

Own-Cluster Reuse Bias in Orthogonal Panel Estimation*

Zixi Zhou[†]

Abstract

Estimating a unit coordinate from the same history used to evaluate an orthogonal panel score can create a drift of order T^{-1} . This drift can shift the $\sqrt{NT}$ limit when N/T converges to a positive constant. In a partially linear model, we separate derivative–influence covariance, score curvature, and propagation of first-step bias. The last component vanishes as the nuisance sieve grows, while the other two may reinforce or cancel. We estimate the drift from the full unit history and construct valid standard errors under independent units, deterministic quasi-uniform coordinates, finite-range dependence, and tensor-product B-splines. Separate results cover sample means and regular smooth scalar moment coordinates. For a sample mean and a fixed sieve, buffered splitting removes covariance but retains curvature; an equal split doubles its leading coefficient. Simulations show cancellation of substantial bias components and improved coverage after correction.

1 Introduction

An orthogonal panel score can have a first-order asymptotic bias when a unit coordinate is estimated from the same history used to evaluate the score. The coordinate error is shared by every observation in the unit. Its covariance with the score derivative and its interaction with score curvature produce a drift of order T^{-1} . When N/T converges to a positive constant, this drift is as large as the sampling error of a $\sqrt{NT}$ -consistent estimator. This paper characterizes that drift and estimates a correction while retaining the full unit history.

The distinction between a shared nuisance function and a unit-specific parameter explains the source of the problem. Regression functions can be learned from observations pooled across units. A persistent coordinate representing risk, productivity, or ability must be learned from repeated measurements on its own unit unless additional cross-unit structure is available. Cross-fitting the shared regressions across units therefore leaves a statistical dependence between the generated coordinate and the score contributions of that unit. We call this dependence *own-cluster reuse*. Within-unit splitting can separate coordinate estimation from score evaluation, but it also shortens the history used to estimate the coordinate.

A simple example shows why a zero expected coordinate derivative does not remove the resulting bias. Let $Z_{it} = \alpha_i + \zeta_{it}$, $D_{it} = v_{it}$, and $Y_{it} = \theta_0 v_{it} + \alpha_i + \varepsilon_{it}$, with centered stationary errors and ε_{it} independent of the joint (v_{it}, ζ_{it}) process. The oracle score at coordinate a is $\psi_{it}(a) = v_{it}(\varepsilon_{it} + \alpha_i - a)$. Its coordinate derivative is $-v_{it}$ and has expectation zero. Nevertheless, substituting the sample mean $\check{\alpha}_i = T^{-1} \sum_{s=1}^T Z_{is}$ gives

$$\frac{1}{T} \sum_{t=1}^T \mathbb{E}_i[\psi_{it}(\check{\alpha}_i)] = -\frac{1}{T^2} \sum_{t=1}^T \sum_{s=1}^T \mathbb{E}_i[v_{it}\zeta_{is}].$$

*This version: October 16, 2026. This work is preliminary and any comments are welcome.

[†]Department of Economics, Duke University. zixi.zhou@duke.edu.

For a joint moving-average process, the leading coefficient of this order- T^{-1} drift is the negative long-run covariance. The expected derivative and the derivative–coordinate-error covariance are different objects.

The general partially linear score adds curvature to this mechanism. A local Taylor expansion multiplies the random score derivative by the coordinate error and the second derivative by its squared error. Both expected contributions have order T^{-1} under the dependence conditions studied here. A smooth scalar moment first step can also be biased at that order, adding its bias multiplied by the population score derivative. We denote these three score-drift components by $B_K^{(1)}$, $B_K^{(2)}$, and $B_K^{(3)}$, with K the nuisance-sieve dimension. For a sample mean, $B_K^{(3)} = 0$. In the partially linear model, the population derivative multiplying first-step bias tends to zero as the sieve approximates the nuisance functions. Derivative–influence covariance and curvature remain as the two possible sources of asymptotic bias. They can reinforce or cancel, so a small total bias can conceal large components with different responses to splitting.

