FEDERAL RESERVE BANK
PHILADELPHIA

Working Papers
RESEARCH DEPARTMENT

(Visualizing) Plausible Treatment Effect Paths

Simon Freyaldenhoven
Federal Reserve Bank of Philadelphia

Christian Hansen
University of Chicago and Visiting Scholar, Federal Reserve Bank of Philadelphia Research Department

WP 25-27

PUBLISHED
September 2025

ISSN: 1962-5361

DOI: https://doi.org/10.21799/frbp.wp.2025.27

(Visualizing) Plausible Treatment Effect Paths

Simon Freyaldenhoven
Federal Reserve Bank of Philadelphia

Christian Hansen
*University of Chicago**

July 31, 2025

Abstract

We consider estimation and inference for treatment effect paths. Examples include dynamic treatment effects, impulse response functions, and event study paths. We present two sets of plausible bounds to help visualize uncertainty associated with these paths. Both plausible bounds are often tighter than traditional confidence intervals, and can provide insights even when traditional (uniform) confidence bands appear uninformative. Our first set of bounds covers the average (or overall) effect rather than the entire path. Our second set of bounds imposes data-driven smoothness restrictions on the treatment path, using post-selection inference (Berk et al. [2013]) to provide formal coverage guarantees.

JEL Codes: C12, C13

KEYWORDS: dynamic treatment effects, post-selection inference, uniform inference

*We thank Thorsten Drautzburg, Paul Goldsmith-Pinkham, Lutz Killian, Christian Leuz, Paul Mohnen, Matthew Notowidigdo, Mikkel Plagborg-Møller, Jon Roth, and Jesse Shapiro, as well as audiences at the Federal Reserve Bank of Philadelphia, the University of Bonn, Georgetown University, Maastricht University, and NY Camp Econometrics for comments. The views expressed herein are those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Philadelphia or the Federal Reserve System.
Emails: simon.freyaldenhoven@phil.frb.org, chansen1@chicagobooth.edu

1 Introduction

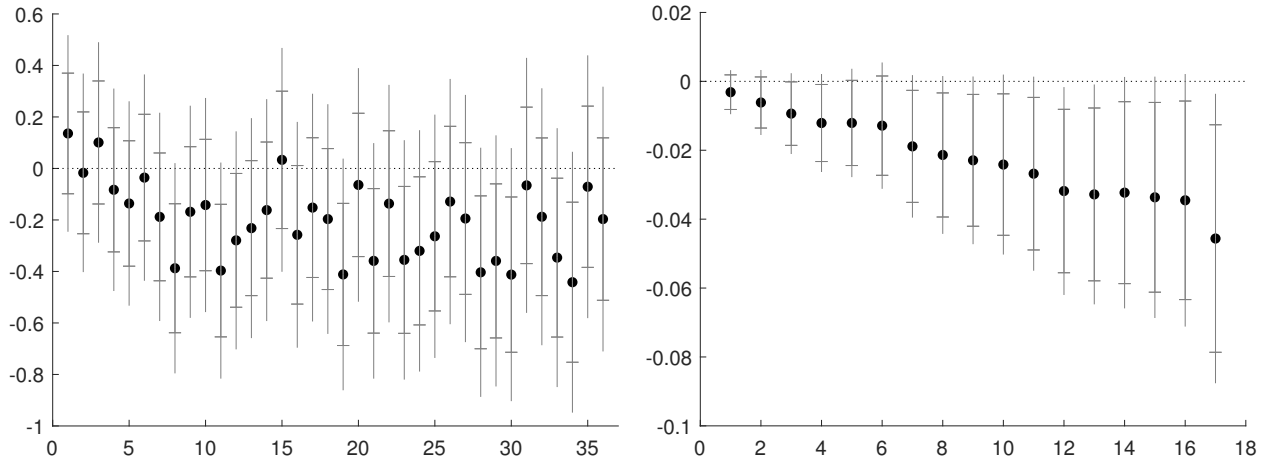
We are interested in the treatment effect path of a policy at discrete horizons $h = 1, ..., H$. Examples include dynamic treatment effects in microeconomics, impulse response functions in macroeconomics, and event study paths in finance. We write $\beta = \{\beta_h\}_{h=1}^H$ for the vector that collects this dynamic treatment effect path up to the fixed maximum horizon of interest H. We assume access to point estimates of the parameters β_h, denoted by $\hat{\beta}_h$, that correspond to the cumulative effect of the policy at horizon $h = 1, \dots H$. Throughout, we assume the vector that collects the estimated dynamic treatment effect path, $\hat{\beta}$, satisfies $\hat{\beta} \sim N(\beta, V_\beta)$ and that we have access to the covariance matrix V_β. Leading examples to obtain such estimates include distributed lag models, local projections, and event studies.[1] We consider both point estimation and uncertainty quantification, though our focus will be on the latter. In particular, we introduce two approaches to visualize the uncertainty about the treatment path, which we call *cumulative* and *restricted* plausible bounds. Both bounds are often substantially tighter than traditional confidence intervals, and can provide useful insights even when traditional (uniform) confidence bands appear uninformative.

The standard approach in economics to quantify and visualize the uncertainty associated with parameter estimates is to construct confidence regions. Intuitively, a confidence region visualizes to the reader what values of the parameter, in this case β, are "plausible" based on the observed data. The idea being that values inside this region appear plausible, while values outside of the region do not. The two predominant confidence regions in practice are pointwise and sup-t confidence regions (e.g. Callaway and Sant'Anna [2021]; Jordà [2023]). A third alternative is the Wald confidence region CR^{Wald}. This region simply collects all parameter values b that are not rejected by a standard Wald test of the null hypothesis that $\beta = b$ at level α. While a confidence region constructed from pointwise confidence intervals does not achieve correct coverage for the vector β, both sup-t and Wald confidence regions achieve valid coverage: $\mathbb{P}(\beta \in CR^{Wald}) = \mathbb{P}(\beta \in CR^{sup-t}) = (1 - \alpha)$.[2]

Sup-t and Wald confidence regions both come with some advantages and disadvantages. Since the Wald region is an ellipsoid, a disadvantage of the Wald confidence region is that

[1] We abstract away from approximation issues, but note that standard asymptotic approximations within these settings along with access to consistent asymptotic variance estimators motivate this setup.

[2] We discuss these regions, and their construction, in more detail in Appendix A. For further discussion of uniform confidence bands, and, in particular, the merits of sup-t confidence bands, also see Freyberger and Rai [2018] and Olea and Plagborg-Møller [2019].



(a) Simulated data with $Cov(\hat{\beta}_i, \hat{\beta}_j) = 0$ for $i \neq j$

(b) Replicated from Bosch and Campos-Vazquez [2014] (Figure 4b)

Figure 1: Two exemplary treatment effect plots, including point estimates, and pointwise and sup-t confidence intervals.

it becomes infeasible to visualize in higher dimensions (i.e. when $H > 3$). The sup-t confidence region has the advantage of being easy to visualize. However, the volume of the sup-t confidence region quickly explodes relative to the volume of the Wald region. For example, when V_β is the identity matrix, the relative volume of the Wald region is less than 10% and around 0.1% of the volume of the sup-t region for $H = 12$ and $H = 24$, respectively. These numbers are generally even smaller if the entries in $\hat{\beta}$ have non-zero correlation.[3] One immediate consequence is that, for even moderate horizons H, the overwhelming majority of paths inside the sup-t bands would be rejected by a simple joint hypothesis test. This property seems unappealing to us and serves as a first indication that sup-t confidence bands may not always be appropriate for visualizing what dynamic treatment effect paths are plausible.

