(Visualizing) Plausible Treatment Effect Paths

Simon Freyaldenhoven

Federal Reserve Bank of Philadelphia

Christian Hansen

University of Chicago[*]

November 12, 2024

Abstract

We consider point estimation and inference for the treatment effect path of a policy. Examples include dynamic treatment effects in microeconomics, impulse response functions in macroeconomics, and event study paths in finance. We present two sets of plausible bounds to quantify and visualize the uncertainty associated with this object. One set of bounds covers the average (or overall) effect rather than the entire treatment path. Our second set of bounds imposes data-driven smoothness restrictions on the treatment path. Post-selection Inference (Berk et al. [2013]) gives us formal coverage guarantees for these bounds. The chosen restrictions also imply novel point estimates that perform well across our simulations. Both plausible bounds are often substantially tighter than traditional confidence intervals, and can provide useful insights even when traditional (uniform) confidence bands appear uninformative.

JEL Codes: C12, C13

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

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 $h = 1, \ldots H$. 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]; Boxell et al. [2024]). 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]

---

[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].
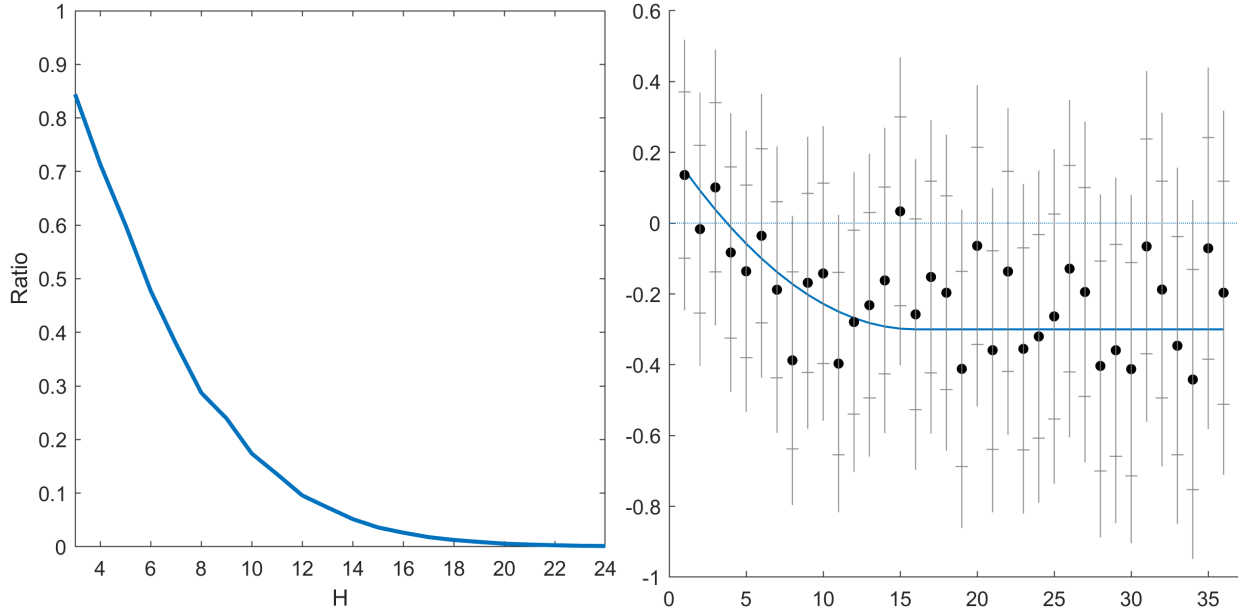
(a) Volume of Wald confidence region relative to sup-t confidence region as a function of $H = \dim(\beta)$.

(b) Exemplary event study plot, including true treatment effect (blue line), point estimates, and pointwise and sup-t confidence intervals.

Figure 1: Illustrative examples of confidence regions.

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 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 Wald region. We illustrate this difference in confidence region volume in Figure 1a, which compares the volume of Wald and sup-t confidence regions when V_β is an identity matrix of dimension H. While the volume of the Wald region is 85% of that of the sup-t region if $H = 3$, this ratio drops to less than 10% and around 0.1% with $H = 12$ and $H = 24$ respectively. These numbers are generally even smaller if the entries in $\hat{\beta}$ have non-zero correlation. It follows that the volume of the sup-t region tends to be orders of magnitude larger than the volume of the Wald region for the typical horizon that is depicted in event studies and impulse responses. 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

2

seems unappealing to us. We thus argue that sup-t confidence bands may not always be appropriate for visualizing what dynamic treatment effect paths are plausible and suggest complementary visualizations.

Figure 1b provides an exemplary visualization meant to mimic event study plots seen in empirical work in a hypothetical example. The object of interest is the treatment path of a policy over a 36-month horizon. The blue line in the figure depicts the true treatment path.[3] In this example, the treatment has an initially positive effect. The cumulative effect then diminishes and quickly becomes negative with dynamics lasting 17 periods before flattening out. We have access to jointly normal estimates $\{\hat{\beta}_h\}_{h=1}^{36}$, with observed point estimates $\hat{\beta}$ given by the black dots. Figure 1b further includes both pointwise 95 percent confidence intervals (inner confidence set as indicated by the dashes) and uniform 95 percent 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 percent of realized samples.