The effect of the drift depends on the estimator’s convergence rate. Under $N/T \rightarrow \kappa_0 \in (0, \infty)$, $\sqrt{NT}/T = \sqrt{N/T} \rightarrow \sqrt{\kappa_0}$. An order- T^{-1} score drift therefore enters the first-order limit. We estimate this drift using fitted coordinate derivatives and observable first-step lag products. With $B_K = B_K^{(1)} + B_K^{(2)} + B_K^{(3)}$ and positive score sensitivity J_K , the correction takes the form

$$\tilde{\theta}_K = \hat{\theta}_K - \frac{\hat{B}_K}{T\hat{J}_K}.$$

The standard error includes the influence-function contribution of the generated coordinate. Under the stated growing-sieve conditions, the corrected estimator has a centered Gaussian limit and valid studentized inference. The finite-sieve target θ_K remains distinct from the structural target θ_0 : sieve approximation bias and own-cluster reuse drift enter through different terms.

The theory separates the mechanism from the conditions needed for feasible inference. Fixed-sieve population expansions identify the two sample-mean channels and compare full-history estimation with buffered splitting. For that population comparison, buffering removes the derivative–influence covariance under fixed-range dependence. It leaves curvature, and an equal split doubles the leading curvature coefficient because each coordinate uses half the history. Growing-sieve results establish feasible correction and studentization under independent units, deterministic quasi-uniform coordinates, independent uniform covariates, bounded finite-range dependence, and tensor-product B-splines. The sample-mean result uses its exact linear representation and a mixed third-moment restriction. The smooth scalar moment result permits a biased first step and dispenses with that cancellation, under a narrower sufficient sieve window and a separate observable unique-root procedure.

Huang [Hua26] studies fixed effects as generated regressors and makes this rate comparison explicit. He identifies the same-unit dependence between an estimated fixed effect and a downstream moment derivative, and explains why ordinary cross-fitting cannot remove it. His target uses one cross-sectional moment per unit and admits $\sqrt{N}$ inference under $\sqrt{N}/T \rightarrow 0$, together with conditions controlling the remaining derivative–error dependence. At that target rate, order- T^{-1} terms vanish. Our target averages panel scores and has rate $\sqrt{NT}$; under $N/T \rightarrow \kappa_0 > 0$, the corresponding terms remain first order. We derive their long-run coefficients and construct a feasible bias correction for inference at rate $\sqrt{NT}$.

This analysis builds on generated-regressor and incidental-parameter theory. Classical two-step results characterize the contribution of a preliminary estimator to the influence function and asymptotic variance [New84, Pag84, MT85, HR13]; later work studies semiparametrically generated covariates and many-regressor first steps [MRS16, CJM19]. Large- T panel corrections address bias

from estimated individual effects [HN04, HK11, FVV11, FVL13, FVW18], and analytical and split-sample corrections can have different higher-order variance properties [HHKN24]. The covariance component is the panel analogue of leave-in bias, while score curvature produces nonlinearity bias [CCJ13, CJM19]. Here a coordinate error learned from a dependent unit history is reused in all T score contributions, so the bias coefficients involve long-run covariances and variances. The drift can remain consequential even at fixed K .

Neyman orthogonality and locally robust estimation control the first-order nuisance directions for which they are imposed [Ney59, BKRW93, New94, CCD⁺18, CEI⁺22]. Higher-order influence functions and higher-order orthogonal scores reduce sensitivity beyond first order [RLTvdV08, vdV14, MSZ18]. Bonhomme, Jochmans, and Weidner [BJW26] construct higher-order orthogonal equations in likelihood models, and Bonhomme, Jochmans, Newey, and Weidner [BJNW26] develop the corresponding approach for parametric moment models. We retain the partially linear profiled score and estimate the drift caused by a specified coordinate-estimation operation. This approach complements locally robust estimation with generated regressors [EPI26], panel DML under dependent sampling [CKMS22, CP26], and inference using noisy unit-level estimates [JW24]. Related work adjusts orthogonal regression moments for estimated two-way fixed effects [FK26].

The simulations separate the bias mechanisms before evaluating inference. In a cancellation design, estimated channel coefficients of -0.440 and 0.440 leave the uncorrected scaled bias near zero, while buffered splitting exposes the remaining curvature. With nonlinear nuisances and $N = T = 200$, correction reduces scaled bias from 0.744 to 0.008 and raises nominal 95% coverage from 89.5% to 95.9%. A ratio-coordinate experiment implements the smooth-moment procedure with a biased first step whose population score multiplier is zero. The experiment illustrates how first-step bias and its propagation to the target can differ.