To illustrate this further, Figure 1 depicts two exemplary treatment effect plots. The object of interest is the treatment path of a policy over the depicted horizon. The point estimates $\hat{\beta}$ are given by the black dots. Both panels further include pointwise 95% confidence intervals (inner confidence set as indicated by the dashes) and uniform 95% sup-t confidence bands (outer confidence set). While the pointwise confidence intervals only permit testing of pre-selected hypotheses for individual coefficients β_h, the sup-t bands contain the entire true path β in 95% of realized samples. Figure 1a depicts a hypothetical example with zero

[3]We illustrate this difference in confidence region volume further in Appendix Figure 3.

correlation between the estimated coefficients.[4] Figure 1b is based on the same estimates as Figure 4b in Bosch and Campos-Vazquez [2014]. In this example, all off-diagonal entries in V_β are positive, and the average correlation between adjacent coefficients is 0.95.

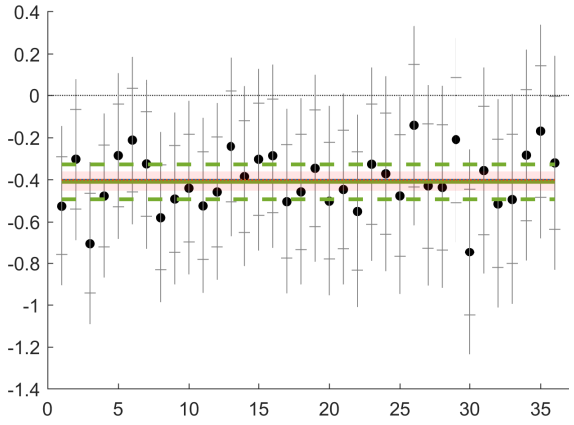
The sup-t region in Figure 1a includes treatment paths that imply an overall positive effect (paths with $\sum_{h=1}^{H} \beta_h > 0$) and treatment paths with very different shapes. In fact, $\beta = 0$ falls inside the sup-t bands, suggesting that the null of "no treatment effect" is plausible. However, a joint test of the null hypothesis that $\beta = 0$ yields a p-value of 1.54×10^{-9}. In contrast, in Figure 1b the sup-t region does not include $\beta = 0$, suggesting that the null of "no treatment effect" is not plausible. However, a joint test of the null hypothesis that $\beta = 0$ yields a p-value of 0.33. These discrepancies between the easy-to-visualize sup-t region and the results of simple joint hypothesis tests again suggest to us that the sup-t confidence region may not always be providing an empirically effective visualization of what treatment effect paths are plausible.

In Figure 2, we therefore introduce two alternative ways to visualize plausible treatment effect paths. Panels 2a-2d are based on simulated data, while panels 2e and 2f are based on published figures in macroeconomics (Nakamura and Steinsson [2018]) and microeconomics (Bosch and Campos-Vazquez [2014]). The inner and outer confidence intervals, respectively, correspond to the usual pointwise and sup-t confidence intervals.[5] In addition, panels 2a-2d include a solid blue line to represent the true treatment effect path and an additional dotted red line, which we explain below. Finally, each plot includes two new features: (i) the shaded red area, and (ii) the dashed and solid green lines. Importantly, these new features shift the goal posts relative to the Wald and sup-t bounds: The inferential target of these bounds is *not* the true treatment path.

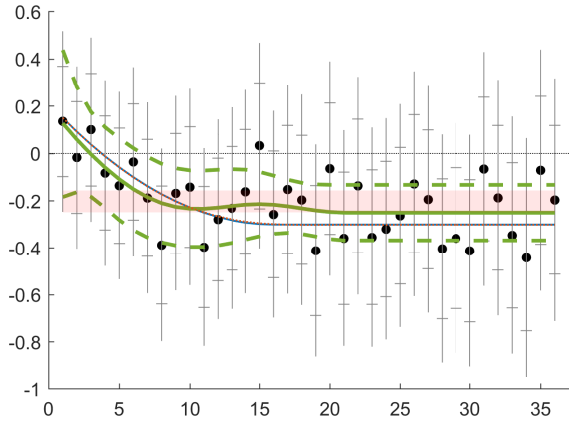
The shaded red area represents our proposed 95% *cumulative plausible bounds*. We construct these bounds so that the average treatment effect across the depicted horizons will be within these bounds for 95% of all realizations of the data. For example, in Figure 2b, these bounds suggest that the average effect of the policy over the 36 periods depicted is between (-0.248, -0.156), and thus that the overall effect of the policy over the 36 periods is strictly negative and inside the window (-8.93, -5.62). In contrast to the standard sup-t region, these bounds suggest that a treatment path with no overall effect of the policy is not

[4] We give more detail on the underlying DGP in Section 4.

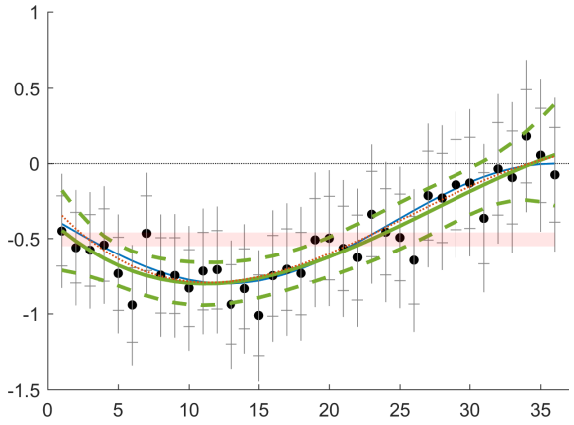
[5] Figure 2b is based on the same estimates $\hat{\beta}$ as Figure 1a. Figure 2f is based on the same estimates $\hat{\beta}$ as Figure 1b.



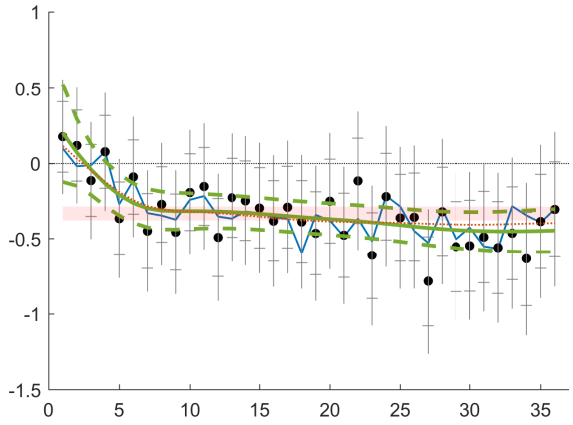
(a) treatment path constant



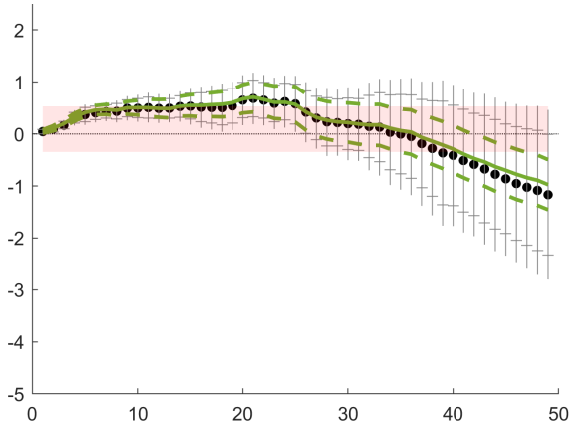
(b) treatment path smooth, eventually flat