We note that the sup-t region in this example 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 compared to β. 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}. More generally, a reasonable conclusion based on the depicted sup-t region might be that the data are largely uninformative about β. On the other hand, related to our discussion of Figure 1a, only around one path out of every 100,000 sampled paths will not be rejected by a joint test if we uniformly sample treatment paths from the sup-t region depicted in Figure 1b. This discrepancy between the easy-to-visualize sup-t region and the results of simple joint hypothesis tests suggests to us that the sup-t confidence region may not 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. Specifically, each panel in Figure 2 presents an event study plot produced from a different data generating process. In each plot, the solid blue line represents the true
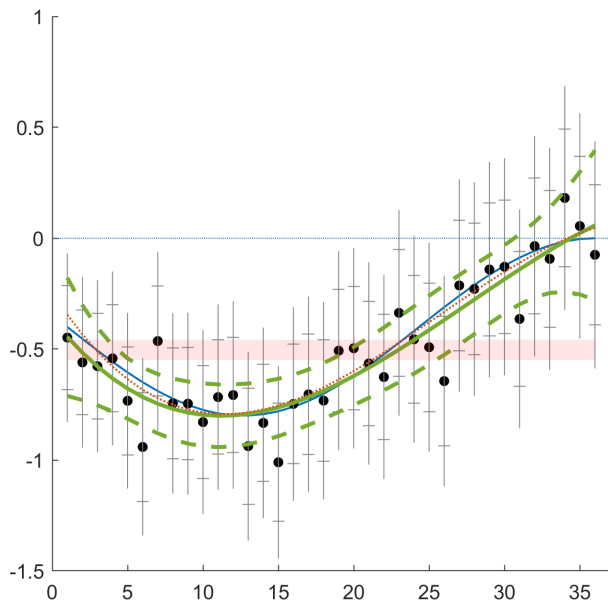
---

[3]We give more detail on the underlying DGP in Section 4.
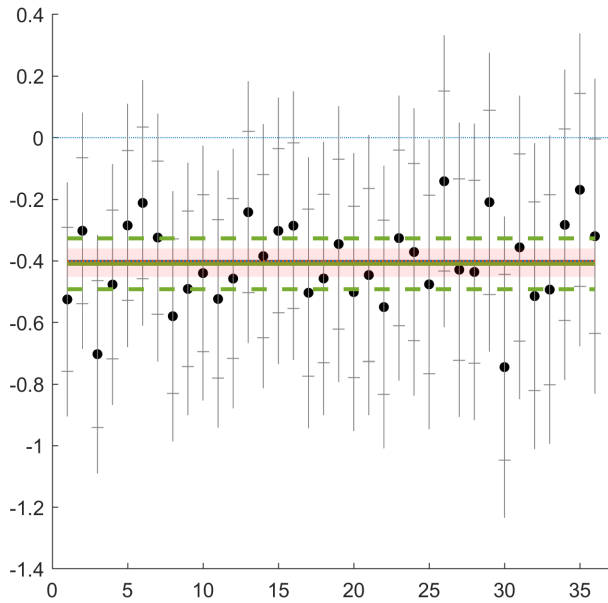
(a) treatment path constant
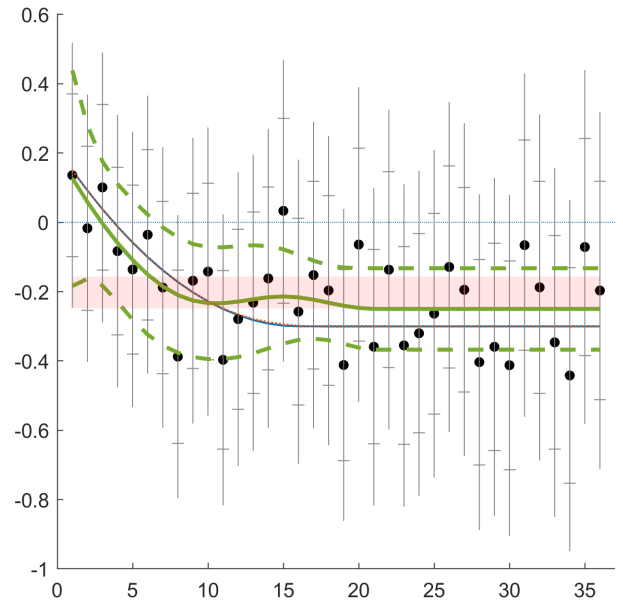
(b) treatment path smooth, eventually flat

(c) treatment path hump-shaped
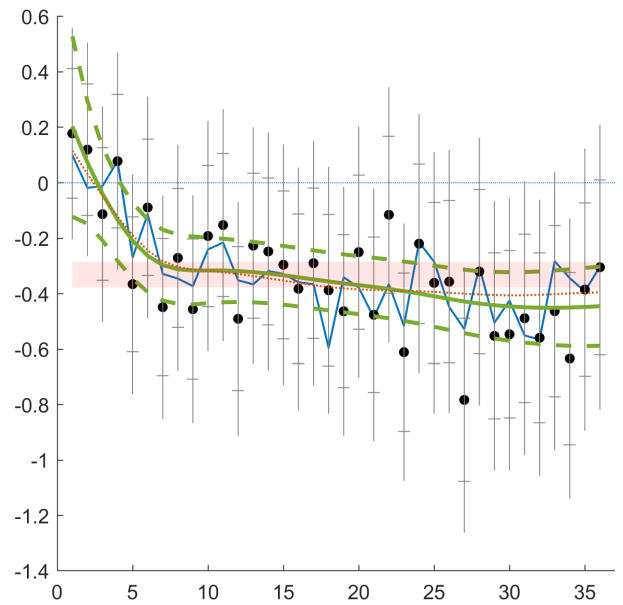
(d) treatment path wiggly

Figure 2: Four exemplary event-study plots including our proposals. Our proposed visualization includes three additional objects. The shaded area provides the *cumulative plausible bounds*. The dashed green lines provide the *restricted plausible bounds*, and the thick solid green line provides the corresponding restricted estimates. We further include the true treatment path (thin blue line) in each generated dataset and its *surrogate* (dotted red line).

treatment effect path. Elements from the "standard" event study plot are provided by the black dots which give point estimates of the treatment effect at each horizon and by the inner and outer confidence intervals corresponding respectively to the usual pointwise and sup-t confidence intervals.[4] In addition, each plot includes three new features: (i) the shaded red area, (ii) the dashed and solid green lines, and (iii) a dotted red line.

