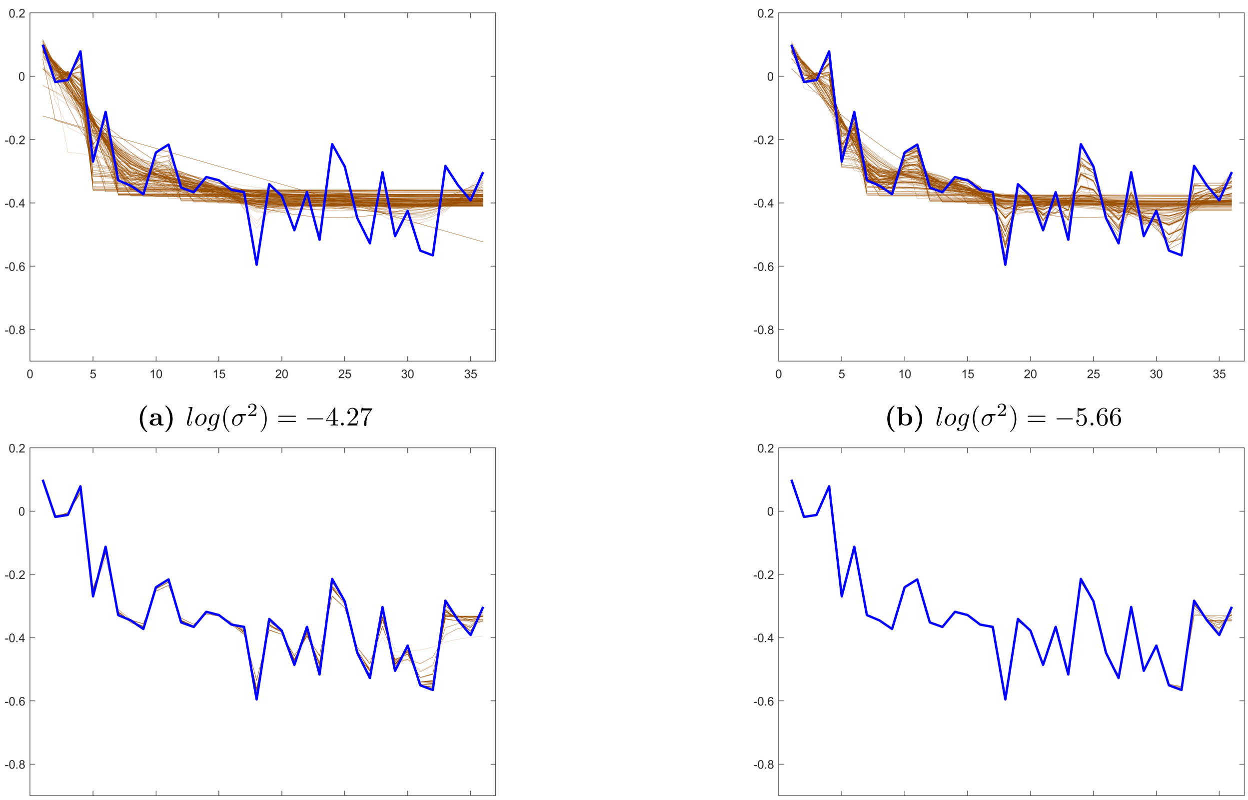


(c) Hump-shaped



(d) Wiggly

Online Appendix Figure 4: Chosen *df* across realizations for restricted estimates as a function of the amount of noise σ^2 in the initial estimates $\hat{\beta}$.