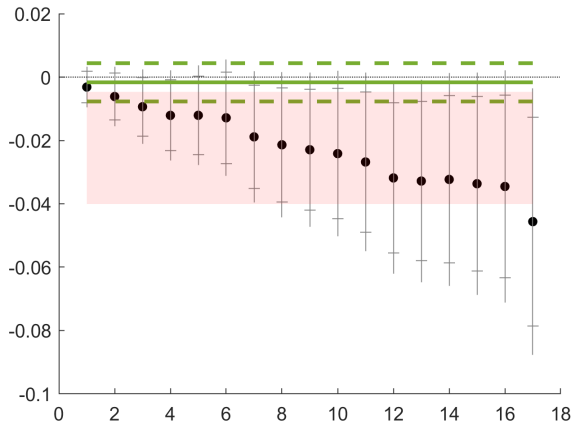
(c) treatment path hump-shaped



(d) treatment path wiggly



(e) Replicated from Nakamura and Steinsson [2018] (Figure A.2)



(f) Replicated from Bosch and Campos-Vazquez [2014] (Figure 4b)

Figure 2: Exemplary treatment effect plots including our proposals. Our proposed visualization includes two additional objects. The shaded red areas provide the *cumulative plausible bounds*. The dashed green lines provide the *restricted plausible bounds*, and the thick solid green line provides the corresponding restricted estimates. In the simulated panels (a)-(d), we further include the true treatment path (thin blue line) and its *surrogate* (dotted red line).

plausible. These cumulative plausible bounds have an alternative interpretation in terms of the overall treatment effect path: The treatment path β (depicted as the solid blue line) will *on average* be within these bounds for 95% of all realizations of the data.

The dashed green lines represent our proposed 95% *restricted plausible bounds*, which are centered around *restricted estimates* provided by the solid green line. These restricted estimates and bounds are motivated by envisioning a researcher who is interested in understanding key features of the treatment effect path but is not concerned with necessarily covering the entire true path at every horizon. However, we also imagine the researcher as being *ex ante* unsure about what the important features are and wanting to use the data to help select a restricted model for summarizing the treatment effect path.

More concretely, we construct the restricted estimates and plausible bounds by using a statistical model selection procedure to select an approximating model from within a pre-specified universe of candidates. We consider a default set of models motivated by a preference for smooth dynamics that eventually die out induced by shrinking first and third differences of $\hat{\beta}$, though the procedure could be applied with any finite, pre-specified universe of models. The restricted estimates are then simply the point estimates of the treatment path based on the selected model. We construct the restricted plausible bounds to provide uniform (95%) coverage accounting for data-dependent model selection by applying Berk et al.'s [2013] post-selection inference (PoSI) to our setting.

Looking at the restricted estimates and restricted plausible bounds in each panel paints a starkly different picture compared to the sup-t intervals. In all cases, the restricted plausible bounds are relatively narrow and seemingly quite informative about the broad features of the treatment effect paths. Figure 2f stands out and merits further discussion. In this instance, our model selection procedure selects a constant treatment effects model. Our restricted estimates then coincide with the MLE estimate of a constant treatment effects model.[6] Remarkably, this estimate is -0.0017, which is outside of the convex hull of the individual estimates $\hat{\beta}_h$.[7] Further, a Wald test of the null hypothesis that $\beta_h = -0.0017$ for $h = 1, \ldots, H$ gives a p-value of 0.36. That is, a traditional joint hypothesis test suggests there is relatively little evidence against this hypothesis, which thus appears relatively "plausible," contrary to what a visual inspection of the traditional treatment effect plot might suggest. This

[6]That is, a model with $\hat{\beta} \sim N(\beta, V_\beta)$, where β is constant across h.

[7]Intuitively, this behavior results from the strong positive correlation in the estimates combined with more precise estimates in early periods. We note that this strong positive correlation in the estimates cannot be inferred from the traditional plot.

difference underscores that traditional treatment effect plots may be ineffective at visualizing the impact of the off-diagonal entries in V_β (we illustrate this further in Appendix A). In contrast, the entire covariance matrix V_β is reflected in our restricted plausible bounds. Accounting for the covariance structure can lead to interestingly different results, improving the informativeness of these plots.

Finally, we reiterate that the inferential target of the restricted plausible bounds in the population is *not* the true treatment path. Rather, the restricted plausible bounds provide uniform coverage of a *surrogate* path given by the approximation that would be obtained by applying the selected model to the true effect path. We depict the selected surrogate for each of the simulated scenarios in Figure 2 with a dotted red line. In Figure 2a and Figure 2b, this surrogate is indistinguishable from the true treatment path. In Figure 2c and Figure 2d, the surrogate differs from the true treatment path but visually captures what seem to be key features of the overall treatment path. Indeed, we suspect many empirical researchers, if given the true treatment path from Figure 2d, would actually be more interested in the smooth approximation provided by the surrogate in this case. That is, we view the fact that the restricted plausible bounds cover a data-dependent approximation to the population treatment effect path as a potentially appealing feature.

In summary, we propose augmenting standard event study plots with two additional elements: a shaded red region (the cumulative plausible bounds) and dashed and solid green lines (the restricted plausible bounds and estimates, respectively). Together with the usual pointwise and sup-t intervals, these visualizations offer a more comprehensive view of plausible effect paths, each serving distinct inferential purposes. Sup-t bands provide a simple, assumption-free summary of plausible paths. Pointwise intervals target effects at specific horizons. Cumulative bounds inform average treatment effects, while restricted bounds and estimates capture approximations of the effect path obtained using data-driven smoothing.

We obtain our restricted plausible bounds and estimates by considering a finite set of candidate models for β. This approach is thus closely related to work that considers parametric models and approximations to β (e.g., Almon [1965], Barnichon and Matthes [2018], Barnichon and Brownlees [2019]). Our restricted estimates are akin to point estimates that could be obtained by taking an empirical Bayes approach within the framework of Shiller [1973], which considers a closely related model universe and takes a fully Bayesian approach. See also the SmIRF estimator of Plagborg-Møller [2016] for a related approach that includes

confidence sets with guaranteed frequentist coverage. In contrast to all the papers above, our inferential target is the population value of the selected surrogate path; see, for example, Genovese and Wasserman [2008] for a general discussion of inference on surrogates.

Alternative approaches to quantifying and visualizing uncertainty about treatment effect paths include Sims and Zha [1999], who argue that conventional pointwise bands should be supplemented with measures of shape uncertainty. Jordà [2009] suggests a method to construct simultaneous confidence regions for impulse responses given propagation trajectories. Freyberger and Reeves [2018] propose a uniformly valid inference method for an unknown function or parameter vector satisfying certain shape restrictions. Inoue and Kilian [2016] suggests a "shotgun plot" which depicts a random sample of B impulse responses contained in the joint Wald confidence set to visualize plausible treatment effect paths. We believe our proposal to provide simple additional visual elements to the usual treatment effect plot provides a useful complement to this existing literature.

2 Cumulative Plausible Bounds

Our first visual feature, the *cumulative plausible bounds*, does not impose any functional form or smoothness assumptions on the underlying treatment path. However, rather than targeting uniform coverage of the full treatment effect path, the cumulative plausible bounds use a weaker notion of "cumulative coverage." These bounds are simply visualizations of the dynamic treatment paths corresponding to the largest and smallest average treatment effect up to horizon H not rejected by a standard hypothesis test. They are simply constructed as $(U, L)^{1-\alpha} = \{U_h^{1-\alpha}, L_h^{1-\alpha}\}_{h=1}^H = \{w'\hat{\beta}/H \pm z_{1-\alpha/2}(w'V_\beta w)^{\frac{1}{2}}/H\}_{h=1}^H$, where $z_{1-\alpha/2}$ is the $(1-\alpha/2)$ quantile of the standard normal distribution, and w is a vector of ones of length H.[8]

These cumulative plausible bounds also have an interpretation in terms of coverage for the event path. The following proposition states that, for a given significance level α, the true treatment path will *on average* be within our bounds for $(1-\alpha)$ of all realizations. This follows immediately from the fact that any path that is not, on average, inside the cumulative plausible bounds implies an overall treatment effect over H periods that is rejected by the corresponding hypothesis test.