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 plausible. These cumulative plausible bounds have an alternative interpretation in terms of the overall treatment effect path. Specifically, the cumulative plausible bounds are such that 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 correspond to our proposed *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 we note that the procedure could be applied with any finite, pre-specified universe of models. The restricted estimates are then simply the point estimates

---

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

of the treatment path based on the selected model. We construct the restricted plausible bounds to provide uniform coverage accounting for data-dependent model selection by directly applying Berk et al.'s (2013) Post-Selection Inference (PoSI) within our setting. Looking at the restricted estimates and restricted plausible bounds in each scenario depicted in Figure 2 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.

Finally, it is important to be clear 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 scenario 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 an appealing feature within our setting.

In summary, our proposal is to augment plots typically provided in empirical papers with two additional visualizations (the shaded red area, and the dashed and solid green lines). The collection of visualizations, including standard pointwise and sup-t intervals, provides a more complete depiction of plausible effect paths with each element of the path being useful for related, but different, inferential goals. We believe that the conclusions a reader may draw when including the additional visualizations are interestingly different than those that can easily be drawn with only the traditional event study plot.

Our paper connects to several strands in the literature. 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. Such approaches have been studied going back at least to Almon [1965], who imposes a parametric

model on distributed lag coefficients. More recently, Barnichon and Matthes [2018] propose approximating impulse responses with a set of basis functions, and Barnichon and Brownlees [2019] propose to shrink impulse response estimates towards polynomials. While related, we differ from these approaches by focusing on inference for a data-dependent surrogate effect path; see, for example, Genovese and Wasserman [2008] for a general discussion of inference on surrogates.

We operationalize our restricted plausible bounds by using data-dependent selection from a universe of candidate models with different fixed degrees of shrinkage over first and third differences. This model universe is closely related to the structure employed in Shiller [1973] which takes a fully Bayesian approach to estimating a distributed lag model under a normal prior on the dth order difference of β. Our restricted estimates are thus akin to point estimates that could be obtained by taking an empirical Bayes approach within the framework of Shiller [1973]. From the empirical Bayes perspective, one could then potentially adapt Armstrong et al. [2022] to the present context to obtain interval estimates. While a potentially interesting avenue for further research, we explicitly do not leverage this connection. Our chief focus is the restricted plausible bounds which provide uniform frequentist coverage for the population value of the selected surrogate.

Our restricted plausible bounds provide uniform coverage for the selected surrogate accounting for data dependent model selection. To maintain these coverage guarantees, we use a version of post-selection inference (PoSI) confidence intervals (Berk et al. [2013]). Given that our inferential target is the population value of the selected model, we note that one could adopt other approaches from the literature on selective inference; see, e.g., Taylor and Tibshirani [2015] and Kuchibhotla et al. [2022] for excellent reviews.

There are a variety of other approaches to quantifying and visualizing uncertainty about treatment effect paths available in the literature. For example, Sims and Zha [1999] argues that conventional pointwise bands common in the literature should be supplemented with measures of shape uncertainty, and proposes such measures. Jordà [2009] provides 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. More generally,

inference for the treatment effect path is tightly tied to more general nonparametric inference problems; see, e.g., Chen et al. [forthcoming] for an interesting recent example that explicitly allows for use of a data-dependent sieve dimension. The "shotgun plot" of Inoue and Kilian [2016], which depicts a random sample of B impulse responses contained in the joint Wald confidence set, provides an interesting alternative approach focused on visualization of plausible treatment effect paths. We believe our proposal to provide simple additional visual elements to the usual event study plot provides a useful complement to this existing literature.

2 Cumulative Plausible Bounds

We first present a simple visual feature, the *cumulative plausible bounds*, that can be added to a standard event study plot. This visualization does not impose any functional form or smoothness assumptions on the underlying treatment path, but, in terms of the full event path, it also will not achieve uniform coverage. Rather, the cumulative plausible bounds use a weaker notion of "cumulative coverage" when thought about in terms of the full event path: The true treatment path will *on average* be within the cumulative plausible bounds in $(1 - \alpha)\%$ of all realizations of the data for a given significance level α. That is, by providing valid inference for the average effect over the horizon H, our cumulative plausible bounds provide a simple visual element that conveys uncertainty about the average treatment effect.

These cumulative plausible bounds are simply visualizations of the dynamic treatment path corresponding to the largest and smallest sum of all treatment effects up to horizon H not rejected by a standard hypothesis test. They can be interpreted as boundary paths that would be consistent with the upper and lower limits of a confidence interval for the overall effect of the policy over H periods. Formally, let

$$ub^{1-\alpha} = \max \sum_{h=1}^{H} \beta_h^* \qquad \text{s.t. } (\beta^* - \hat{\beta})' V_{\beta}^{-1} (\beta^* - \hat{\beta}) = \kappa^{(1-\alpha)}, \tag{1}$$

where $\kappa^{(1-\alpha)}$ denotes the inverse of the chi-square cdf with one degree of freedom at chosen significance level $(1 - \alpha)$. We further define $lb^{1-\alpha}$ analogously, replacing the max in (1) with min. Since both $ub^{1-\alpha}$ and $lb^{1-\alpha}$, corresponding to the upper and lower limit of the overall

effect are scalars, there are infinitely many treatment paths that correspond to these bounds on the overall treatment effect. To visualize the bounds, we use $(U, L)^{1-\alpha} = \{U_h^{1-\alpha}, L_h^{1-\alpha}\}_{h=1}^{H}$, where $U_h^{1-\alpha} = \frac{ub^{1-\alpha}}{H}$ and $L_h^{1-\alpha} = \frac{lb^{1-\alpha}}{H}$.[5] We choose this visualization as the interval $(U, L)^{1-\alpha}$ is a $(1-\alpha)\%$ Wald confidence interval for the average effect of the policy over the horizon H, $\frac{1}{H}\sum_{h=1}^{H}\beta_h$.

The following trivial proposition clarifies how coverage of these cumulative plausible bounds relates to coverage of the treatment path,

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

$$\mathbb{P}\left(\frac{\sum_{h=1}^{H} L_h^{1-\alpha}}{H} < \frac{\sum_{h=1}^{H}\beta_h}{H} < \frac{\sum_{h=1}^{H} U_h^{1-\alpha}}{H}\right) = (1-\alpha). \tag{2}$$

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

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

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

Since a Wald test for the cumulative effect has correct size, the result follows immediately. $\square$

Proposition 1 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.

