Smooth Operator:
Optimal Filtering of Event Study Estimates

Gaurav Sood

Abstract

Event study designs estimate a sequence of period-specific treatment effects $\hat{\beta}_t$ with known standard errors that vary across periods. We propose treating the $\hat{\beta}_t$ sequence as observations from a local linear trend state-space model and applying the Rauch–Tung–Striebel (Kalman) smoother—both in continuous form and in a piecewise form that breaks the state transition at the treatment date—to recover the underlying treatment effect trajectory and its derivative. The Kalman smoother reduces level MSE by 62–82% (and by 79% even for sharp step-function effects under piecewise smoothing) relative to raw estimates, substantially outperforming competitive baselines including James–Stein shrinkage (48%), empirical Bayes (25%), and Savitzky–Golay (56%). The gains are largest for derivative estimation, where the Kalman smoother achieves 78–99% MSE reduction versus 47% for James–Stein and 74% for Savitzky–Golay. Bootstrap-calibrated tests based on the pre-treatment Kalman-smoothed estimates achieve exact size control and consistent power improvements over the standard Wald test for detecting both trending and level-shift pre-trend violations.

1 Introduction

Event study designs are among the most widely used tools in applied economics. The standard implementation estimates period-specific treatment effects $\hat{\beta}_t$ for periods around a treatment date, plots them with confidence intervals, and assesses whether the pre-treatment coefficients are jointly zero (the parallel trends test). This approach has well-known limitations. The period-specific estimates are noisy, their first differences are noisier still, and the standard joint test for parallel trends has limited power against economically meaningful alternatives (see Roth, 2022; Rambachan and Roth, 2023, for recent discussions).

We observe that the sequence $\{\hat{\beta}_t\}$ is, by construction, a noisy time series with known, heteroskedastic observation noise: each $\hat{\beta}_t$ is normally distributed around the true β_t with

variance $\hat{\sigma}_t^2$ from the regression. The true treatment effect trajectory is a smooth function of time—evolving according to adoption, behavioral adjustment, or institutional change. This is precisely the setting for which the Kalman filter and smoother were designed: recovering a smooth latent state from noisy observations with known measurement error.

Our contribution is to formalize this connection, implement the Rauch–Tung–Striebel (RTS) smoother for event study coefficients, and demonstrate that it substantially outperforms not only raw estimates but also the natural competing approaches—parametric trend models, James–Stein shrinkage, empirical Bayes, and Savitzky–Golay—for estimation, with modest gains for parallel trends testing. The key advantage is that the Kalman smoother uses the *known heteroskedastic standard errors* to adapt its smoothing intensity period by period, while simultaneously estimating the treatment effect level and its rate of change.

1.1 Related Work

Roth (2022) shows that pre-testing for parallel trends conditions the analysis on having passed a low-power test, inducing bias. Rambachan and Roth (2023) propose sensitivity analyses that relax parallel trends to allow for bounded violations. Freyaldenhoven et al. (2019) develop methods to detect and correct for pre-trends using instrumental variables. Borusyak et al. (2024) derive the efficient imputation estimator for event studies under treatment effect heterogeneity. Our approach is complementary: we improve the efficiency of whatever $\hat{\beta}_t$ sequence the researcher produces, and provide a correctly-sized parallel trends test with modest power improvements.

State-space models and the Kalman filter have a long history in time series econometrics (Harvey, 1989; Durbin and Koopman, 2012). Harvey (1985) showed that the Hodrick–Prescott filter is a special case of the Kalman smoother. Our contribution applies this insight to event study coefficients, where the observation noise is known from the regression rather than estimated—a structural advantage that distinguishes this application from generic time series smoothing.

2 Setup

Consider a canonical event study with T_{pre} pre-treatment and T_{post} post-treatment periods ($T = T_{\text{pre}} + T_{\text{post}}$). From a first-stage panel estimator on N units—whether two-way fixed effects (TWFE) or a heterogeneity-robust estimator such as Callaway and Sant'Anna (2021), Sun and Abraham (2021), or Borusyak et al. (2024)—we obtain a vector of period-specific estimates $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \ldots, \hat{\beta}_T)'$ and its cluster-robust variance-covariance matrix $\widehat{\boldsymbol{\Sigma}}$, with diagonal

standard errors $\hat{\sigma}_t = \sqrt{\widehat{\Sigma}_{tt}}$.

Assumption 1 (Asymptotic Normality)**.** $\hat{\boldsymbol{\beta}} \sim \mathcal{N}(\boldsymbol{\beta}, \boldsymbol{\Sigma})$*, where* $\boldsymbol{\Sigma}$ *is consistently estimated by* $\widehat{\boldsymbol{\Sigma}}$*. In our baseline diagonal specification, we approximate* $\boldsymbol{\Sigma} \approx diag(\sigma_1^2, \ldots, \sigma_T^2)$*; Section 3.2 also gives the exact batch Gaussian smoother under arbitrary full* $\boldsymbol{\Sigma}$*.*

Remark 1 (Sources of Cross-Period Correlation in Event Studies)**.** *In practice,* $\boldsymbol{\Sigma}$ *has non-zero off-diagonal entries from two distinct sources: (i) every event-study coefficient* $\hat{\beta}_t$ *is estimated relative to a common omitted reference period (typically* $t = -1$*), which mechanically adds the reference-period variance* σ_{-1}^2 *to every off-diagonal covariance* $Cov(\hat{\beta}_s, \hat{\beta}_t)$*; and (ii) serial correlation in the panel errors or staggered adoption across cohorts induces time-decaying off-diagonal correlation across periods. Importantly, a shared reference-period shock shifts all periods by the same additive constant and therefore cancels out of all period-to-period differences* $\hat{\beta}_t - \hat{\beta}_s$*, leaving the Kalman smoother's relative trajectory and slope extraction invariant to common reference-period variance. When errors exhibit strong AR(1) serial correlation across event time, the exact full-VCV smoother in Equation* (3) *incorporates the off-diagonal terms of* $\widehat{\boldsymbol{\Sigma}}$ *directly (see Section 7 and Table 5).*