(a) $log(\sigma^2) = -4.27$

(b) $log(\sigma^2) = -5.66$

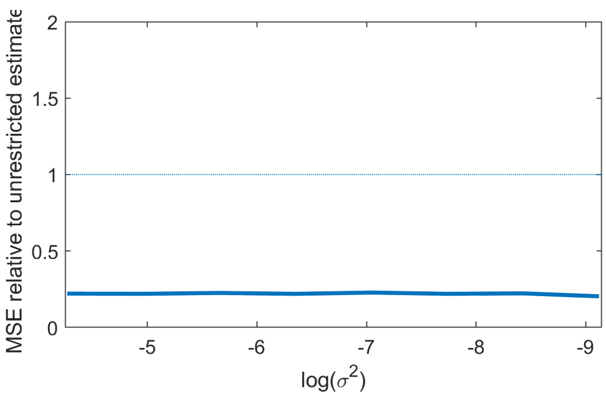
(c) $log(\sigma^2) = -7.04$

(d) $log(\sigma^2) = -8.43$

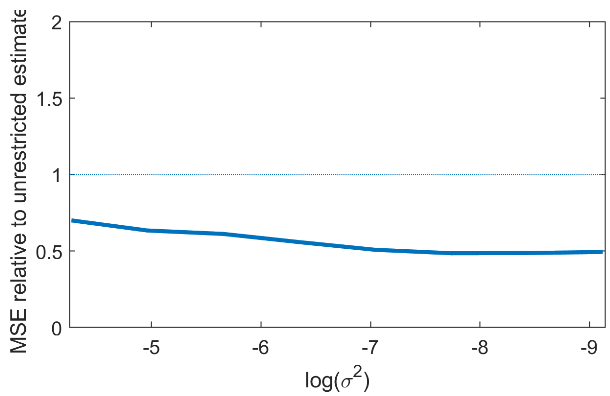
Online Appendix Figure 5: Illustration of 1,000 chosen surrogates for Wiggly DGP for various levels of σ^2, the amount of noise in the initial estimates $\hat{\beta}$.

E Simulation results with positively correlated estimates

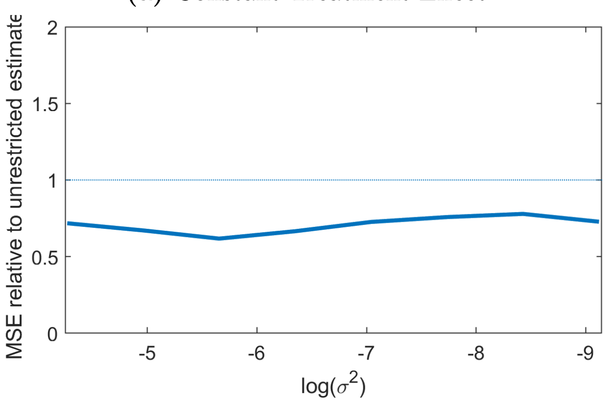
In this appendix, we repeat the simulation experiment reported in the main test using a covariance matrix V_β capturing positively correlated estimates. In particular, recall that $V_\beta = \sigma^2 * diag(S)R\ diag(S)$, where $S_h = (100+h)/100$, and R is a $H \times H$ Toeplitz matrix with $Rij = R_{i+1,j+1} = \rho^{i-j}$. In the results that follow, we set $\rho = 0.8$.