---

[5]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.

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 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 or ad hoc 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 instead of implicitly using the data to select a restricted model, which leads to invalid inference.

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 model β_M is close to, but potentially simpler, than β. We note that the surrogate model 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 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.[6] 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]). To avoid this problem, 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 $\beta_{M(\hat{\beta})}$. 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, 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 specifically consider a model universe motivated from 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 estimated horizon.

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

In practice, we introduce 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 $\hat{\beta}$ as input, we define the following object:

$$
\begin{aligned}
\tilde{\beta}(\lambda_1, \lambda_2, K) &= \arg\min_b \; Q(b, \lambda_1, \lambda_2, K) \\
&= \arg\min_b \; \underbrace{(\hat{\beta} - b)' V^{-1} (\hat{\beta} - b)}_{\text{distance from } \hat{\beta}} + \lambda_1 \underbrace{b' D_1' W_1(K) D_1 b}_{\substack{\text{penalty on first difference} \\ \text{after horizon K}}} + \lambda_2 \underbrace{b' D_3' W_3 D_3 b}_{\substack{\text{penalty on} \\ \text{third difference}}}, \quad (3)
\end{aligned}
$$

---

[6]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."

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 B for further details.)

Solving (3) provides a closed form solution[7] for $\tilde{\beta}(\lambda_1, \lambda_2, K) := \tilde{\beta}(M)$ given by

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

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

$$\tilde{\beta}(M) - P(M)\beta \sim N(0, V_M), \tag{4}$$

where $V_M = P(M) V_\beta P(M)'$. Here, $P(M)\beta = \beta_M$ can be thought of as a particular surrogate model of β. Intuitively, $P(M)\beta$ corresponds to a "projection" of the true treatment path β into a lower dimensional space. Given (4), it would be straightforward to construct a confidence region for $\{\beta_{M,h}\}_{h=1}^{H}$, where $\beta_{M,h}$ denotes the hth 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.

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" $\hat{\beta} - \tilde{\beta}(M) =$

---

[7]Rather than use (3), one could alternatively consider a constrained optimization with tuning parameters c_1, c_2, and K, explicitly bounding the first and third difference, by using

$$\hat{\beta}(c_1, c_2, K) = \arg\min_b \underbrace{(\hat{\beta} - b)' V^{-1}(\hat{\beta} - b)}_{\text{distance from } \hat{\beta}}$$

$$\text{such that} \quad \underbrace{b' D_1' W_1(K) D_1 b \leq c_1}_{\substack{\text{small first difference,} \\ \text{after horizon K}}} \quad \text{and} \quad \underbrace{b' D_3' W_2 D_3 b \leq c_2}_{\text{small third difference}}.$$

However, this formulation does not have a closed form solution, making it less appealing to us as a practical matter in the present setting.

12

$\hat{\beta} - P(M)\hat{\beta} = (I - P(M))\hat{\beta}$. We use this residual formulation to define an analog of model degrees of freedom given by $\mathrm{df}(M) = \mathrm{trace}(P(M))$. We then select a model that minimizes a BIC analog over $\mathcal{M}$ where $\mathcal{M}$ denotes the universe of values for $M = (\lambda_1, \lambda_2, K)$:

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

We tie the researcher's hands by pre-specifying $\mathcal{M}$, the universe of models considered. In our implementation, $\mathcal{M}$ includes both a constant, linear, quadratic and cubic treatment effect model (with one, two, three, and four degrees of freedom respectively), as well as the unrestricted estimates $\hat{\beta}$ (with H degrees of freedom). $\mathcal{M}$ further includes surrogates 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 B though 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. We believe these features make the specific structure and estimation approach we employ a good default.*

Given the selected surrogate $\hat{M}$, we define the *restricted estimates* as $\tilde{\beta}(\hat{M})$. However, we cannot directly apply (4) to obtain a valid confidence region for the population value of the surrogate effect path $\beta_{\hat{M}}$ 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 $(\ell_h(X), u_h(X)) = (\tilde{\beta}(\hat{M})_h - C^\alpha V_{\hat{M}}^{1/2}(h,h), \tilde{\beta}(\hat{M})_h + C^\alpha V_{\hat{M}}^{1/2}(h,h))$ 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 hth 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 hth 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.[8]

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 $\beta_{\hat{M}}$:*

$$\mathbb{P}[\beta_{\hat{M}} \in CR^{POSI}] \geq 1 - \alpha. \tag{5}$$

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

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

$\square$

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 $\mathbb{P}(\beta_{\hat{M}} = \beta_M = \beta) = 1$. That is, in cases where model selection is effectively non-random and the selected model 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 $\mathbb{P}(\beta_{\hat{M}} = \beta_M = \beta) = 1$ under a sequence of models where $\sigma^2 \to 0$ and surrogate treatment 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 question the utility of this viewpoint from providing a useful finite sample approximation and because we view the surrogate as an economically interesting summary of the treatment path that incorporates economically motivated shrinkage with an aim of conveying key features that can be picked out relative to noise captured by V_β.*

---

[8]An implication is that one could select the model using methods other than minimizing our BIC analog. We found that minimizing this BIC analog performed well across our simulations and believe it provides a good default.