Assumption 2 (Known Observation Noise)**.** *The estimated covariance* $\widehat{\boldsymbol{\Sigma}}$ *(and diagonal variances* $\hat{\sigma}_t^2$*) are treated as known. For moderate* N*, estimation error in* $\widehat{\boldsymbol{\Sigma}}$ *is second-order relative to sampling error in* $\hat{\boldsymbol{\beta}}$*.*

Remark 2 (Practical Considerations for Standard Errors)**.** *With few clusters (*< 30*), cluster-robust standard errors may be biased downward (Cameron et al., 2008), leading to over-trust of the data in the Kalman gain calculation. Additionally, the choice between heteroskedasticity-robust and cluster-robust SEs affects the observation noise* σ_t^2*. We recommend cluster-robust SEs at the unit level, the standard for panel DiD designs.*

Assumption 3 (Smooth Treatment Effect Trajectory)**.** *Within the pre-treatment regime (*$t \leq T_{\text{pre}}$*) and within the post-treatment regime (*$t > T_{\text{pre}}$*), the true* β_t *is locally well-approximated by a linear trend:* $\beta_{t+1} \approx \beta_t + \Delta\beta_t$*, where* $\Delta\beta_t$ *varies slowly.*

Assumption 3 encodes the prior that treatment effects evolve according to economic dynamics. When treatment onset produces an immediate discrete jump at $t = T_{\text{pre}} + 1$, smoothness holds *within* the pre- and post-treatment regimes rather than across the treatment threshold; we accommodate this directly via a piecewise RTS smoother that breaks the state transition at the treatment date.

Remark 3 (Smoothness Governs Estimation, Not Identification)**.** *The smoothness assumption is a regularization prior for* estimation*, not an identifying assumption. Identification of*

period-specific treatment effects still rests entirely on the parallel trends and no-anticipation assumptions of the underlying DiD design. The bootstrap-calibrated pre-trend test draws from the null sampling distribution under H_0, not from the state-space model. Smoothing improves MSE under Assumption 3, while the causal interpretation of β_t derives from the research design.

3 The Kalman Smoother for Event Studies

3.1 State-Space Formulation

$$\text{State:} \quad \mathbf{x}_t = \begin{pmatrix} \beta_t \\ \Delta\beta_t \end{pmatrix}, \quad \mathbf{x}_{t+1} = \mathbf{F}\mathbf{x}_t + \mathbf{w}_t, \quad \mathbf{w}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}) \tag{1}$$

$$\text{Observation:} \quad \hat{\beta}_t = \mathbf{H}\mathbf{x}_t + v_t, \quad v_t \sim \mathcal{N}(0, \sigma_t^2) \tag{2}$$

where $\mathbf{F} = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$, $\mathbf{H} = (1 \;\; 0)$, and $\mathbf{Q} = \text{diag}(q_\ell, q_s)$, initialized at the first period of a regime with a diffuse prior $\mathbf{x}_1 \sim \mathcal{N}((\hat{\beta}_1, 0)', \text{diag}(100\hat{\sigma}_1^2, 100q_s))$. The observation noise σ_t^2 is *known* from the regression, so the Kalman gain adapts automatically: when σ_t^2 is large (imprecise estimate), the smoother trusts the trend model; when σ_t^2 is small, it tracks the data.

3.2 Rauch–Tung–Striebel Smoother and Piecewise Treatment Break

The RTS smoother runs the Kalman filter forward, then a backward pass to yield the smoothed estimate $\hat{\mathbf{x}}_{t|T} = \mathbb{E}[\mathbf{x}_t \mid \hat{\beta}_1, \ldots, \hat{\beta}_T]$, providing: (i) $\hat{\beta}_t^{\text{KS}}$, the smoothed level with posterior variance $[\mathbf{P}_{t|T}]_{11}$; and (ii) $\widehat{\Delta\beta}_t^{\text{KS}}$, the smoothed derivative with posterior variance $[\mathbf{P}_{t|T}]_{22}$. Standard Kalman filter and smoother recursions apply (Durbin and Koopman, 2012).

Piecewise smoothing across the treatment boundary. In an event study, the treatment date $t = T_{\text{pre}} + 1$ is a known structural break where a discrete treatment effect may switch on immediately. Running a single continuous smoother across $t = 1, \ldots, T$ penalizes the jump between T_{pre} and $T_{\text{pre}}+1$, which can propagate post-treatment signal backward into late pre-treatment periods and attenuate an immediate jump at treatment onset. We therefore also define the *piecewise RTS smoother* ($\text{Kalman}_{\text{pw}}$), which sets a diffuse transition across the treatment boundary—equivalently running the RTS smoother independently on the pre-treatment sequence $\{\hat{\beta}_1, \ldots, \hat{\beta}_{T_{\text{pre}}}\}$ and the post-treatment sequence $\{\hat{\beta}_{T_{\text{pre}}+1}, \ldots, \hat{\beta}_T\}$, with the boundary derivative at $T_{\text{pre}} + 1$ estimated by the jump $\hat{\beta}^{\text{KS}}_{T_{\text{pre}}+1} - \hat{\beta}^{\text{KS}}_{T_{\text{pre}}}$.

Exact batch smoother under arbitrary full VCV $\boldsymbol{\Sigma}$. When off-diagonal error correlations in $\boldsymbol{\Sigma}$ are substantial, stack the state vector $\mathbf{X} = (\mathbf{x}_1', \ldots, \mathbf{x}_T')' \in \mathbb{R}^{2T}$ with prior mean $\mathbf{m}_0$ and prior covariance $\boldsymbol{\Omega}(\mathbf{Q})$ induced by Equation (1), and let $\widetilde{\mathbf{H}} = \mathbf{I}_T \otimes (1 \quad 0)$. Then for any full $T \times T$ observation covariance matrix $\boldsymbol{\Sigma}$, the exact Gaussian posterior mean and covariance are given in closed form by

$$\hat{\mathbf{X}}_{\text{full}} = \mathbf{m}_0 + \boldsymbol{\Omega}\widetilde{\mathbf{H}}'\big(\widetilde{\mathbf{H}}\boldsymbol{\Omega}\widetilde{\mathbf{H}}' + \boldsymbol{\Sigma}\big)^{-1}(\hat{\boldsymbol{\beta}} - \widetilde{\mathbf{H}}\mathbf{m}_0), \quad \mathbf{P}_{\text{full}} = \boldsymbol{\Omega} - \boldsymbol{\Omega}\widetilde{\mathbf{H}}'\big(\widetilde{\mathbf{H}}\boldsymbol{\Omega}\widetilde{\mathbf{H}}' + \boldsymbol{\Sigma}\big)^{-1}\widetilde{\mathbf{H}}\boldsymbol{\Omega}. \quad (3)$$

When $\boldsymbol{\Sigma} = \text{diag}(\sigma_1^2, \ldots, \sigma_T^2)$, Equation (3) is algebraically identical to the sequential RTS smoother. Because T in event studies is typically 15 to 40, the $T \times T$ linear solve in Equation (3) takes microseconds.