[8]Instead of using bounds $(U, L)^{1-\alpha}$ that are constant across h, one could alternatively depict bounds that reflect the shape of the unrestricted estimates. See Appendix B.

Proposition 1. *The true treatment path β will on average (over H) be within the cumulative plausible bounds for $(1-\alpha)$% of all realizations:*

$$\mathbb{P}\left(\sum_{h=1}^{H} L_h^{1-\alpha} < \sum_{h=1}^{H} \beta_h < \sum_{h=1}^{H} U_h^{1-\alpha}\right) = (1-\alpha). \tag{1}$$

Proof. By construction, a t-test on the average treatment effect with significance level $(1-\alpha)$ will reject

a) any treatment path $\tilde{\beta}_h$ with $\frac{\sum_{h=1}^{H}\tilde{\beta}_h}{H} > \frac{\sum_{h=1}^{H} U_h^{1-\alpha}}{H}$ (i.e., with $\sum_{h=1}^{H}\tilde{\beta}_h > \sum_{h=1}^{H} U_h^{1-\alpha}$),

b) any treatment path $\tilde{\beta}_h$ with $\frac{\sum_{h=1}^{H}\tilde{\beta}_h}{H} < \frac{\sum_{h=1}^{H} L_h^{1-\alpha}}{H}$ (i.e., with $\sum_{h=1}^{H}\tilde{\beta}_h < \sum_{h=1}^{H} L_h^{1-\alpha}$).

Since a t-test for the average effect has correct size, the result follows immediately. □

3 Restricted Plausible Bounds

The second idea we pursue is to present confidence regions that cover approximations of the true effect path that have "reasonable shapes." We term these confidence regions *restricted plausible bounds*. Here, we define "reasonable shapes" by pre-specifying a universe of models. We then use data-dependent model selection to choose a good representation for $\hat{\beta}$ from among this set. Intuitively, this approach is related to directly imposing a functional form restriction as is often done in empirical work, for example, by

- specifying a parametric model for β, e.g., imposing a constant treatment effects model $(\beta_h = \beta_{h'}\ \forall h, h')$,
- aggregating the underlying dataset over time (e.g., monthly to quarterly), which effectively restricts β to "step functions,"
- estimating an impulse response function (IRF) via a vector auto regression (VAR), which restricts the IRF to functional forms compatible with the chosen VAR (cf. the discussions in Li et al. [2024] and Olea et al. [2024]).

One key feature of our approach is that we do not rely on a fixed functional form restriction or make use of some other implicit device to choose a restricted model. Rather, we select a model, and then take model selection explicitly into account when constructing confidence

bounds. That is, we propose a model selection procedure that is explicit, transparent, and will allow us to maintain formal coverage guarantees.

Before we formally define our proposal, we introduce some necessary notation. We first borrow from the nonparametric statistics literature to introduce the notion of a "surrogate" (cf. Genovese and Wasserman [2008]). A surrogate path β_M is close to, but potentially simpler than, β. We note that the surrogate path is a population object that approximates β, the true treatment path. For example, we may define a constant treatment effects surrogate of β as $\beta_s = \arg\min_b (\beta - b)'(\beta - b)$ s.t. $\Delta b = 0$. If the surrogate model M is fixed a priori (and not itself a function of the data), inference for β_M is straightforward, though we stress that any inferential statements in this case will be about β_M and not β.[9] However, failing to take into account that the data is used to select the surrogate creates a problem for inference (e.g., Leeb and Pötscher [2005] or Roth [2022]). In our setting the surrogate is explicitly a function of the data (or more precisely, of the unrestricted estimates $\hat{\beta}$), and we may thus write $\beta_{M(\hat{\beta})}$ to denote a data-dependent surrogate path. In a first step, we use the data to select the surrogate model. In a second step, we then create a uniformly valid confidence region for the selected surrogate path, taking into account that the choice of surrogate is also random (i.e., a function of the data).

Given that we are doing model selection from a specified universe of models, a key choice is the specific model universe we consider. We consider a model universe motivated by the following economic intuition:

1. The dynamics of the treatment effect die off eventually. That is, after K periods, the treatment effect is constant. We treat K as unknown and allow K to be as large as H, thus allowing dynamics across the entire depicted horizon.

2. The dynamic treatment path is "smooth," where we measure smoothness using the third differences of the treatment path.

In practice, we use shrinkage over first and third differences of $\hat{\beta}$ to implement 1. and 2.

Formally, we assume that the estimates of the treatment path $\hat{\beta}$ are jointly normal with $\hat{\beta} \sim N(\beta, V_\beta)$, where $V_\beta = \sigma^2 V$, $\sigma^2 = \frac{1}{H}\sum_{h=1}^{H} V_\beta(h,h)$, and V is positive-definite. Taking $\hat{\beta}$

[9] Targeting a simple surrogate function is akin to the standard approach in economics of estimating linear models even when the conditional expectation function is not believed to be linear. One can think of the linear model as a "surrogate model" capturing the best linear predictor. Inference will then be about the linear surrogate, and not the "truth."

as input, we define the following object:

$$\begin{aligned}\tilde{\beta}(\lambda_1, \lambda_2, K) &= \arg\min_b \; Q(b, \lambda_1, \lambda_2, K) \\ &= \arg\min_b \; \underbrace{(\hat{\beta} - b)' V^{-1} (\hat{\beta} - b)}_{\text{distance from } \hat{\beta}} + \lambda_1 \underbrace{b' D_1' W_1(K) D_1 b}_{\substack{\text{penalty on first difference} \\ \text{after horizon K}}} + \lambda_2 \underbrace{b' D_3' W_3 D_3 b,}_{\substack{\text{penalty on} \\ \text{third difference}}} \end{aligned} \tag{2}$$

where

- D_1 and D_3 are the $(H \times H-1)$ and $(H \times H-3)$ first and third difference operators,
- $\eta = (\lambda_1,\, \lambda_2,\, K)$ are tuning parameters,
- $W_1(K)$ and W_3 are weighting matrices, where $W_1(K)$ only places weight on first differences for horizon $K \leq H$ and beyond (see Appendix C for further details).

Solving (2) provides a closed form solution[10] for $\tilde{\beta}(\lambda_1, \lambda_2, K) := \tilde{\beta}(M)$ given by

$$\begin{aligned}\tilde{\beta}(M) &= \left(V^{-1} + \lambda_1 D_1' W_1(K) D_1 + \lambda_2 D_3' W_3 D_3\right)^{-1} V^{-1} \hat{\beta} \\ &= P(M)\hat{\beta}.\end{aligned}$$

For fixed $M = (\lambda_1, \lambda_2, K)$, it immediately follows that

$$\tilde{\beta}(M) - P(M)\beta \sim N(0, V_M), \tag{3}$$

where $V_M = P(M) V_\beta P(M)'$. Here, $P(M)\beta = \beta_M$ defines a particular surrogate path for β. Intuitively, $P(M)\beta$ corresponds to a "projection" of the true treatment path β into a lower dimensional space. Given (3), it would be straightforward to construct a confidence region for $\{\beta_{M,h}\}_{h=1}^H$, where $\beta_{M,h}$ denotes the h^{th} entry in vector β_M, for a given, fixed value of the tuning parameters. However, knowing *ex ante* what values to use for λ_1, λ_2, and K seems challenging. We thus use model selection to choose λ_1, λ_2, and K — or, equivalently, to choose the surrogate model M.