Remark 3. *Throughout we work with the unrestricted estimates $\hat{\beta}$. Our motivation is to augment visualizations of event-study paths within the setting where $\hat{\beta} \sim N(\beta, V_\beta)$ provides a reasonable approximation. We further note that the interpretation of event-study plots as currently displayed with point estimates and pointwise confidence intervals essentially relies on exactly this approximation. 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 such as difference-in-differences with relatively few treatment or control observations 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 after 17 periods (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 Online Appendix Table 1.[9] 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 $\hat{\beta} \sim N(\beta, V_\beta)$.[10]

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 Online Appendix C for more detail). The largest

---

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

[10]In the figures that follow, V_β is diagonal with its entries specified in Online Appendix C. We repeat our exercise with more general covariance matrices in Online Appendix E.
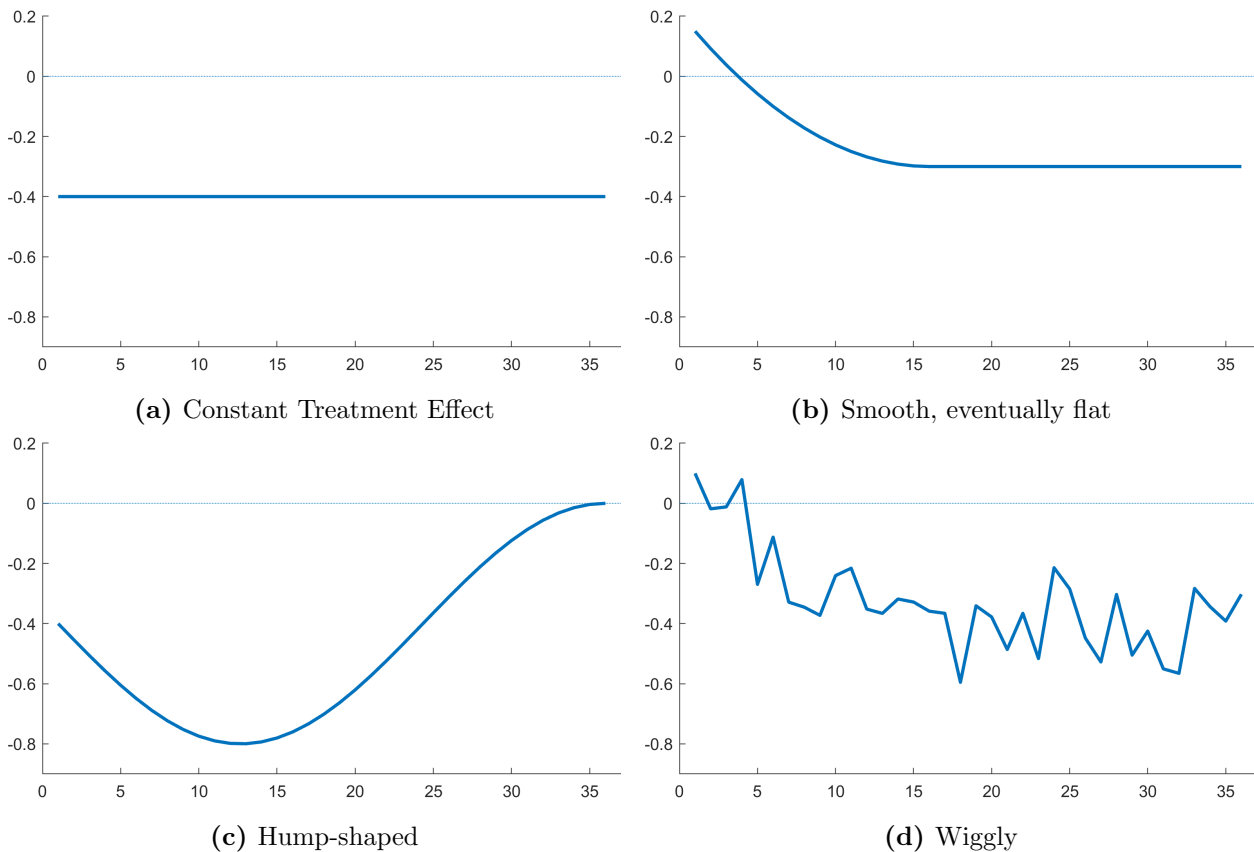
(a) Constant Treatment Effect

(b) Smooth, eventually flat

(c) Hump-shaped

(d) Wiggly

Figure 3: Four exemplary treatment effect paths β

value of σ^2 in Figure 4 (corresponding to the left most point) thus represents relatively noisy estimates.[11] 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 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.

---

[11]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, Online Appendix Figure 2 also illustrates a single realization of the right most point in Figure 4.
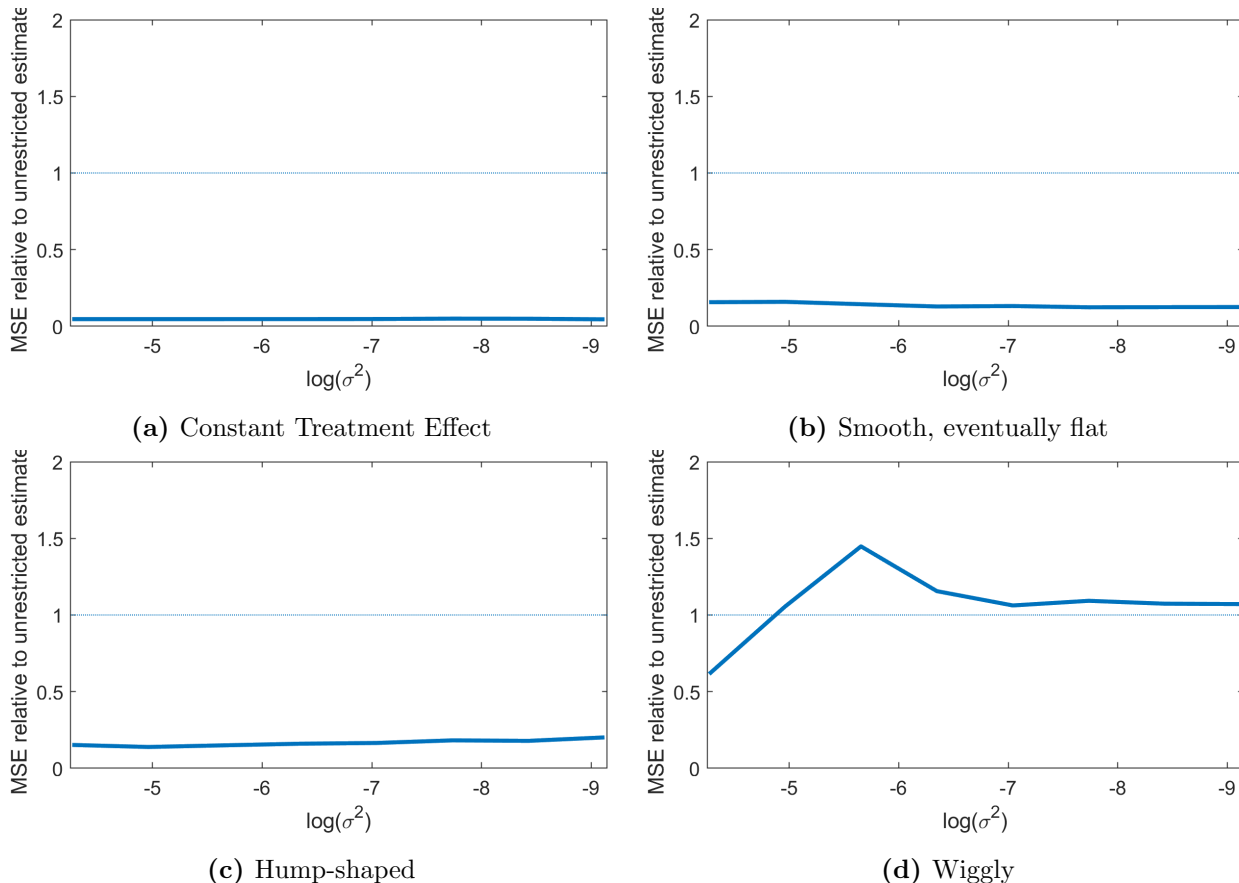
16