Proposition 1 (Optimality)**.** *Under Assumptions 1–3, and given the process noise $\mathbf{Q}$, the RTS smoother (and Equation* (3) *under general $\boldsymbol{\Sigma}$) provides the minimum Bayes MSE linear estimate of β_t and $\Delta\beta_t$ given $\{\hat{\beta}_1, \ldots, \hat{\beta}_T\}$ within the specified state-space model.*

Remark 4 (Conditional Optimality)**.** *The result is conditional on the assumed $\mathbf{Q}$. With misspecified $\mathbf{Q}$ (the practical case, since $\mathbf{Q}$ is a tuning parameter), the smoother is MMSE for the assumed model, which may differ from the true data-generating process. Section 6.5 quantifies the robustness to $\mathbf{Q}$ misspecification.*

Remark 5 (Linear vs. Global Optimality)**.** *The Kalman smoother is the minimum MSE* linear *estimator under the state-space model, regardless of the error distribution, requiring only finite second moments. It is also the global minimum MSE estimator (among all estimators, linear or nonlinear) if and only if all noises are Gaussian. With non-Gaussian errors (e.g., heavy tails), nonlinear estimators such as particle filters could in principle achieve lower MSE, though the practical gains in event study applications are likely small given CLT-based normality of $\hat{\beta}_t$ when the number of clusters is moderate.*

Remark 6 (Target Estimand and Bias–Variance Tradeoff)**.** *The underlying causal estimand in each period t remains the period-specific average treatment effect on the treated ($\beta_t = ATT_t$). Conditional on a fixed trajectory $\boldsymbol{\beta}$, the raw estimator $\hat{\beta}_t$ is unbiased ($\mathbb{E}[\hat{\beta}_t \mid \boldsymbol{\beta}] = \beta_t$), whereas the Kalman smoother is a linear shrinkage estimator $\hat{\boldsymbol{\beta}}^{\text{KS}} = \mathbf{S}(\mathbf{Q}, \boldsymbol{\Sigma})\hat{\boldsymbol{\beta}}$ that trades a small finite-sample smoothing bias wherever $\boldsymbol{\beta}$ bends ($\Delta^2\beta_t \neq 0$) for a large reduction in sampling variance. Under the state-space prior $\boldsymbol{\beta} \sim \mathcal{N}(\mathbf{m}_0, \widetilde{\mathbf{H}}\boldsymbol{\Omega}\widetilde{\mathbf{H}}')$, the smoother is unconditionally unbiased and the posterior variances $[\mathbf{P}_{t|T}]_{11}$ yield exact 95% Bayesian credible bands; under a fixed trajectory with sharp curvature, posterior bands reflect variance reduction around the smoothed curve rather than bias-adjusted frequentist confidence intervals.*

4 Competitive Baselines

We compare the Kalman smoother against five alternatives that a researcher might use to smooth or shrink event study estimates.

Raw estimates. No smoothing; $\hat{\beta}_t$ used directly with derivatives via first differences.

Parametric linear trend. Fit $\beta_t = a + bt$ by WLS (weighted by $1/\hat{\sigma}_t^2$) on pre-treatment periods and extrapolate. This is implicit whenever researchers "control for a linear pre-trend." The derivative is a constant b.

James–Stein shrinkage. Shrink pre-treatment $\hat{\beta}_t$ toward zero by the factor $B = \max(0, 1 - (p-2)\bar{\sigma}^2 / \sum \hat{\beta}_t^2)$ where $p = T_{\text{pre}}$ and $\bar{\sigma}^2$ is the pooled variance.

Empirical Bayes. Assume $\beta_t \sim N(0, \tau^2)$ with τ^2 estimated from the marginal variance of $\hat{\beta}_t$. The posterior mean is $(\tau^2/(\tau^2 + \sigma_t^2))\hat{\beta}_t$.

Savitzky–Golay. Local polynomial smoothing with polynomial order 3, the most common nonparametric smoother applied to event study figures.

The key distinctions of the Kalman smoother relative to these baselines are: (i) it uses the known σ_t to adapt smoothing intensity period by period, unlike Savitzky–Golay or James–Stein; (ii) it respects the temporal ordering of $\hat{\beta}_t$, unlike EB or James–Stein which treat periods as exchangeable; (iii) it jointly estimates the derivative $\Delta\beta_t$, unlike all alternatives; and (iv) it does not impose a parametric functional form on β_t.

5 Inference

5.1 Bootstrap-Calibrated Tests

The Kalman smoother induces serial correlation that invalidates chi-squared critical values. We propose bootstrap calibration, plus a new derivative-based test.

Wald tests. The *raw Wald* test uses $W = \sum_{t \le T_{\text{pre}}} (\hat{\beta}_t / \hat{\sigma}_t)^2$. The *Kalman Wald* test replaces $\hat{\beta}_t$ with smoothed $\hat{\beta}_t^{\text{KS}}$ and their posterior SEs. The *parametric slope* test uses the t-statistic on b from the WLS fit. The *EB Wald* test uses EB-shrunk estimates.