[10] For intuition, note that the problem in (2) is closely related to the following constrained optimization with tuning parameters c_1, c_2, and K, explicitly bounding the first and third difference:

$$\begin{aligned}\hat{\beta}(c_1, c_2, K) = \arg\min_b \; & \underbrace{(\hat{\beta} - b)' V^{-1} (\hat{\beta} - b)}_{\text{distance from } \hat{\beta}} \\ \text{such that} \quad & \underbrace{b' D_1' W_1(K) D_1 b \leq c_1}_{\substack{\text{small first difference,} \\ \text{after horizon K}}} \qquad \text{and} \qquad \underbrace{b' D_3' W_2 D_3 b \leq c_2}_{\text{small third difference}}.\end{aligned}$$

However, this formulation is computationally more challenging, making it less appealing.

Specifically, we use the estimated $\hat{\beta}$ and an object akin to an information criterion to select the surrogate M. First, note that we can construct the "residuals" $\hat{\beta} - \tilde{\beta}(M) = \hat{\beta} - P(M)\hat{\beta} = (I - P(M))\hat{\beta}$. We use this residual formulation to define an analog of model degrees of freedom given by $\text{df}(M) = \text{trace}(P(M))$. We then select a model that minimizes a BIC analog over $\mathcal{M}$, where $\mathcal{M}$ denotes the universe of values for $M = (\lambda_1, \lambda_2, K)$:

$$\hat{M} = \underset{M \in \mathcal{M}}{\arg\min}\ (\hat{\beta} - \tilde{\beta}(M))' V_{\beta}^{-1} (\hat{\beta} - \tilde{\beta}(M)) + \log(H)\text{df}(M).$$

We tie the researcher's hands by pre-specifying $\mathcal{M}$, the universe of models considered. In our implementation, $\mathcal{M}$ includes surrogate models corresponding to a constant, linear, quadratic and cubic treatment effect path (with one, two, three, and four degrees of freedom, respectively), as well as an unrestricted model corresponding to the unrestricted estimates $\hat{\beta}$ (with H degrees of freedom). $\mathcal{M}$ further includes surrogate models corresponding to surrogate paths of the form $P(M)\beta = \beta_M$ using a grid over $(\lambda_1, \lambda_2, K)$. We discuss our implementation in more detail in Appendix C and visualize the model universe $\mathcal{M}$ for four exemplary treatment paths β in Appendix Figure 5, but note that $\mathcal{M}$ does not depend on $\hat{\beta}$ or σ.

Remark 1. *One could use other model universes and shrinkage methods. Examples of alternative approaches include Barnichon and Matthes [2018] and Barnichon and Brownlees [2019]. We have chosen a class that we believe will be a reasonable representation of beliefs in many applications. Our approach with this model class is also particularly easy computationally, which allows us to nest a large universe of models. Likewise, one could select the surrogate model M using methods other than minimizing our BIC analog. We found that the specific structure and estimation we employ performed well across our simulations.*

Given the selected surrogate $\hat{M}$, we define the *restricted estimates* as $\tilde{\beta}(\hat{M})$. However, we cannot directly apply (3) to obtain a valid confidence region for the population value of the surrogate path $\beta_{\hat{M}} = P(\hat{M})\beta$ because $\hat{M}$ was selected by looking at the data, $\hat{\beta}$. Thus, in a second step, we use valid post-selection inference (Berk et al. [2013]), which explicitly accounts for data-dependent (and thus random) model selection, to construct a uniformly valid confidence region for $\beta_{\hat{M}}$. These confidence intervals, our *restricted plausible bounds*, are rectangular regions of the form $CR^{POSI} = \{\ell_h(X), u_h(X)\}_{h=1}^{H}$ for $[\ell_h(X), u_h(X)] = [\tilde{\beta}(\hat{M})_h \pm C^{\alpha} V_{\hat{M}}^{1/2}(h,h)]$, where $\tilde{\beta}(\hat{M})_h$ denotes the restricted estimate of the effect at horizon h and $V_{\hat{M}}^{1/2}(h,h)$ is the square root of the h^{th} diagonal entry of $V_{\hat{M}}$. To ensure uniform validity we use the "PoSI constant" of Berk et al. [2013] as C^{α}, defined as the minimal value that

satisfies

$$\mathbb{P}\left(\max_{M\in\mathcal{M}}\max_{h}|t_{h\cdot M}|\leq C^{\alpha}\right)\geq(1-\alpha),$$

where $t_{h\cdot M}=V_M^{-1/2}(h,h)\xi_h$, and ξ_h is the h^{th} element of multivariate normal vector ξ with mean $\mathbf{0}_H$ and variance V_M. Importantly, C^{α} depends on $\mathcal{M}$, the universe of models considered, but not on the model selection procedure.

The following proposition is a direct application of Berk et al. [2013].

Proposition 2. *For any treatment path β, we obtain valid coverage for its surrogate $\beta_{\hat{M}}$:*

$$\mathbb{P}[\beta_{\hat{M}}\in CR^{POSI}]\geq 1-\alpha. \tag{4}$$

Proof. This follows immediately from the guarantees in Berk et al. [2013]:

$$\mathbb{P}(\beta_M\in CR^{POSI}|\hat{M}=M)\geq 1-\alpha.$$

□

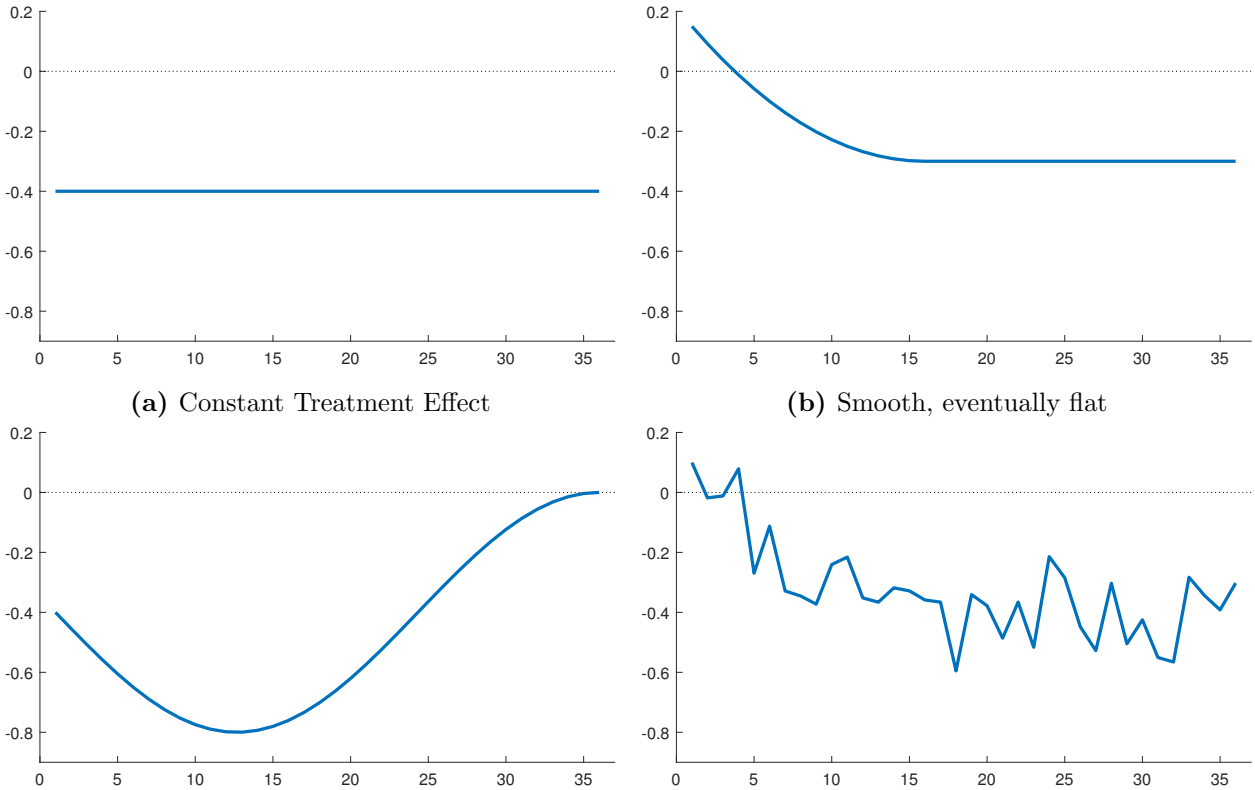
Proposition 2 guarantees that our restricted plausible bounds cover the selected surrogate to the truth in at least $(1-\alpha)\%$ of sample realizations.