Section 2 defines the model, orthogonal score, and sieve target. Section 3 explains the sample-mean bias channels and buffered splitting. Sections 4 and 5 give the feasible correction and its asymptotic theory. Section 6 treats smooth scalar moment coordinates. Section 7 reports the numerical evidence, and Section 8 concludes. The appendices supply the fixed-sieve expansions and the proof details for both feasible procedures.

2 Model and Orthogonal Score

2.1 Panel model and unit coordinates

For units $i = 1, \dots, N$ and dates $t = 1, \dots, T$, observe $W_{it} = (Y_{it}, D_{it}, X_{it}, Z_{it})$. A scalar unit-specific parameter α_i enters both regression functions; we refer to it as the unit coordinate. The model is

$$Z_{it} = \alpha_i + \zeta_{it}, \tag{2.1}$$

$$D_{it} = m_0(X_{it}, \alpha_i) + v_{it}, \tag{2.2}$$

$$Y_{it} = \theta_0 D_{it} + h_0(X_{it}, \alpha_i) + \varepsilon_{it}. \tag{2.3}$$

The scalar θ_0 is the target. The functions m_0, h_0 are shared across units; each α_i is learned from its own repeated observations. Let $\mathcal{C}_i$ contain the persistent unit information, including α_i . Write $\mathbb{E}_i$ and $\mathbb{P}_i$ for expectation and probability under the conditional innovation law given $\mathcal{C}_i$. The sigma-fields $\mathcal{C}_i \vee \sigma(W_{it} : t \in \mathbb{Z})$ are independent across units. Within a unit, observations can be serially dependent.

The normalization $\alpha_i = \mathbb{E}_i[Z_{it}]$ gives $\mathbb{E}_i[\zeta_{it}] = 0$. For the baseline model, impose

$$\mathbb{E}[v_{it} \mid X_{it}, \mathcal{C}_i] = 0, \quad \mathbb{E}[\varepsilon_{it} \mid D_{it}, X_{it}, \mathcal{C}_i] = 0. \tag{2.4}$$

Persistent information beyond α_i can affect higher moments and serial dependence, while these restrictions fix the conditional means. The raw coordinate estimate is

$$\check{\alpha}_i = \frac{1}{T} \sum_{t=1}^T Z_{it}, \quad \Delta_i = \check{\alpha}_i - \alpha_i = \frac{1}{T} \sum_{t=1}^T \zeta_{it}. \quad (2.5)$$

The fixed-sieve results use this raw mean. For growing sieves, $\hat{\alpha}_i$ denotes its clipped version, specified below. Section 6 replaces the measurement equation by a known smooth scalar moment $g(W_{it}, a)$.

Conditional on $\mathcal{C} = \sigma(\mathcal{C}_i : 1 \leq i \leq N)$, population projection coefficients and targets are held fixed when taking $\mathbb{E}_i$. A coordinate estimated from unit i is integrated jointly with that unit's observations. In particular, taking this expectation preserves the reuse dependence. We write sample averages and population averages as explicit sums. For scalar or Euclidean-vector Q , let $\|Q\|_{L_i^r} = (\mathbb{E}_i \|Q\|^r)^{1/r}$.

2.2 The orthogonal score and its sieve target

Define the reduced-form function $\ell_0 = \theta_0 m_0 + h_0$ and the shared nuisance pair $\eta_0 = (m_0, \ell_0)$. We use the partialling-out score [CCD⁺18],

$$\psi(W; \theta, \eta, a) = \{D - m(X, a)\} [Y - \ell(X, a) - \theta \{D - m(X, a)\}], \quad \eta = (m, \ell).$$

At the true coordinate and nuisance functions it equals $v_{it}\varepsilon_{it}$. Its expected first derivative in either regression-function direction is zero under (2.4). This orthogonality controls estimation of the shared functions. Estimation of a coordinate using its own history requires a separate expansion.

Let $\mathcal{S}_K = \{p_K^\top c : c \in \mathbb{R}^K\}$ be a differentiable sieve. Define m_K, ℓ_K as the least-squares projections of D, Y onto $\mathcal{S}_K$ under the panel-grid population criterion $N^{-1} \sum_i \mathbb{E}_i [T^{-1} \sum_t \{Q_{it} - f(X_{it}, \alpha_i)\}^2]$, for $Q = D, Y$, respectively. These targets depend on the current design and on N ; we suppress that dependence in their indices. Set $\eta_K = (m_K, \ell_K)$ and $h_K = \ell_K - \theta_0 m_K$. The coefficient implementation and projection measures appear in Appendix A; continuous Lebesgue projection and panel-grid projection remain distinct.