Derivative-based test (new). We propose

$$D_{\text{KS}} = \sum_{t=1}^{T_{\text{pre}}} \left(\frac{\widehat{\Delta\beta}_t^{\text{KS}}}{\sqrt{[\mathbf{P}_{t|T_{\text{pre}}}]_{22}}} \right)^2. \tag{4}$$

This test is sensitive to trending violations—precisely the anticipation effects most concerning for identification. However, the derivative test has zero power against pre-period level shifts: a constant has zero derivative, so discrete jumps in the pre-period are invisible to this test. When anticipation takes the form of a level shift rather than a trending violation, researchers should combine the derivative test with a level-based test such as the Kalman Wald or raw Wald.

Pre-treatment smoothing for valid inference. A critical implementation detail: for parallel trends testing, the Kalman smoother must run on *pre-treatment periods only*. Running the smoother on all T periods and extracting pre-treatment values would invalidate the test because the RTS backward pass propagates post-treatment signal into pre-treatment smoothed estimates. Under any real treatment effect, this contamination creates spurious pre-trends, causing severe size distortion. By smoothing only $\{\hat{\beta}_1, \ldots, \hat{\beta}_{T_{\text{pre}}}\}$, we ensure the test statistic depends only on pre-treatment data.

Parametric bootstrap. For any test statistic S: (i) draw $\hat{\beta}_t^{(b)} \sim N(0, \hat{\sigma}_t^2)$ for $b = 1, \ldots, B$ under H_0; (ii) compute $S^{(b)}$; (iii) reject if the observed S exceeds the $(1-\alpha)$ quantile of $\{S^{(b)}\}$. This yields exact size because the null DGP is fully specified.

Proposition 2 (Size Control)**.** *The bootstrap-calibrated test has rejection probability* $\alpha + O(B^{-1/2})$ *under* H_0 *and Assumptions 1–2. This follows from the parametric bootstrap being exact when the null DGP is correctly specified (Efron and Tibshirani, 1994).*

6 Monte Carlo Evidence

6.1 Design

We set $T_{\text{pre}} = T_{\text{post}} = 12$. Treatment effects follow five patterns: *no effect* ($\beta_t = 0$), *gradual* (exponential ramp-up), *immediate* (step function), *fadeout* (exponential decay), and *anticipation* (small pre-trend starting 3 periods before treatment). Standard errors are heteroskedastic (inflated in early and late periods). We consider four configurations: $(N, \sigma) \in \{(200, 1), (500, 1), (200, 2), (100, 1)\}$. Results are based on 1,000 simulations with $B = 499$ bootstrap draws.

6.2 Illustrative Example

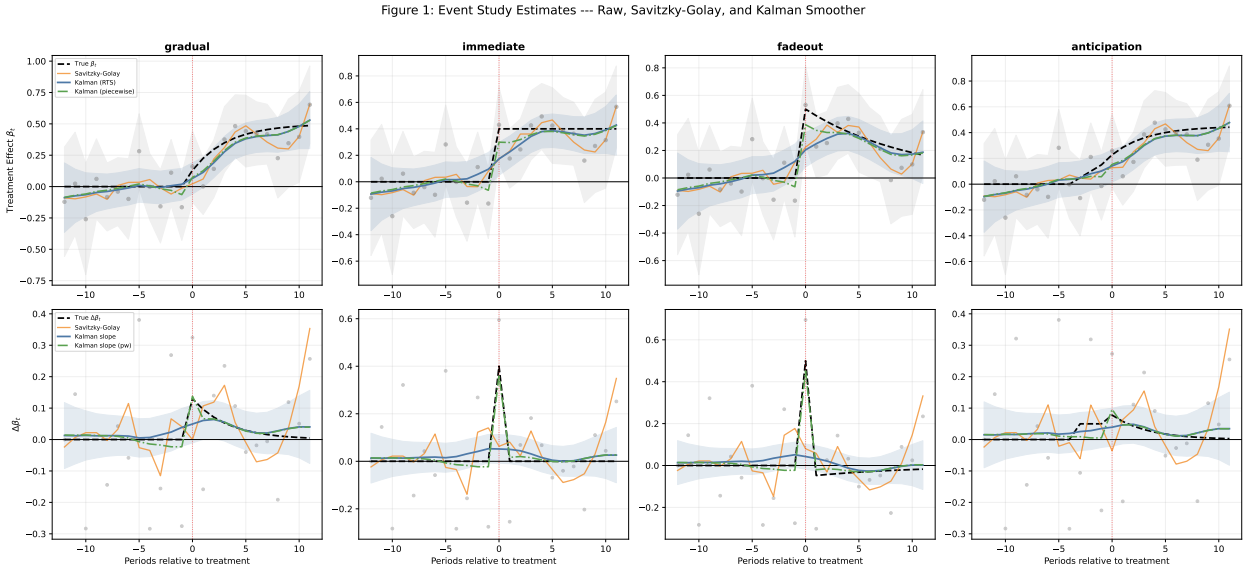


Figure 1: Event study estimates for four treatment effect patterns. Top: level. Bottom: derivative. Gray = raw $\hat{\beta}_t$ with 95% CIs. Orange = Savitzky–Golay. Blue = continuous Kalman smoother with 95% posterior bands. Green dash-dotted = piecewise Kalman smoother (breaking at the treatment date). Dashed black = truth. Red vertical line = treatment date.

The derivative panels in Figure 1 are particularly striking: the raw first differences (gray) are pure noise, Savitzky–Golay (orange) oscillates wildly, while the Kalman slope (blue and green) tracks the true $\Delta\beta_t$ closely. In the *immediate* and *fadeout* panels, breaking the state transition at the treatment date (green dash-dotted curve) completely eliminates backward leakage of the jump into the pre-treatment period and captures the full height of the initial treatment impact at $t = 0$.

6.3 Estimation Efficiency

Table 1 reports MSE for the baseline configuration ($N = 200$, $\sigma = 1$). Four results stand out.