Remark 2. *An immediate consequence of Proposition 2 is that $\mathbb{P}(\beta\in CR^{POSI})\geq 1-\alpha$ if $\mathbb{P}\left(\beta_{\hat{M}}=\beta_M=\beta\right)=1$. That is, in cases where model selection is effectively non-random and the selected surrogate path coincides with β, the restricted plausible bounds will also provide valid coverage for the true treatment path. Given the form of our BIC type objective for selecting $\hat{M}$ and that the unrestricted estimates are always included in our default model universe, one could provide conditions for $\mathbb{P}\left(\beta_{\hat{M}}=\beta_M=\beta\right)=1$ under a sequence of models where $\sigma^2\to 0$ and surrogate paths were well-separated – e.g. where $\|\beta-\beta_M\|\geq\delta>0$ for all candidate models M such that $\beta_M\neq\beta$. While technically possible, we i) question the utility of this perspective in offering a useful finite sample approximation and ii) view the surrogate as an economically interesting summary of the treatment path in itself.*

Remark 3. *Throughout, we work with the unrestricted estimates $\hat{\beta}$. An alternative would be to estimate the restricted models directly on the data. In settings where $\hat{\beta}\sim N(\beta,V_\beta)$ provides a good approximation, we suspect that such an approach will yield qualitatively similar results. It may be interesting to explore directly estimating restricted models in settings where the approximation $\hat{\beta}\sim N(\beta,V_\beta)$ is questionable.*

4 Numerical Results

In this section, we illustrate the properties of our restricted plausible estimates and bounds as well as our cumulative plausible bounds in simulation experiments with treatment paths generated to resemble treatment path dynamics that practitioners may encounter. We illustrate these treatment paths in Figure 3. We consider a constant treatment effect path (cf. Figure 3a); a treatment path that smoothly declines before flattening out (cf. Figure 3b); a hump-shaped treatment path with dynamics that continue for the entire H periods (cf. Figure 3c); and a "wiggly" treatment path (cf. Figure 3d). We describe the exact DGP for each of the four panels in more detail in Appendix Table 1.[11] The object of interest is the treatment path over a 36-month horizon. We have access to jointly normal estimates $\{\hat{\beta}_h\}_{h=1}^{36}$. For each of these four treatment paths, we then draw 1,000 realizations of $\hat{\beta} \sim N(\beta, V_\beta)$.[12]



(a) Constant Treatment Effect

(b) Smooth, eventually flat

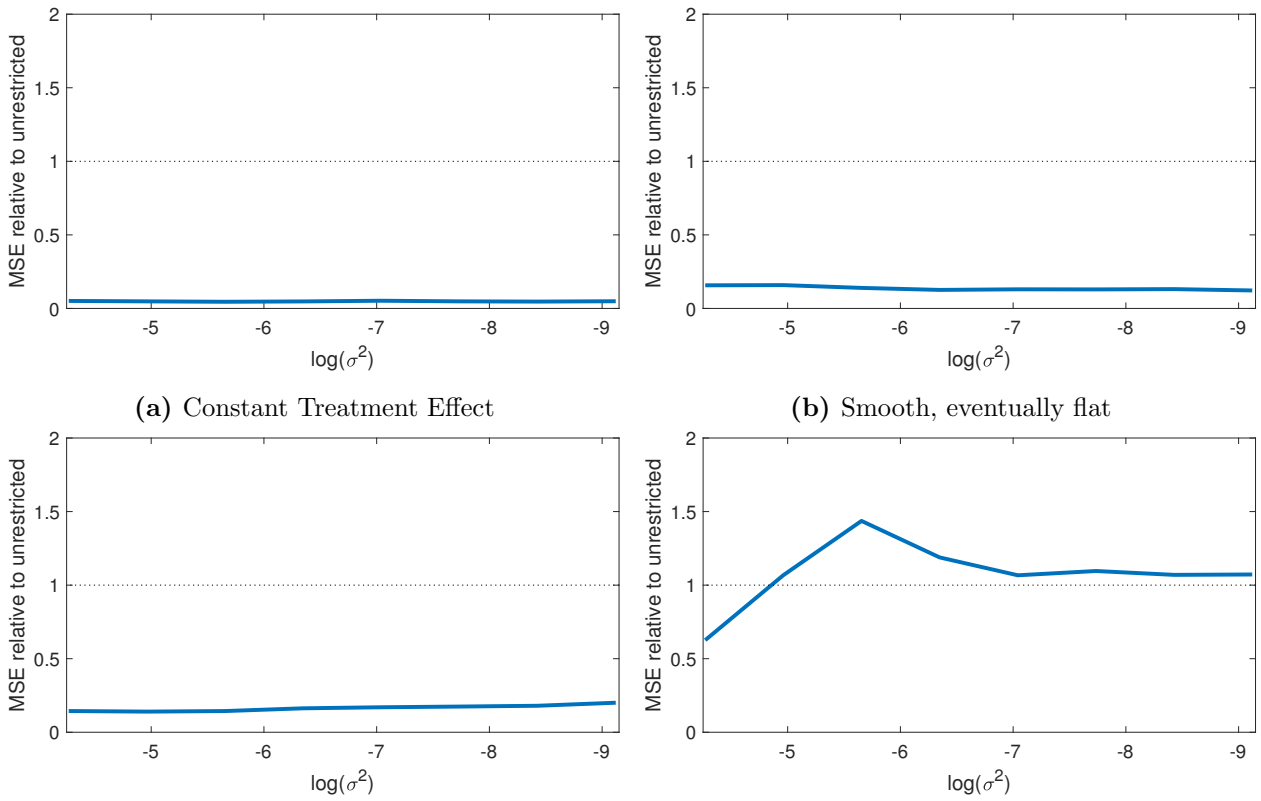
(c) Hump-shaped

(d) Wiggly

Figure 3: Four exemplary treatment effect paths β.

[11] Figure 2 provides one example realization from each of these DGPs.

[12] In the figures that follow, V_β is diagonal with its entries specified in Appendix D. We repeat our exercise with more general covariance matrices in Appendix F.



(a) Constant Treatment Effect

(b) Smooth, eventually flat

(c) Hump-shaped

(d) Wiggly

Figure 4: Relative performance of restricted and unrestricted estimators. Depicted is the ratio in mean-squared error between restricted and unrestricted estimates, $\frac{MSE(\tilde{\beta}(\hat{M}))}{MSE(\hat{\beta})}$, as a function of the amount of noise σ^2 in the initial estimates $\hat{\beta}$.