Write $u_{D,it,K} = D_{it} - m_K(X_{it}, \alpha_i)$ and $u_{Y,it,K} = Y_{it} - \ell_K(X_{it}, \alpha_i)$. Define the positive sensitivity and finite-sieve target by

$$J_K = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \mathbb{E}_i [u_{D,it,K}^2] > 0, \\ \theta_K = \frac{1}{NTJ_K} \sum_{i=1}^N \sum_{t=1}^T \mathbb{E}_i [u_{D,it,K} u_{Y,it,K}].$$

The projection equations make the expected score orthogonal along sieve directions at (θ_K, η_K) . They need not give orthogonality along all infinite-dimensional directions at a fixed K .

The finite-sieve target differs from θ_0 through projection error. Let $r_{m,K} = m_0 - m_K$ and $r_{h,K} = h_0 - h_K$. The conditional-mean restrictions and projection equations give

$$\theta_K - \theta_0 = \frac{1}{NTJ_K} \sum_{i=1}^N \sum_{t=1}^T \mathbb{E}_i [r_{m,K}(X_{it}, \alpha_i) r_{h,K}(X_{it}, \alpha_i)]. \quad (2.6)$$

For the tensor-spline construction below, let $d_W = d_X + 1$ denote the dimension of (x, a) . If J_K is bounded away from zero and the projection errors have respective L^2 orders $K^{-\tau_m/d_W}$ and $K^{-\tau_h/d_W}$, then $|\theta_K - \theta_0| \leq CK^{-(\tau_m + \tau_h)/d_W}$.

Appendix A.3 derives the identity. This approximation bias is separate from the reuse bias studied next.

Define $\psi_{it,K}(a) = \psi(W_{it}; \theta_K, \eta_K, a)$ and $\psi_{it,K}^{(r)}(a) = \partial_a^r \psi_{it,K}(a)$ for $r = 1, 2, 3$. Without an argument, these expressions are evaluated at $a = \alpha_i$. The derivative holds the observation, target, and nuisance functions fixed. In contrast, $\hat{\psi}_{it,K}$ evaluates the score at all fitted arguments. Thus $\psi_{it,K}(\hat{\alpha}_i)$ and $\hat{\psi}_{it,K}$ are different objects.

For later computation, write $u = D - m(X, a)$ and $e = Y - \ell(X, a) - \theta u$. The two derivatives that enter the bias are

$$\begin{aligned}\partial_a \psi &= -m_a e - u(\ell_a - \theta m_a), \\ \partial_a^2 \psi &= -m_{aa} e - u(\ell_{aa} - \theta m_{aa}) + 2m_a(\ell_a - \theta m_a).\end{aligned}$$

Here $m_a = \partial_a m$ and $m_{aa} = \partial_a^2 m$, with the same convention for ℓ . These formulas also apply to the fitted sieve functions.

3 Own-Cluster Reuse Bias

3.1 Two channels from estimating a sample mean

The linear example in the Introduction has $\psi_{it,K}^{(1)} = -v_{it}$ and $\psi_{it,K}^{(2)} = 0$ when the sieve reproduces constants and the coordinate. Its expected derivative is zero, while the empirical derivative and coordinate error remain correlated.

More generally, Taylor expansion around α_i gives

$$\frac{1}{T} \sum_t \{\psi_{it,K}(\check{\alpha}_i) - \psi_{it,K}(\alpha_i)\} = \Delta_i \frac{1}{T} \sum_t \psi_{it,K}^{(1)} + \frac{\Delta_i^2}{2T} \sum_t \psi_{it,K}^{(2)} + R_{iT}.$$

The first term pairs a score derivative with the mean coordinate noise. The second pairs curvature with the coordinate-error variance. Both have expectation of order T^{-1} under the following fixed-sieve conditions.