(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 $\hat{\beta}$.

Figure 3d).

We next turn our attention to inference and first look at coverage. We again consider the four previous DGPs and vary the amount of noise in the estimation of β by varying σ^2. The result is depicted in Figure 5, where we set $\alpha = 0.05$. 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.

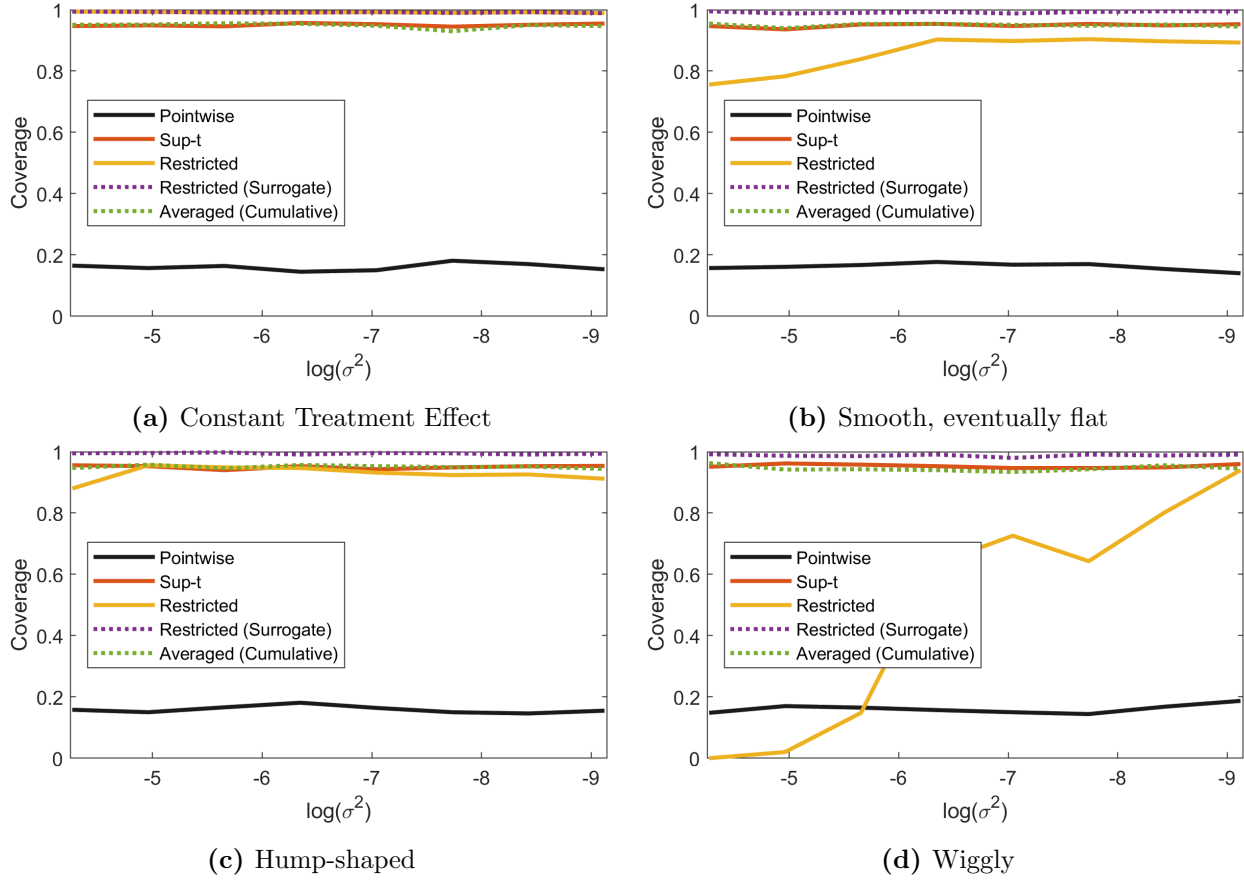Across all four DGPs, the pointwise confidence intervals only cover the true treatment

Figure 5: Coverage properties of various confidence regions as a function of the amount of noise σ^2 in the initial estimates $\hat{\beta}$.