We first compare the point estimation properties of the unrestricted estimates $\hat{\beta}$ with our restricted estimates $\tilde{\beta}(\hat{M})$ for each of these four scenarios. In particular, Figure 4 depicts the ratio in mean-squared error, $MSE_{\tilde{\beta}(\hat{M})}/MSE_{\hat{\beta}}$, as a function of σ^2, which scales the covariance matrix of the estimates, V_β (see Appendix D for more detail).

The largest value of σ^2 in Figure 4 (corresponding to the left most point) thus represents relatively noisy estimates.[13] We conclude that in most cases our restricted estimator has excellent point estimation properties when the target is the true treatment path β. In the three panels that have a smooth treatment path (Figures 4a-4c), the MSE of our restricted estimate is a full order of magnitude lower compared to the unrestricted estimate. In Figure 4d, our restricted estimate has a lower MSE when estimates are very noisy, a higher MSE

[13]Figure 2 was created from a single realization of the the left most point in Figure 4. To give the reader a sense of the scale of the x-axes, Appendix Figure 4 also illustrates a single realization of the right most point in Figure 4.

for intermediate sizes of σ^2, and a similar MSE when σ^2 is small. However, we suspect that a lower-dimensional summary of β, as provided by the surrogate path, may in fact be the policy relevant object in cases where the true treatment path exhibits complicated dynamics as in this panel (cf. Figure 3d).

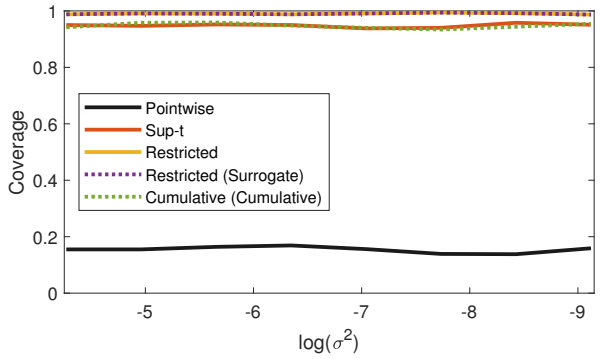
Figure 5 summarizes inference results, where we set $\alpha = 0.05$. We again consider the four previous DGPs and vary the amount of noise in the estimation of β by varying σ^2.

Coverage results are depicted in Figures 5a-5d. The pointwise, sup-t, and restricted coverage numbers, as indicated by the three solid lines, represent the empirical analogue to the usual notion of uniform coverage for β: It reflects the fraction of simulations in which the true treatment path β falls entirely inside the pointwise confidence intervals (black), sup-t intervals (red), and restricted plausible bounds (yellow), respectively.

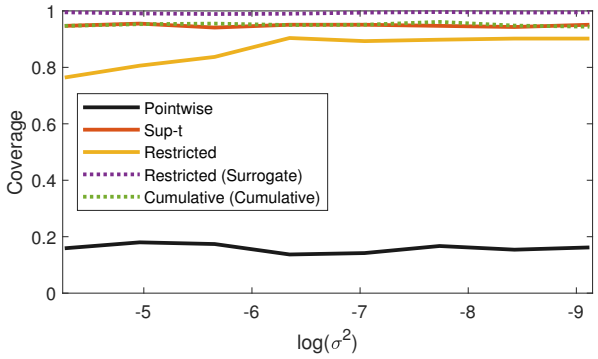
Across all four DGPs, the pointwise confidence intervals cover the true treatment path in only 15-20% of simulations. On the other hand, the sup-t intervals achieve nominal coverage across all DGPs. The restricted plausible bounds are not constructed to provide uniform coverage for the true treatment path. It is thus not surprising that these bounds do not cover the entire treatment path in 95% of simulations across all DGPs. However, we do see that their coverage appears to converge towards 95% as the amount of noise in the initial estimates decreases (cf. Remark 2). Further, the restricted plausible bounds perform substantially better than the pointwise confidence intervals in covering the true treatment path when the true path is smooth. In the scenario of a wiggly treatment path, the restricted bounds exhibit poor coverage properties when the amount of noise in the initial estimates is large.

Proposition 2 guarantees that the plausible bounds cover the selected surrogate to the truth in at least 95% of realizations. In Figures 5a-5d, the restricted plausible bounds' coverage of the surrogate is illustrated by the dashed purple line. We see that this coverage is above 95% for all levels of σ^2 across all DGPs. We reemphasize that, in the scenario of a wiggly treatment path, a smooth approximation to the true path, for which we obtain valid coverage, may be a policy relevant object. Finally, we note that, in line with Proposition 1, the cumulative bounds (as indicated by the dashed green lines) indeed cover the cumulative effect of the policy in 95% of simulations. That is, the true treatment path is on average within these bounds in 95% of simulations.

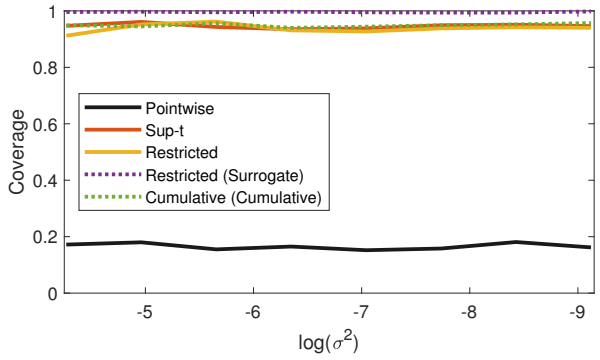
In order to cover the true path in a higher fraction of simulations, the sup-t bands are