Assumption 3.1 (Fixed-sieve sample-mean regularity). *The following conditions hold for a common fixed $\delta > 0$ and a common compact interval $\mathcal{A}$. Write $I_i = [\alpha_i - \delta, \alpha_i + \delta]$.*

(i) Sampling and dependence. *For fixed K , the unit arrays are independent across i . Conditional on $\mathcal{C}_i$, the joint time array containing ζ_{it} , the target score map $a \mapsto \psi_{it,K}(a)$ and the projection-moment maps $a \mapsto p_K(X_{it}, a)\{D_{it} - m_K(X_{it}, a)\}$ and $a \mapsto p_K(X_{it}, a)\{Y_{it} - \ell_K(X_{it}, a)\}$ on $\mathcal{A}$, their first three coordinate derivatives on I_i , and the map $a \mapsto p_K(X_{it}, a)$ with its first coordinate derivative on $\mathcal{A}$ is strictly stationary and L -dependent for a fixed L .*

(ii) Local smoothness. *The variables in part (i) are uniformly bounded on their displayed domains, the highest displayed coordinate derivative is uniformly Lipschitz on I_i , and $Z_{it} \in \mathcal{A}$ almost surely. Moreover, $I_i \subset \mathcal{A}$ uniformly in i .*

(iii) Coordinate localization. *The generated coordinate is the raw sample mean in (2.5). The true coordinates remain in a compact interior set, and there are constants $C, c > 0$, independent of i, N , and T , such that*

$$\sup_{1 \leq i \leq N} \mathbb{P}(\check{\alpha}_i \notin I_i | \mathcal{C}_i) \leq C \exp(-cT). \quad (3.1)$$

The three moment maps in part (i) form the stacked score used in Appendix A. These conditions concern bounded oracle maps and local derivatives; fitted arrays using a whole history do not

themselves inherit finite-range dependence. The bounded envelope controls the exponentially rare failure of local coordinate localization.

Let $\mu_{iK} = \mathbb{E}_i[\psi_{i0,K}^{(1)}]$, $\nu_{iK} = \mathbb{E}_i[\psi_{i0,K}^{(2)}]$, and $\Omega_i = \sum_{|j| \leq L} \mathbb{E}_i[\zeta_{i0}\zeta_{ij}]$. For the sample mean the first-step influence is $\xi_{it} = \zeta_{it}$. The leading score drift is

$$B_K^{(1)} = \frac{1}{N} \sum_{i=1}^N \sum_{|j| \leq L} \mathbb{E}_i[\psi_{i0,K}^{(1)} \zeta_{ij}],$$

$$B_K^{(2)} = \frac{1}{2N} \sum_{i=1}^N \nu_{iK} \Omega_i, \quad B_K = B_K^{(1)} + B_K^{(2)}.$$

Parenthesized superscripts identify bias channels. An ordinary power applies to the whole coefficient, as in $(B_K)^2$. A third channel enters for a biased first step in Section 6.

Proposition 3.2 (Population drift and its estimator consequence). *Fix K and suppose Assumption 3.1 holds uniformly in i, N . As $T \rightarrow \infty$, allowing $N = N_T$,*

$$\frac{1}{N} \sum_{i=1}^N \sum_{t=1}^T \mathbb{E}_i[\psi_{it,K}(\check{\alpha}_i) - \psi_{it,K}(\alpha_i)] = B_K + o(1). \quad (3.2)$$

If also Assumption C.2 holds and $N/T \rightarrow \kappa \in (0, \infty)$, the globally reported raw-coordinate profile root of Appendix D.5 satisfies

$$\sqrt{NT}(\hat{\theta}_K - \theta_K) = \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T \phi_{it,K} + \sqrt{\frac{N}{T}} \frac{B_K}{J_K} + o_p(1), \quad (3.3)$$

where the centered influence is

$$\phi_{it,K} = J_K^{-1} \{ \psi_{it,K} - \mathbb{E}_i[\psi_{it,K}] + \mu_{iK} \zeta_{it} \}. \quad (3.4)$$

The population conclusion requires no estimator-stability assumption. The second conclusion adds that condition to control the profile root. Appendix C.2 proves both parts, with the centered sampling expansion isolated as Lemma A.1. The centering in (3.4) is unit specific: the population target sets the average expected score to zero, rather than every unit's expected score separately. The term $\mu_{iK} \zeta_{it}$ affects sampling variation; the coefficients $B_K^{(1)}, B_K^{(2)}$ affect centering.