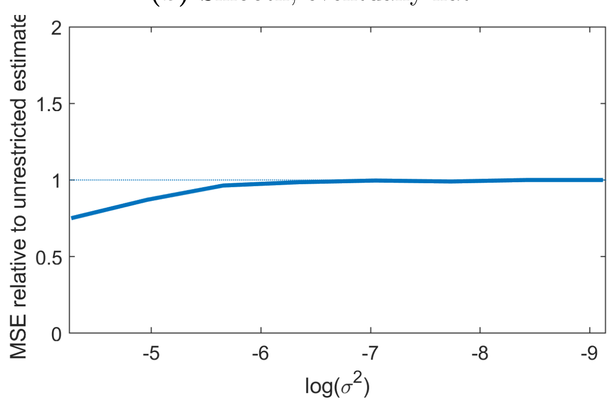
(a) Constant Treatment Effect



(b) Smooth, eventually flat

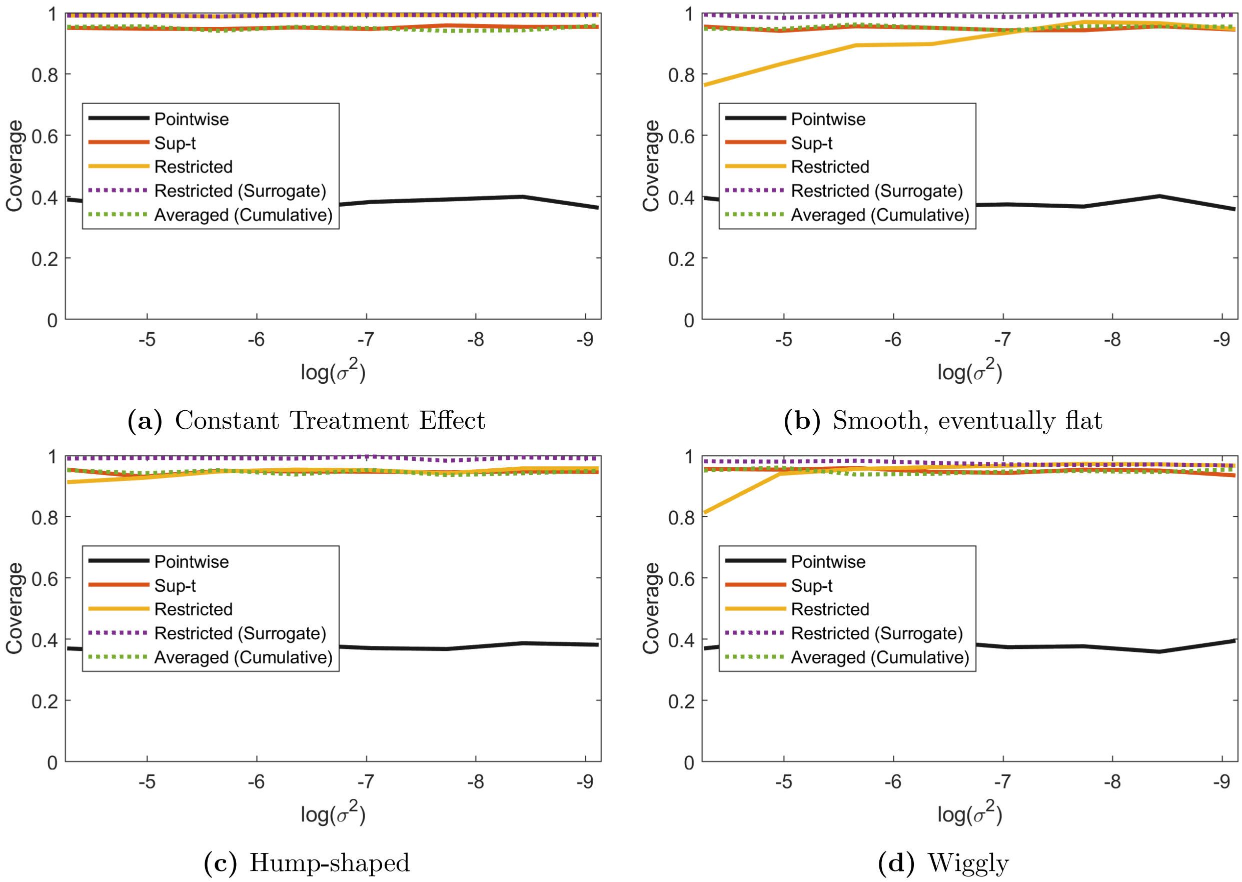


(c) Hump-shaped



(d) Wiggly

Online Appendix Figure 6: Relative performance of restricted and unrestricted estimators. Depicted is the ratio $\frac{MSE(\tilde{\beta}(\hat{M}))}{MSE(\hat{\beta})}$ as a function of the amount of noise in the initial estimates $\hat{\beta}$.