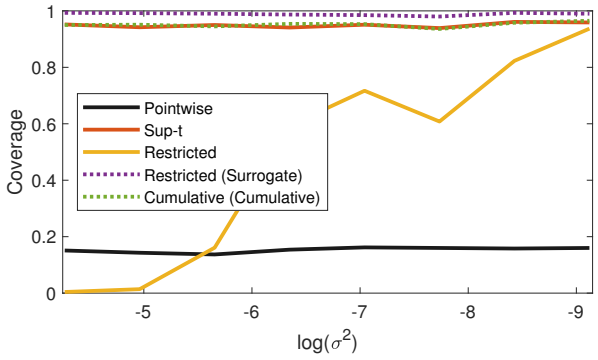
(a) Coverage. Constant Treatment Effect



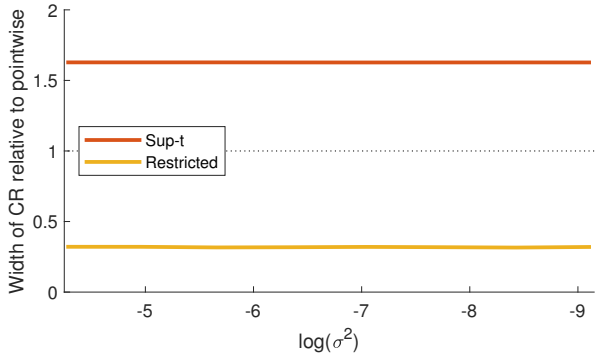
(b) Coverage. Smooth, eventually flat



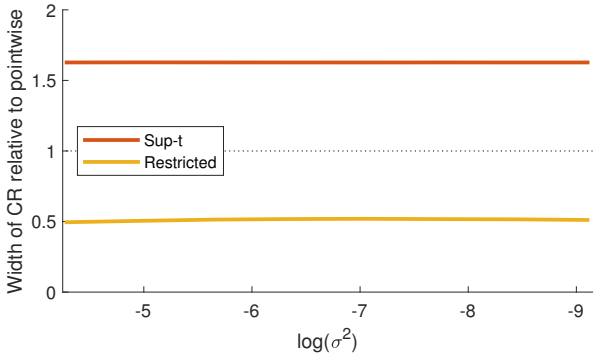
(c) Coverage. Hump-shaped



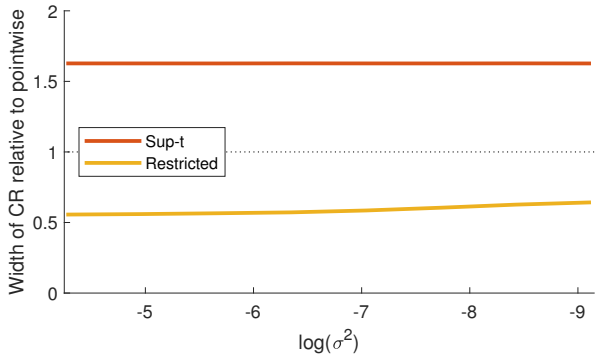
(d) Coverage. Wiggly



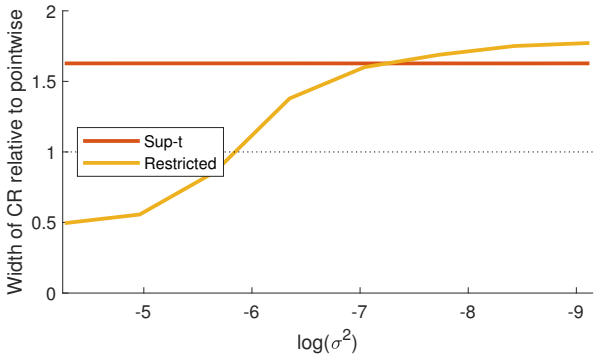
(e) Width. Constant Treatment Effect



(f) Width. Smooth, eventually flat



(g) Width. Hump-shaped



(h) Width. Wiggly

Figure 5: Inference results. Panels (a)-(d) provide coverage properties of various confidence regions as a function of the amount of noise σ^2 in the initial estimates $\hat{\beta}$. Panels (e)-(h) provide width of sup-t and plausible bounds relative to the width of pointwise intervals.

generally wider than the pointwise bands (cf. Figure 1). We illustrate this across our simulations in Figures 5e-5h, which depict the average width of the sup-t bands and the restricted bounds, relative to the pointwise intervals. Across most scenarios, the restricted bounds are narrower than both pointwise and sup-t bands, despite offering improved coverage. For example, in the constant treatment effect case, restricted bounds are less than half as wide as pointwise bands and less than a quarter as wide as sup-t bands, yet still achieve close to 95% coverage. The restricted bounds are only wider than the sup-t bands when (i) the treatment path is wiggly and (ii) there is very little noise. In this region of the parameter space, our model selection mechanism results in relatively little smoothing. Our restricted plausible bounds are then slightly more conservative than the sup-t bands, which follows immediately from the results in Berk et al. [2013].[14]

5 Conclusion

We propose restricted and cumulative plausible bounds as new visualization tools for dynamic treatment effect paths. By targeting objects other than uniform coverage of the entire path, these bounds can be substantially tighter than standard pointwise and uniform confidence bands, and thus provide more informative summaries when traditional bands are uninformative. Our approach can also yield markedly different conclusions when there is strong correlation in the estimates. As a byproduct, our restricted estimates deliver improved point estimates in settings where the effect path is smooth.

It may be interesting to explore the notion of explicitly covering data-dependent surrogates in other economically relevant settings. Although we have focused on treatment path estimation, our approach — shrinking unrestricted models toward lower-dimensional subspaces using data-driven ridge penalties — applies more broadly. For example, an alternative application may include shrinking estimates towards theoretical restrictions implied by economic theory as in Fessler and Kasy [2019]. Further, it would be interesting to explore related ideas in more structural settings, such as selecting among or regularizing moment conditions in GMM-type frameworks.

[14]We illustrate the model selection of our procedure further in Appendix Figures 6-8. Essentially, we frequently select surrogate paths β_M with significantly reduced degrees of freedom. This effective dimension reduction explains both the improvement in point estimation properties of $\tilde{\beta}(\hat{M})$ relative to the unrestricted estimates $\hat{\beta}$ documented in Figure 4, and the decrease in width of the confidence intervals documented in Figure 5.

References

Shirley Almon. The distributed lag between capital appropriations and expenditures. *Econometrica*, 33(1):178–196, 1965.

Regis Barnichon and Christian Brownlees. Impulse response estimation by smooth local projections. *Review of Economics and Statistics*, 101(3):522–530, 2019.

Regis Barnichon and Christian Matthes. Functional approximation of impulse responses. *Journal of Monetary Economics*, 99:41–55, 2018.

Richard Berk, Lawrence Brown, Andreas Buja, Kai Zhang, and Linda Zhao. Valid post-selection inference. *The Annals of Statistics*, 41(2):802–837, 2013.

Mariano Bosch and Raymundo M Campos-Vazquez. The trade-offs of welfare policies in labor markets with informal jobs: The case of the "seguro popular" program in Mexico. *American Economic Journal: Economic Policy*, 6(4):71–99, 2014.

Brantly Callaway and Pedro H.C. Sant'Anna. Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2):200–230, 2021.

Pirmin Fessler and Maximilian Kasy. How to use economic theory to improve estimators: Shrinking toward theoretical restrictions. *Review of Economics and Statistics*, 101(4): 681–698, 2019.

Joachim Freyberger and Yoshiyasu Rai. Uniform confidence bands: Characterization and optimality. *Journal of Econometrics*, 204(1):119–130, 2018.

Joachim Freyberger and Brandon Reeves. Inference under shape restrictions. *Available at SSRN 3011474*, 2018.

Christopher Genovese and Larry Wasserman. Adaptive confidence bands. *The Annals of Statistics*, 36(2):875–905, 2008.

Atsushi Inoue and Lutz Kilian. Joint confidence sets for structural impulse responses. *Journal of Econometrics*, 192(2):421–432, 2016.

Òscar Jordà. Simultaneous confidence regions for impulse responses. *The Review of Economics and Statistics*, 91(3):629–647, 2009.

Òscar Jordà. Local projections for applied economics. *Annual Review of Economics*, 15(1): 607–631, 2023.

Hannes Leeb and Benedikt M. Pötscher. Model selection and inference: Facts and fiction. *Econometric Theory*, 21(1):21–59, 2005.

Dake Li, Mikkel Plagborg-Møller, and Christian K Wolf. Local projections vs. VARs: Lessons from thousands of DGPs. *Journal of Econometrics*, 244(2):105722, 2024.

Emi Nakamura and Jón Steinsson. Identification in macroeconomics. *Journal of Economic Perspectives*, 32(3):59–86, 2018.

José Luis Montiel Olea and Mikkel Plagborg-Møller. Simultaneous confidence bands: Theory, implementation, and an application to SVARs. *Journal of Applied Econometrics*, 34(1): 1–17, 2019.

José Luis Montiel Olea, Mikkel Plagborg-Møller, Eric Qian, and Christian K Wolf. Double robustness of local projections and some unpleasant varithmetic. Technical report, National Bureau of Economic Research, 2024.

Mikkel Plagborg-Møller. *Essays in macroeconometrics*. PhD thesis, 2016.

Jonathan Roth. Pretest with caution: Event-study estimates after testing for parallel trends. *American Economic Review: Insights*, 4(3):305–322, 2022.

Robert J Shiller. A distributed lag estimator derived from smoothness priors. *Econometrica*, 41(4):775–788, 1973.

Christopher A Sims and Tao Zha. Error bands for impulse responses. *Econometrica*, 67(5): 1113–1155, 1999.