3.2 Buffered splitting

Consider two contiguous date blocks separated by a buffer wider than L : $\mathcal{I}_{1,T} = \{1, \dots, \lfloor T/2 \rfloor - L\}$ and $\mathcal{I}_{2,T} = \{\lfloor T/2 \rfloor + L + 1, \dots, T\}$. Their sizes are T_1, T_2 , with $T_1 + T_2 = T - 2L$. For evaluation in block r , use the other block's raw mean $\check{\alpha}_i^{(-r)}$. Define $S_{iK} = (T - 2L)^{-1} \sum_{r=1}^2 \sum_{t \in \mathcal{I}_{r,T}} \psi_{it,K}(\check{\alpha}_i^{(-r)})$, and let $S_{iK,0}$ use α_i instead.

Proposition 3.3 (Splitting removes covariance and retains curvature). *Under Assumption 3.1, with K fixed, each fold's linear derivative term has expectation zero exactly, and*

$$\frac{T}{N} \sum_{i=1}^N \mathbb{E}_i[S_{iK} - S_{iK,0}] = 2B_K^{(2)} + o(1). \quad (3.5)$$

For a one-way buffered split with a coordinate-training block of size $T_1/T \rightarrow \lambda \in (0, 1)$, the corresponding coefficient is $\lambda^{-1} B_K^{(2)} + o(1)$.

The proof in Appendix C.3 uses independence between the separated blocks to remove reuse covariance. The shorter training history increases coordinate-error variance, doubling the leading curvature coefficient for equal folds. The proposition compares expected scores at fixed K .

4 Feasible Bias Correction

4.1 Profile estimation

For the growing-sieve sample-mean procedure, clip the raw mean as $\hat{\alpha}_i = \Pi_{[b_T, 1-b_T]}(\check{\alpha}_i)$, where $b_T = C_b \sqrt{\log N/T}$ and the fixed C_b satisfies Assumption 5.1(iii). At these coordinates, regress D and Y on $p_K(X_{it}, \hat{\alpha}_i)$ using all panel observations. Denote the fitted functions by $\hat{m}_K, \hat{\ell}_K$ and their residuals by $\hat{u}_{D,it,K} = D_{it} - \hat{m}_K(X_{it}, \hat{\alpha}_i)$ and $\hat{u}_{Y,it,K} = Y_{it} - \hat{\ell}_K(X_{it}, \hat{\alpha}_i)$. Set $\hat{\eta}_K = (\hat{m}_K, \hat{\ell}_K)$. On the acceptance event $\mathcal{G}_K$, the profile estimator is

$$\hat{J}_K = \frac{1}{NT} \sum_{i,t} \hat{u}_{D,it,K}^2, \quad \hat{\theta}_K = \frac{\sum_{i,t} \hat{u}_{D,it,K} \hat{u}_{Y,it,K}}{\sum_{i,t} \hat{u}_{D,it,K}^2}.$$

Acceptance requires the generated-coordinate Gram matrix to have minimum eigenvalue at least c_Σ , and the residual second moment to be at least $c_V/2$. The deterministic admissible thresholds and complete candidate and fallback rules are given in Appendix A.2. On rejection, the sample-mean procedure reports the fixed vector specified there, including its slope $\theta_K^\dagger$.

4.2 Bias channels and standard errors

Let L be a supplied fixed upper bound on dependence, and $I_T(j) = \{t : 1 \leq t \leq T, 1 \leq t+j \leq T\}$. Set $\hat{\xi}_{it} = Z_{it} - \hat{\alpha}_i$ for this sample-mean procedure. The fitted derivatives $\hat{\psi}_{it,K}^{(r)}$ use $(\hat{\theta}_K, \hat{\eta}_K, \hat{\alpha}_i)$. Write $\hat{\mu}_{iK} = T^{-1} \sum_t \hat{\psi}_{it,K}^{(1)}$ and $\hat{\nu}_{iK} = T^{-1} \sum_t \hat{\psi}_{it,K}^{(2)}$. Estimate the two channels on $\mathcal{G}_K$ by

$$\begin{aligned} \hat{B}_K^{(1)} &= \frac{1}{NT} \sum_{i=1}^N \sum_{|j| \leq L} \sum_{t \in I_T(j)} \hat{\psi}_{it,K}^{(1)} \hat{\xi}_{i,t+j}, \\ \hat{\Omega}_i &= \frac{1}{T} \sum_{|j| \leq L} \sum_{t \in I_T(j)} \hat{\xi}_{it} \hat{\xi}_{i,t+j}, \quad \hat{B}_K^{(2)} = \frac{1}{2N} \sum_{i=1}^N \hat{\nu}_{iK} \hat{\Omega}_i. \end{aligned}$$