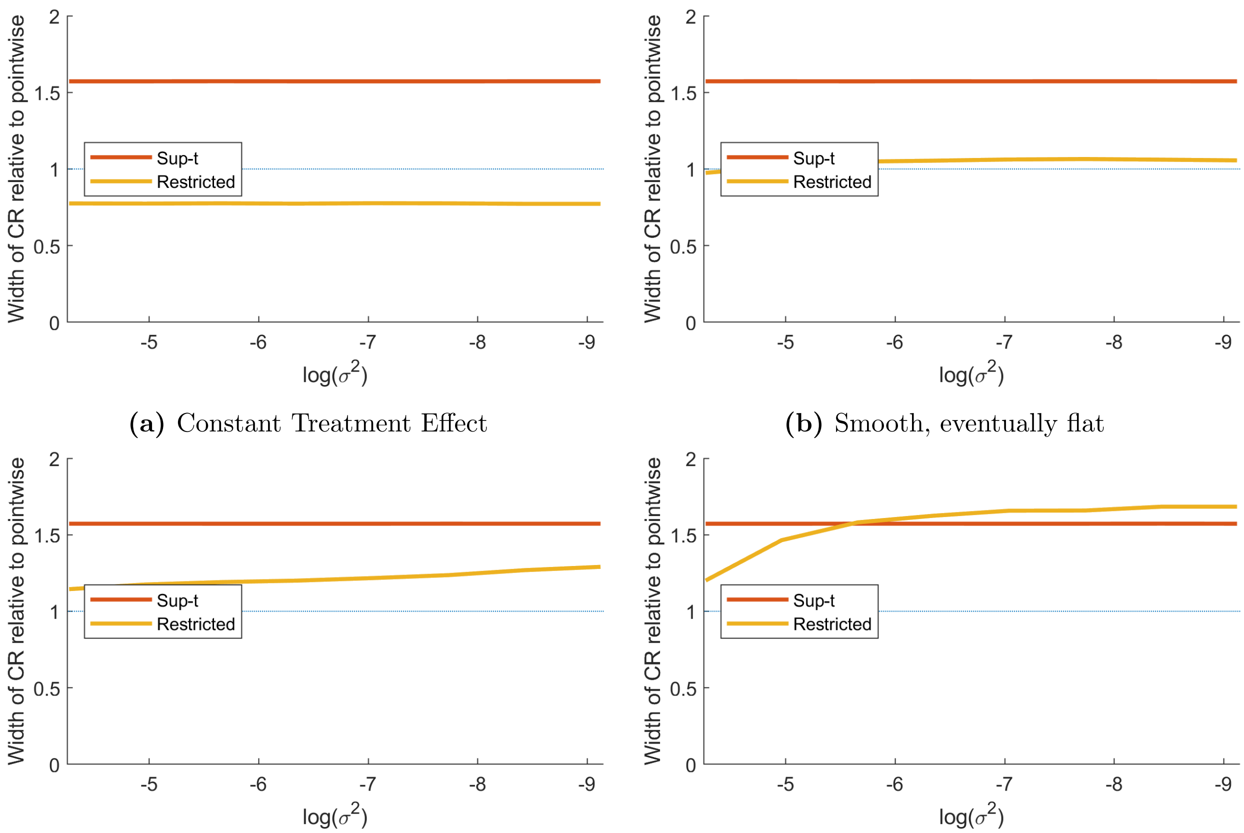
(a) Constant Treatment Effect

(b) Smooth, eventually flat

(c) Hump-shaped

(d) Wiggly

Online Appendix Figure 7: Coverage properties of various confidence regions as a function of the amount of noise in the initial estimates $\hat{\beta}$.