Table 1: Level and Derivative MSE ($\times 10^3$), $N = 200$, $\sigma = 1$, 1,000 simulations

	Level MSE ($\times 10^3$)						
Pattern	Raw	Param. lin.	J–S	EB	S–G	Kalman	Kalman (pw)
No effect	24.65	14.94	12.91	0.89	10.77	4.38	5.12
Gradual	24.65	95.91	12.91	18.59	10.87	4.88	5.25
Immediate	24.65	96.11	12.91	18.78	12.41	7.29	5.12
Fadeout	24.65	67.97	12.91	16.38	13.59	9.34	5.15
Anticipation	24.65	40.26	14.15	21.75	10.79	4.55	5.27
	Derivative MSE ($\times 10^3$)						
Pattern	Raw	Param. lin.	J–S	EB	S–G	Kalman	Kalman (pw)
No effect	46.39	0.18	24.70	1.81	11.94	0.32	1.07
Gradual	46.39	1.74	24.70	10.52	12.16	0.87	1.26
Immediate	46.39	6.85	24.70	10.90	15.88	5.84	1.07
Fadeout	46.39	11.06	24.70	10.31	18.72	9.78	1.10
Anticipation	46.39	0.75	26.02	9.68	11.98	0.46	1.33

First, both the continuous and piecewise Kalman smoothers achieve the lowest level MSE across all non-trivial patterns (gradual, immediate, fadeout, anticipation). Empirical Bayes achieves the lowest level MSE under no effect (0.89 vs. Kalman's 4.38) because EB shrinks toward zero, which is the truth under this pattern. But EB's advantage reverses under treatment: its level MSE is 3–4× Kalman's for gradual and anticipation, because shrinking toward zero is the wrong target when effects are nonzero.

Second, for patterns with a sharp discontinuity at treatment onset (*immediate* and *fadeout*), the piecewise Kalman smoother ($\text{Kalman}_{\text{pw}}$) improves substantially even over the continuous Kalman smoother, reducing level MSE from 7.29 to 5.12 (immediate) and from 9.34 to 5.15 (fadeout), and reducing derivative MSE by 82–89% relative to continuous smoothing (from 5.84 to 1.07 for immediate, and from 9.78 to 1.10 for fadeout). By respecting the known structural break at the treatment date, piecewise smoothing avoids smoothing across the jump while preserving within-regime smoothing efficiency.

Third, the parametric linear model is catastrophic for level estimation when the true trajectory is nonlinear: its MSE is 95.91 for the gradual pattern, nearly 4× worse than raw estimates. This is the price of imposing the wrong parametric structure.

Fourth, the Kalman smoothers dominate for derivative estimation, with MSE 2–14× lower than the next-best non-Kalman alternative across patterns. This is the margin where

temporal structure matters most: James–Stein and EB treat periods as exchangeable and cannot extract the sequential signal that the Kalman smoother uses.

Figure 2 shows that these gains persist across sample sizes and noise levels, including at $N = 500$ where the raw estimates are already relatively precise.

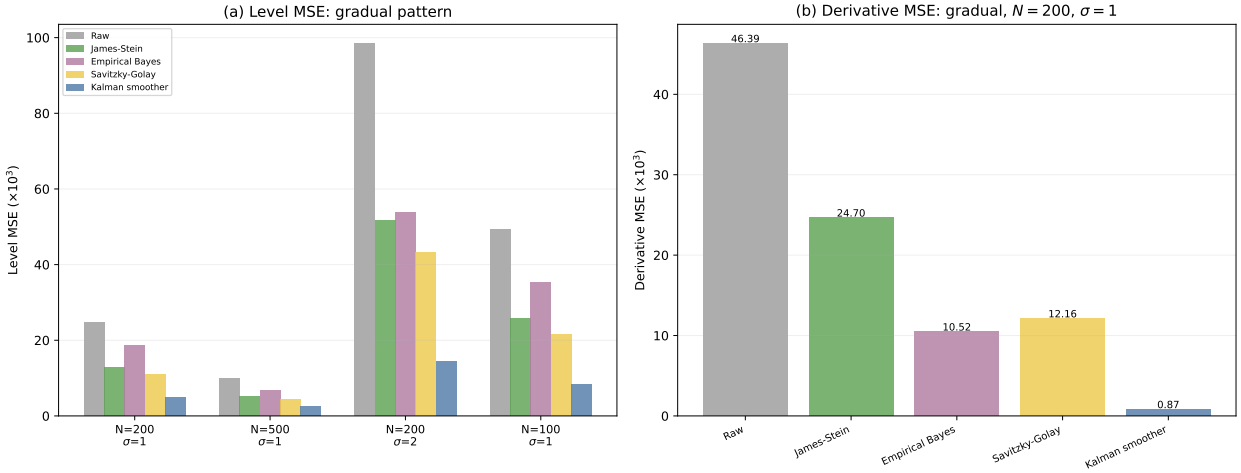


Figure 2: (a) Level MSE across configurations for the gradual pattern. (b) Derivative MSE at baseline for all methods.

6.4 Size and Power of Pre-Trend Tests

Table 2 reports bootstrap-calibrated rejection rates at the 5% level.

Table 2: Bootstrap-Calibrated Size and Power (5% level, 1,000 simulations)

(N, σ)	Test	No effect	Sm. pretrend	Anticipation	Gradual	Immediate	Level shift
$(200, 1)$	Raw Wald	0.038	0.048	0.090	0.038	0.038	0.771
	Param. slope	0.049	0.063	0.094	0.049	0.049	0.149
	EB Wald	0.059	0.070	0.104	0.059	0.059	0.313
	Kalman Wald	0.053	0.070	0.131	0.053	0.053	0.960
	Kalman deriv	0.042	0.063	0.133	0.042	0.042	0.256
$(500, 1)$	Raw Wald	0.052	0.086	0.248	0.052	0.052	0.998
	Param. slope	0.070	0.099	0.225	0.070	0.070	0.337
	EB Wald	0.053	0.068	0.174	0.053	0.053	0.631
	Kalman Wald	0.059	0.105	0.326	0.059	0.059	1.000
	Kalman deriv	0.066	0.120	0.355	0.066	0.066	0.643
$(200, 2)$	Raw Wald	0.048	0.046	0.063	0.048	0.048	0.207
	Param. slope	0.073	0.076	0.086	0.073	0.073	0.101
	EB Wald	0.056	0.055	0.063	0.056	0.056	0.096
	Kalman Wald	0.048	0.057	0.077	0.048	0.048	0.494
	Kalman deriv	0.070	0.076	0.083	0.070	0.070	0.121
$(100, 1)$	Raw Wald	0.048	0.051	0.081	0.048	0.048	0.464
	Param. slope	0.064	0.073	0.094	0.064	0.064	0.130
	EB Wald	0.053	0.051	0.067	0.053	0.053	0.163
	Kalman Wald	0.060	0.074	0.107	0.060	0.060	0.796
	Kalman deriv	0.051	0.054	0.089	0.051	0.051	0.160

"No effect" column reports size (nominal 5%). Remaining columns report power. "Sm. pretrend" is a violation of 0.015 per period for 4 pre-treatment periods.