These uncentered lag products are divided by T , including at boundary dates. With $\hat{B}_K = \hat{B}_K^{(1)} + \hat{B}_K^{(2)}$, report

$$\tilde{\theta}_K = \hat{\theta}_K - \frac{\hat{B}_K}{T \hat{J}_K} \quad (4.1)$$

on acceptance and $\tilde{\theta}_K = \theta_K^\dagger$ on rejection. Since $\partial_\theta \psi = -\{D - m(X, a)\}^2$, positive score drift moves the uncorrected root upward. The correction therefore subtracts the score coefficient divided by the positive sensitivity.

To estimate sampling variation, use $\hat{\phi}_{it,K} = \hat{J}_K^{-1}(\hat{\psi}_{it,K} + \hat{\mu}_{iK} \hat{\xi}_{it})$ and its within-unit sample mean $\bar{\phi}_{iK} = T^{-1} \sum_t \hat{\phi}_{it,K}$. The raw variance statistic is

$$\hat{q}_K = \frac{1}{NT} \sum_{i=1}^N \sum_{|j| \leq L} \sum_{t \in I_T(j)} (\hat{\phi}_{it,K} - \bar{\phi}_{iK})(\hat{\phi}_{i,t+j,K} - \bar{\phi}_{iK}).$$

This heteroskedasticity and autocorrelation consistent calculation retains all possibly nonzero covariance lags. On $\mathcal{G}_K$, report $\hat{\sigma}_K^2 = \hat{q}_K$ if $\hat{q}_K > 0$ and $\hat{\sigma}_K^2 = 1$ otherwise; on rejection set $\hat{\sigma}_K^2 = 1$. In every case $\hat{\sigma}_K = \sqrt{\hat{\sigma}_K^2}$. The standard error of $\tilde{\theta}_K$ is $\hat{\sigma}_K/\sqrt{NT}$.

5 Asymptotic Theory

5.1 Primitive conditions

The growing-sieve result uses a sufficient tensor-spline benchmark. Quasi-uniform coordinates and independent uniform covariates control projection and quadrature. Bounded finite-range dependence supplies concentration and moment bounds. A mixed third-moment cancellation permits the wider sample-mean sieve window.

Let $W^+ = [0, 1]^{d_X} \times [0, 1]$, $d_W = d_X + 1$, and $b_T = C_b \sqrt{\log N/T}$. The spline mesh is λ_K .

Assumption 5.1 (Growing-sieve conditions). *The following conditions hold.*

(i) Coordinate design and covariates. *The coordinates are deterministic, and the benchmark has no additional persistent unit-level randomness: $\mathcal{C}_i = \sigma(\alpha_i)$. Thus $\mathbb{E}_i$ integrates over the common covariate and disturbance law specified below. Let $\alpha_{(1)} \leq \dots \leq \alpha_{(N)}$ be their ordered values and let F_N be their empirical distribution function. All coordinates lie in $[2b_T, 1 - 2b_T]$. If F_T^I denotes the distribution function of the uniform law on $[2b_T, 1 - 2b_T]$, then, for fixed constants $0 < c_\alpha < C_\alpha < \infty$,*

$$\begin{aligned} \frac{c_\alpha}{N} &\leq \alpha_{(j+1)} - \alpha_{(j)} \leq \frac{C_\alpha}{N}, \quad j = 1, \dots, N-1, \\ \sup_{a \in \mathbb{R}} |F_N(a) - F_T^I(a)| &\leq \frac{C_\alpha}{N}. \end{aligned} \tag{5.1}$$

The variables in $\{X_{it}\}$ are independent and uniformly distributed on $[0, 1]^{d_X}$ and are independent of the coordinate design and all disturbance processes.