path in 15-20% of simulations. On the other hand, the sup-t intervals achieve nominal coverage across 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 Figure 5, 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 often be the 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 to cover the true path in a higher fraction of simulations, the sup-t bands are generally wider than the pointwise bands (cf. Figure 1). We illustrate this across our simulations in Figure 6, which depicts the average width of the sup-t bands and the restricted bounds, relative to the pointwise intervals. While the sup-t bands are wider than the pointwise bands across the parameter space, the restricted bounds are in fact narrower throughout much of the parameter space, despite their improved coverage properties. For example, in the "Constant Treatment Effect" scenario, the restricted bounds are less than half as wide as the pointwise bands and less than a quarter as wide as the sup-t bands for all noise levels σ^2. Yet, they cover close to 95% of all realizations, while the pointwise confidence intervals only cover the true treatment path in 15-20% of simulations.[12] 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 infers that the wiggles in the treatment path are large enough relative to the noise level that heavy smoothing will incur a large cost in terms of losing fit. With relatively little smoothing, our restricted plausible bounds are then slightly more conservative than the standard sup-t bands, which follows immediately from the results in Berk et al. [2013].

Finally, we shed some more light on the model selection of our procedure in Figure 7.

---

[12]As we show in Online Appendix Figure 4, we select surrogate models β_M with significantly reduced degrees of freedom, relative to the unrestricted estimates, in the first three DPGs. 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 6.
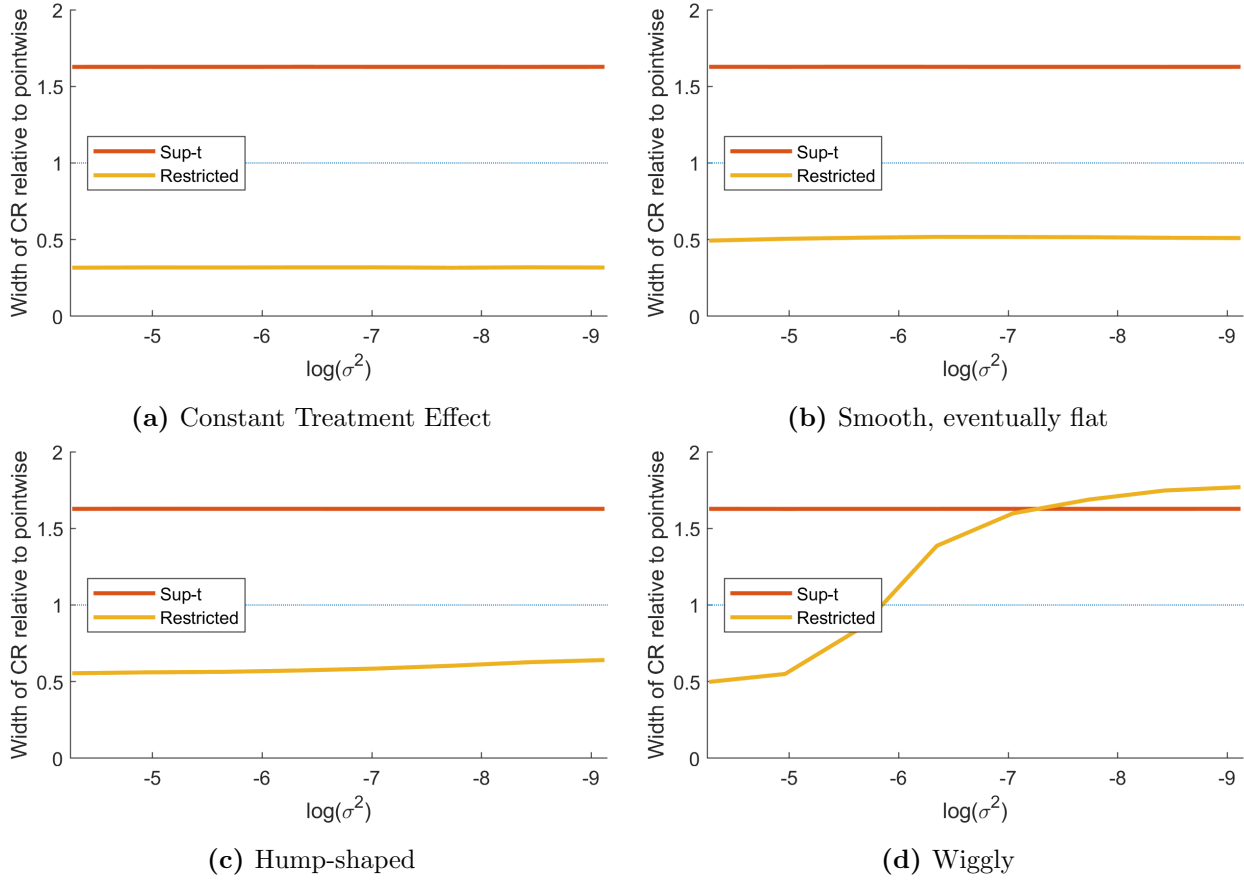
Figure 6: Average width of confidence regions relative to pointwise confidence intervals as a function of the amount of noise σ^2 in the initial estimates $\hat{\beta}$.

Each panel depicts the 1,000 selected surrogates across our simulations in light brown. Since the chosen surrogates will depend on the precision of the initial estimates $\hat{\beta}$ we fix σ^2 at 0.014, corresponding to the amount of noise in the examples in Figure 2. When the true treatment path is smooth (Panels 7a-7c), the chosen surrogate tends to closely approximate the truth. In contrast, the chosen surrogate is often simpler than β while retaining its main features when the true treatment path is not smooth (Panel 7d).[13]

---

[13]In Online Appendix Figure 5, we further illustrate how the chosen surrogates depend on σ^2 for our Wiggly DGP.