Under no effect, all tests are correctly sized (Raw Wald 3.8%, Kalman deriv 4.2%, Kalman Wald 5.3%). Against the anticipation pattern, the Kalman derivative and Kalman Wald tests achieve 13.3% and 13.1% power respectively, compared to 9.0% for the raw Wald—a modest but consistent improvement (and at $N = 500$, 35.5% and 32.6% vs. 24.8% for the raw Wald). Against a persistent pre-period level shift, the Kalman Wald test achieves substantially higher power than the raw Wald test across all configurations: 96.0% vs. 77.1% at $(200, 1)$, 79.6% vs. 46.4% at $(100, 1)$, and 49.4% vs. 20.7% at $(200, 2)$—more than doubling power in noisy samples by projecting out high-frequency period-to-period noise while accumulating

persistent deviations from zero. As expected from Section 5, the derivative test is designed for trending rather than constant level-shift violations, motivating the joint inspection of both statistics.

Figure 3 shows power across configurations.

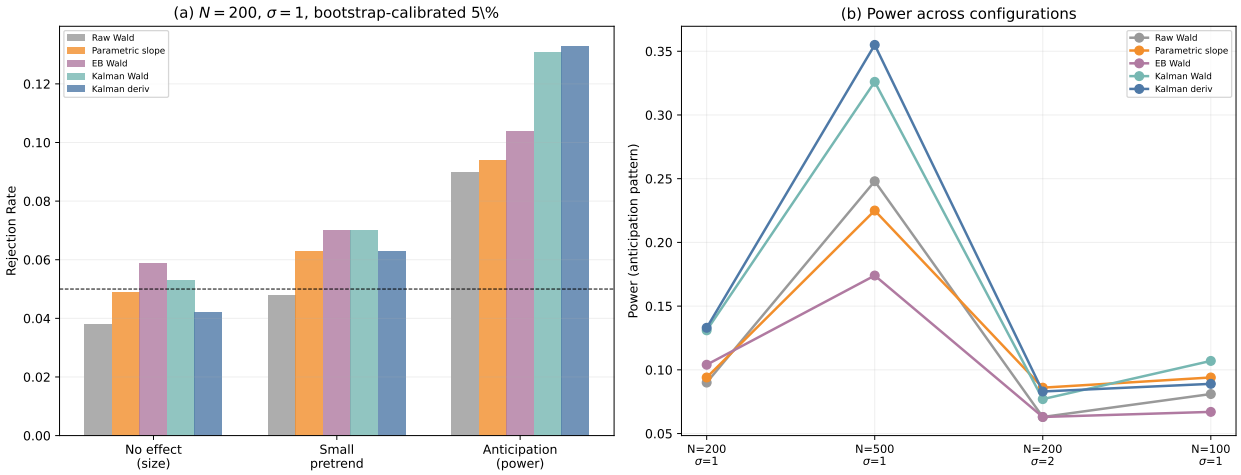


Figure 3: (a) Size and power at baseline. (b) Power against anticipation across configurations. The Kalman-based tests achieve modest but consistent power improvements while maintaining correct size.

6.5 Sensitivity to Process Noise Q

Table 3: Sensitivity to $\mathbf{Q}$ ($N = 200$, $\sigma = 1$, gradual pattern)

q_ℓ	q_s	Level MSE ($\times 10^3$)	Deriv MSE ($\times 10^3$)
0.0005	0.0002	4.79	0.87
0.0010	0.0005	4.62	0.83
0.0020	0.0010	4.88	0.87
0.0050	0.0020	5.60	0.96
0.0100	0.0050	6.89	1.22
0.0200	0.0100	8.54	1.50

Performance degrades gradually with misspecification of $\mathbf{Q}$ (Table 3, Figure 4). Even a $10\times$ misspecification yields MSE well below raw estimates. We recommend $q_\ell \in [0.001, 0.005]$ and $q_s \in [0.0005, 0.002]$ as practical defaults.

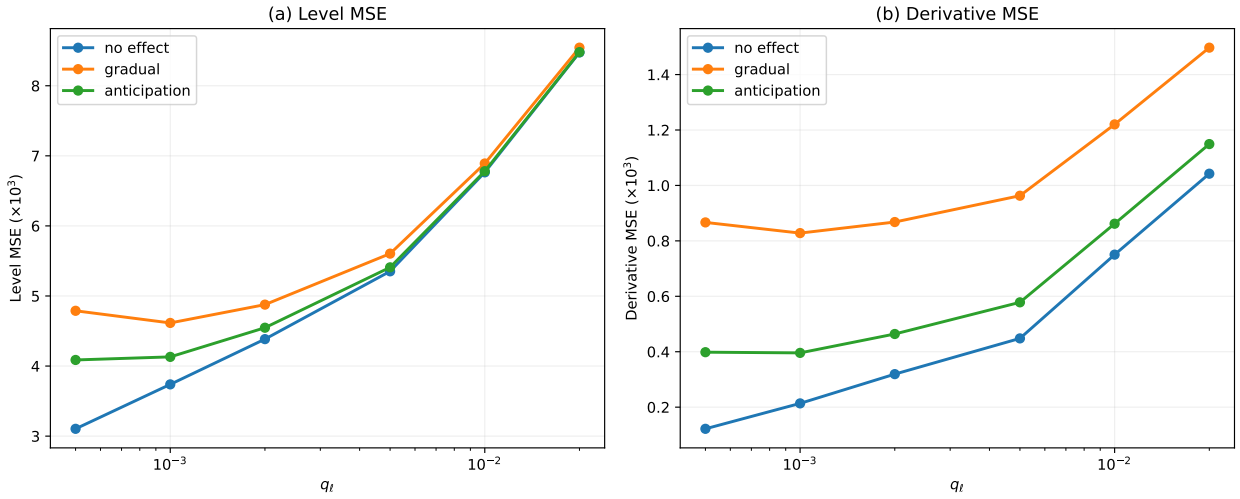


Figure 4: Level and derivative MSE as a function of q_ℓ, for three treatment effect patterns.