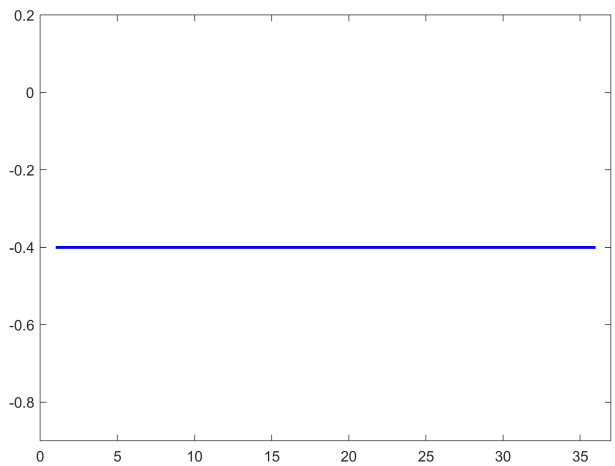
(a) Constant Treatment Effect

(b) Smooth, eventually flat

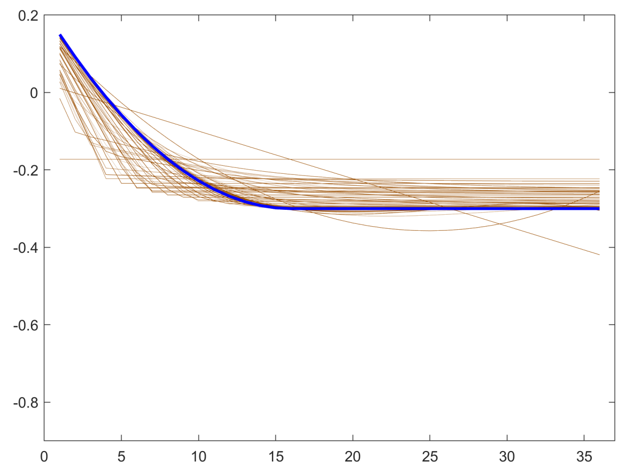
(c) Hump-shaped

(d) Wiggly

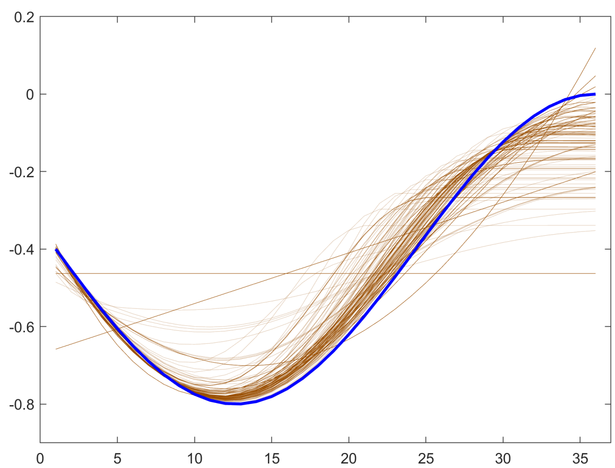
Online Appendix Figure 8: Average width of confidence regions relative to pointwise CIs as a function of the amount of noise in the initial estimates $\hat{\beta}$.



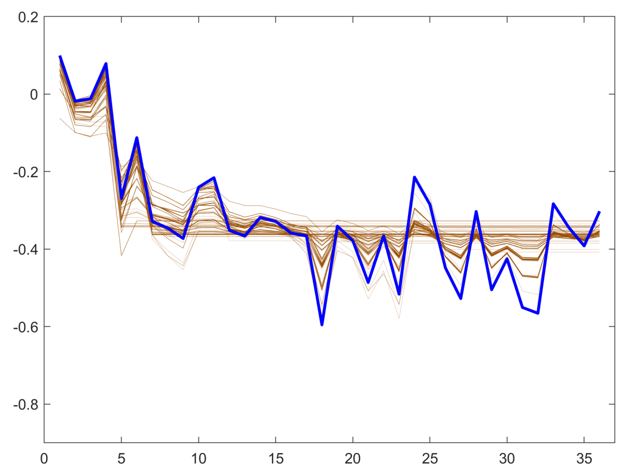
(a) Constant Treatment Effect



(b) Smooth, eventually flat



(c) Hump-shaped



(d) Wiggly

Online Appendix Figure 9: Illustration of the 1,000 chosen surrogates for $\sigma^2 = 0.014$ ($\log(\sigma^2) = -4.27$) under positive correlation in the point estimates.