(a) Constant Treatment Effect

(b) Smooth, eventually flat
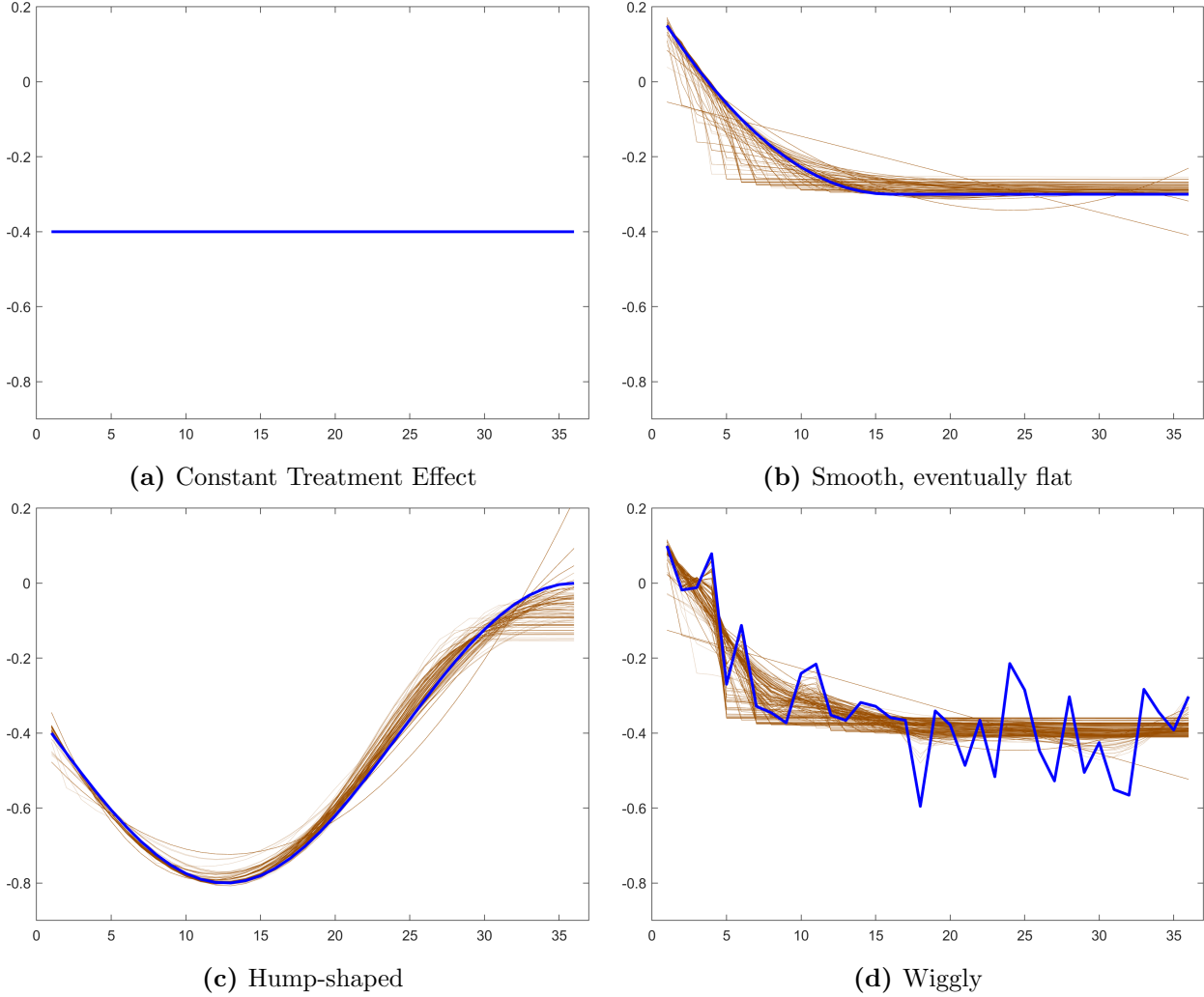
(c) Hump-shaped

(d) Wiggly

Figure 7: Illustration of the 1,000 chosen surrogates for $\sigma^2 = 0.014$ ($\log(\sigma^2) = -4.27$), corresponding to the left most point in Figures 4-6, and thus a large amount of noise in $\hat{\beta}$.

5 Conclusion

We are interested in (joint inference on) the dynamic treatment effect of a policy. We propose two visualization devices, which we term restricted and cumulative plausible bounds, to include in standard visualizations of estimated effect paths. Because both bounds explicitly target objects other than uniformly covering the entire treatment path, they can be substantially tighter than standard pointwise and uniform confidence bands. Our bounds may thus

provide useful insights about features of treatment effect paths when traditional confidence bands appear uninformative. As an auxiliary benefit, producing our restricted plausible bounds also provides additional restricted estimates that capture simplified representations of the dynamic effect path. Unsurprisingly, these restricted estimates have improved point estimation properties relative to the unrestricted estimates in settings where the treatment effect path is smooth.

It would be intersting to explore the notion of explicitly covering data-dependent surrogates in other settings. While alternative model universes or model selection rules may be needed outside of the present setting, our proposal can easily be adapted to do so. For example, it would be interesting to explore related ideas in more structural settings.

References

Shirley Almon. The distributed lag between capital appropriations and expenditures. *Econometrica*, pages 178–196, 1965.

Timothy B Armstrong, Michal Kolesár, and Mikkel Plagborg-Møller. Robust empirical bayes confidence intervals. *Econometrica*, 90(6):2567–2602, 2022.

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*, pages 802–837, 2013.

Levi Boxell, Matthew Gentzkow, and Jesse M. Shapiro. Cross-country trends in affective polarization. *The Review of Economics and Statistics*, 106(2):557–565, 03 2024.

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

Xiaohong Chen, Timothy Christensen, and Sid Kankanala. Adaptive estimation and uniform confidence bands for nonparametric structural functions and elasticities. *The Review of Economic Studies*, forthcoming.

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*, pages 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.

Arun K Kuchibhotla, John E Kolassa, and Todd A Kuffner. Post-selection inference. *Annual Review of Statistics and Its Application*, 9(1):505–527, 2022.

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*, page 105722, 2024.

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

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*, pages 775–788, 1973.

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

Jonathan Taylor and Robert J Tibshirani. Statistical learning and selective inference. *Proceedings of the National Academy of Sciences*, 112(25):7629–7634, 2015.