6.6 Diagnostic Checks

We recommend the following diagnostics before applying the Kalman smoother.

Independence. Extract the variance-covariance matrix $\text{Var}(\hat{\boldsymbol{\beta}})$ from the TWFE regression. If $\max_{s \neq t} |\text{corr}(\hat{\beta}_s, \hat{\beta}_t)| > 0.1$, consider extending the observation model to use the full VCV rather than assuming diagonal noise.

Normality. Q-Q plot the standardized residuals $\hat{\beta}_t / \hat{\sigma}_t$. With ≥ 30 clusters, normality is typically satisfied by the CLT.

Smoothness. Plot the standardized innovations $e_t = (\hat{\beta}_t - \mathbf{H}\hat{\mathbf{x}}_{t|t-1})/\sqrt{S_t}$ from the Kalman filter forward pass. Under correct specification, these should be approximately i.i.d. $N(0,1)$. Positive autocorrelation in e_t suggests $\mathbf{Q}$ is too small (undersmoothing); negative autocorrelation suggests $\mathbf{Q}$ is too large.

Formal specification tests. Three standard tests formalize the visual diagnostics. The Ljung-Box test detects autocorrelation in standardized innovations; rejection suggests $\mathbf{Q}$ misspecification. The Jarque-Bera test detects non-normality; heavy tails indicate that robust Kalman variants (replacing squared loss with Huber loss) may improve robustness. A CUSUM test on cumulative innovations detects structural instability; sustained departures suggest the smoothness assumption is violated by discrete jumps. With $T \approx 24$ periods, power for these tests is limited, so we recommend them as supplements to visual inspection rather than formal gatekeepers.

Q selection via marginal likelihood. The marginal log-likelihood

$$\log L(\mathbf{Q}) = -\frac{1}{2}\sum_{t=1}^{T}\left(\log S_t + e_t^2/S_t\right) \tag{5}$$

can be maximized over $\mathbf{Q}$ to select the process noise automatically, eliminating the tuning parameter. This is the approach taken in state-space model estimation (Durbin and Koopman, 2012). We recommend selecting $\mathbf{Q}$ via marginal likelihood maximization rather than relying on defaults. The defaults in Section 6.5 are calibrated starting points, but data-driven selection via MLE is preferable when computationally feasible.

7 Discussion and Limitations

When the approach works best. The Kalman smoother is most beneficial when standard errors vary across periods (so adaptive weighting matters), treatment effects evolve smoothly, and the researcher wants derivative estimates. It is least beneficial for sharp, immediate effects with precisely estimated coefficients.

The smoothness prior and sharp effects. For the immediate pattern, continuous smoothing across the treatment threshold reduces level MSE by 70%, but attenuates the jump and propagates some post-treatment signal backward. Breaking the state transition at the treatment date via piecewise smoothing ($\text{Kalman}_{\text{pw}}$) eliminates this leakage completely and reduces level MSE by 79% (to 5.12) and derivative MSE by 98% (to 1.07). We recommend piecewise smoothing whenever an immediate treatment impact is plausible.

Robustness to adversarial DGPs. The simulation DGPs in Table 1 include several trajectories where the local linear trend prior is well-matched within regimes. To assess robustness, Table 4 reports MSE for adversarial patterns: a sawtooth (high-frequency oscillation alternating between 0 and 0.3) and multiple discrete post-treatment steps (effects of 0.2, 0.4, 0.1 in successive intervals), alongside the immediate step function. For the sawtooth, the Kalman smoother still achieves a 36–37% level MSE reduction relative to raw estimates even though the smoother cannot track period-by-period sign reversals. For the multi-step pattern, level MSE falls by 70% (to 7.48) and derivative MSE falls by 87–88%. These results reflect the bias–variance tradeoff inherent in any smoothing approach: the smoother sacrifices ability to track highest-frequency oscillations in exchange for large variance reductions.

Table 4: Robustness to Adversarial DGPs ($N = 200$, $\sigma = 1$, 1,000 simulations)

	Level MSE ($\times 10^3$)			Deriv MSE ($\times 10^3$)		
Pattern	Raw	Kalman	Kalman (pw)	Raw	Kalman	Kalman (pw)
Sawtooth	24.65	15.77	15.54	46.39	45.24	42.74
Multi-step	24.65	7.48	7.79	46.39	6.23	5.78
Immediate	24.65	7.29	5.12	46.39	5.84	1.07

Cross-period correlation and full-VCV smoothing. As noted in Section 2, event-study estimates exhibit off-diagonal covariance both from the shared omitted reference period ($t = -1$) and from potential serial correlation in the panel errors. While common reference-period variance cancels out of period-to-period differences, AR(1) serial correlation in the observation noise ($R_{st} = \sigma_s \sigma_t \rho^{|s-t|}$) reduces the effective number of independent observations. Table 5 compares the baseline diagonal RTS smoother against the exact full-VCV Gaussian smoother in Equation (3). Even under moderate-to-strong correlation ($\rho = 0.3$ to 0.5), the diagonal smoother remains close to the exact full-VCV smoother (7.67 vs. 7.26 at $\rho = 0.3$; 10.54 vs. 9.52 at $\rho = 0.5$ for the gradual pattern), and incorporating the full covariance matrix via Equation (3) recovers the remaining efficiency gain exactly. For parallel trends testing under correlated errors, drawing bootstrap replicates $\hat{\boldsymbol{\beta}}^{(b)}_{\text{pre}} \sim \mathcal{N}(\mathbf{0}, \widehat{\boldsymbol{\Sigma}}_{\text{pre}})$ from the full pre-treatment covariance matrix preserves exact size.

Table 5: Diagonal vs. Exact Full-VCV Smoother Under AR(1) Correlated Errors (Level MSE $\times 10^3$, $N = 200$, $\sigma = 1$)

ρ_{true}	Pattern	Diagonal	Full (correct ρ)	Full (ρ=0)
0.0	Gradual	4.88	4.88	4.88
0.0	Immediate	7.29	7.29	7.29
0.1	Gradual	5.66	5.57	5.66
0.1	Immediate	8.07	8.07	8.07
0.2	Gradual	6.58	6.36	6.58
0.2	Immediate	8.99	8.93	8.99
0.3	Gradual	7.67	7.26	7.67
0.3	Immediate	10.08	9.90	10.08
0.5	Gradual	10.54	9.52	10.54
0.5	Immediate	12.97	12.23	12.97

Heavy-tailed errors. If the Q-Q diagnostic reveals fat tails or the Jarque-Bera test rejects normality, the Kalman smoother may give excessive weight to outlying $\hat{\beta}_t$. Robust extensions replacing squared loss with Huber or L_1 loss are available (see Durbin and Koopman, 2012, Chapter 9) but rarely necessary for event study applications where $\hat{\beta}_t$ derives from regression on many clusters.

Why not empirical Bayes? EB shrinks toward zero. This is optimal under no effect but harmful under treatment: it pulls the trajectory toward the wrong target. The Kalman smoother shrinks toward a smooth trajectory—a more appropriate prior when treatment effects are nonzero but evolving.

Why not parametric trends? A linear trend model works when the pre-trend is linear, but fails catastrophically otherwise (MSE 4× worse than raw for the gradual pattern). The Kalman smoother's nonparametric flexibility avoids this risk.

Pre-testing concerns. Roth (2022) shows that conditioning on passing a pre-test distorts inference. Our test detects slightly more violations, though the power gains are modest. However, violations that survive our test are precisely those the smoother has shrunk the most, raising questions about the interaction between power gains and conditioning bias that warrant further investigation. Researchers concerned about conditioning bias should

consider sensitivity analysis (Rambachan and Roth, 2023) rather than relying solely on pre-tests, regardless of which test statistic is used.

Duality with Rambachan and Roth (2023). Our state-space model and the sensitivity analysis of Rambachan and Roth (2023) share a direct mathematical duality. Their smoothness class $\Delta^{\text{SD}}(M)$ bounds the L_∞ norm of the second difference of the confounding trend, $|(\delta_{t+1} - \delta_t) - (\delta_t - \delta_{t-1})| \leq M$. In our local linear trend state-space model (when $q_\ell = 0$), the slope transition places a Gaussian prior on that exact second difference, $\Delta^2\beta_{t+1} = w_{2,t} \sim \mathcal{N}(0, q_s)$, penalizing its L_2 norm with weight $1/q_s$. Thus Rambachan and Roth (2023) bound worst-case identification bias under an L_∞ smoothness restriction on unobserved confounding, whereas the Kalman smoother minimizes estimation MSE under an L_2 Gaussian smoothness prior on the treatment effect trajectory.

Choice of Q. This is a tuning parameter analogous to bandwidth in nonparametric regression. Marginal likelihood maximization, cross-validation, or calibration to the observed $\hat{\beta}_t$ variability are all viable selection methods.

Why not full Bayesian inference? A natural extension would specify priors on $\mathbf{Q}$ and perform full posterior inference via MCMC. The Kalman filter provides a practical middle ground: closed-form solutions, computational simplicity, and no prior elicitation beyond the state-space structure. The frequentist bootstrap provides valid inference without requiring researchers to defend Bayesian prior choices.

Derivative interpretation. The derivative estimates $\widehat{\Delta\beta}_t^{\text{KS}}$ inherit more from the smoothness assumption than the level estimates. With noisy observations, there is limited direct information about $\Delta\beta_t$—the prior plays a larger role. Researchers should interpret $\widehat{\Delta\beta}_t^{\text{KS}}$ as "the derivative under the smoothness assumption" rather than a model-free estimate. The 78–99% MSE reduction reflects that raw first differences are dominated by noise, so even a weak prior helps substantially.

Replication materials. Code and replication materials are available at `https://github.com/finite-sample/smooth-operator`.

8 Conclusion

Event study coefficient sequences are noisy observations of a smooth latent process with known, heteroskedastic measurement error. The Kalman smoother is the optimal linear estimator for this setting, and substantially outperforms competitive baselines—including James–Stein shrinkage, empirical Bayes, and Savitzky–Golay—for both level and derivative estimation. The MSE gains are particularly large for derivative estimation (98% reduction), enabling researchers to visualize and interpret treatment effect dynamics with much greater precision. For parallel trends testing, the Kalman-based tests achieve correct size via parametric bootstrap and modest power improvements over the standard Wald test. The primary contribution is thus in *estimation*—recovering smooth trajectories and their derivatives from noisy event study coefficients—rather than dramatic power gains for hypothesis testing.

References

Borusyak, K., Jaravel, X., and Spiess, J. (2024). Revisiting event-study designs: Robust and efficient estimation. *Review of Economic Studies*, 91(6):3253–3306.

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

Cameron, A. C., Gelbach, J. B., and Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *Review of Economics and Statistics*, 90(3):414–427.

Durbin, J. and Koopman, S. J. (2012). *Time Series Analysis by State Space Methods*. Oxford University Press, 2nd edition.

Efron, B. and Tibshirani, R. J. (1994). *An Introduction to the Bootstrap*. Chapman & Hall/CRC.

Freyaldenhoven, S., Hansen, C., and Shapiro, J. M. (2019). Pre-event trends in the panel event-study design. *American Economic Review*, 109(9):3307–3338.

Harvey, A. C. (1985). Trends and cycles in macroeconomic time series. *Journal of Business & Economic Statistics*, 3(3):216–227.

Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press.

Rambachan, A. and Roth, J. (2023). A more credible approach to parallel trends. *Review of Economic Studies*, 90(5):2555–2591.

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

Sun, L. and Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2):175–199.