(ii) Independent units and within-unit dependence. *Fix an integer $L \geq 1$ that does not vary with (N, T) . Viewed as random elements indexed by $\mathbb{Z}$, the joint disturbance histories $\{(\zeta_{it}, v_{it}, \varepsilon_{it}) : t \in \mathbb{Z}\}$ are independent across i and have the same law. The bivariate process $\{(\zeta_{it}, v_{it}) : t \in \mathbb{Z}\}$ is bounded, centered, strictly stationary, and L -dependent: for every integer s , the σ -fields generated by dates $t \leq s$ and $t \geq s + L + 1$ are independent. The process $\{\varepsilon_{it} : t \in \mathbb{Z}\}$ is bounded, centered, strictly stationary, and L -dependent, and is independent of the entire (ζ, v) process. Assume $\sigma_v^2 = \mathbb{E}[v_{it}^2] > 0$ and $0 < \text{Var}(\varepsilon_{it}) < \infty$. Assume the mixed third-moment restriction*

$$\mathbb{E}[v_{i0} \zeta_{ij_1} \zeta_{ij_2}] = 0, \quad (j_1, j_2) \in \mathbb{Z}^2. \tag{5.2}$$

(iii) Generated coordinate. *The data satisfy (2.1)–(2.3). The raw mean $\check{\alpha}_i$ is clipped as $\hat{\alpha}_i = \Pi_{[b_T, 1-b_T]}(\check{\alpha}_i)$. Write $d_i = \hat{\alpha}_i - \alpha_i$ and $c_i = d_i - \Delta_i$. The finite-range Hoeffding constant c_{fm} satisfies $c_{\text{fm}} C_b^2 > 1$.*

(iv) Smoothness and sieve. *The functions satisfy $m_0 \in C^{\tau_m}(W^+)$ and $h_0 \in C^{\tau_h}(W^+)$ with $3 < \tau_m, \tau_h \leq p + 1$. The basis is an open-clamped, maximum-smoothness tensor-product B-spline basis of degree $p \geq 4$, with quasi-uniform positive knot spans and Lebesgue- L^2 normalization. Write $p_K(x, a) = p_{K_x}^x(x) \otimes p_{K_a}^a(a)$, where $K = K_x K_a$, $K_x \asymp \lambda_K^{-d_X}$, and $K_a \asymp \lambda_K^{-1}$. When $d_X = 0$, take $K_x = 1$. Thus $K \asymp \lambda_K^{-d_W}$.*

(v) *Asymptotic growth.*

$$\begin{aligned} \frac{N}{T} &\rightarrow \kappa_0 \in (0, \infty), & N\lambda_K &\rightarrow \infty, \\ \frac{b_T}{\lambda_K} &\rightarrow 0, & \frac{K \log K}{NT} &\rightarrow 0. \end{aligned} \quad (5.3)$$

Use the supplied range L for both bias and variance lag sums, with weight one at every integer lag $|j| \leq L$ and zero elsewhere. The exact endpoint grid $2b_T + (i-1)(1-4b_T)/(N-1)$ satisfies the coordinate-design condition; alignment between coordinates and spline knots is unnecessary.

The mixed third-moment restriction allows dependence between v and ζ . It eliminates a mixed third-moment term in the nuisance coefficient expansion. Joint central symmetry of the (ζ, v) process is sufficient, although stronger than the restriction. Bounded finite moving averages also satisfy the dependence condition. For innovations independent and identically distributed over (i, t) , $\mathbb{E}[u_{it}^v(u_{it}^\zeta)^2] = 0$ suffices for the mixed cancellation; ε innovations must be independent of the entire (u^ζ, u^v) process. Appendix A.5 gives the covariance calculation.

5.2 Channel limits and the main theorem

Let $\Omega_{v\zeta} = \sum_{|j| \leq L} \mathbb{E}[v_{i0}\zeta_{ij}]$ and $\Omega_\zeta = \sum_{|j| \leq L} \mathbb{E}[\zeta_{i0}\zeta_{ij}]$. The limiting coefficient has a direct primitive expression:

$$\begin{aligned} B = & -\Omega_{v\zeta} \int_{[0,1]^{d_X}} \{h_0(x, 1) - h_0(x, 0)\} dx \\ & + \Omega_\zeta \int_{[0,1]^{d_W}} \partial_a m_0(x, a) \partial_a h_0(x, a) dx da. \end{aligned} \quad (5.4)$$

Proposition 5.2 (Primitive limits of the two channels). *Under Assumption 5.1 and $\lambda_K \rightarrow 0$, $B_K^{(1)}$ converges to the first term in (5.4). The curvature and total coefficients satisfy*

$$B_K^{(2)} \rightarrow \Omega_\zeta \int_{[0,1]^{d_W}} \partial_a m_0(x, a) \partial_a h_0(x, a) dx da, \quad (5.5)$$

$$B_K \rightarrow B. \quad (5.6)$$

The exact clipped-coordinate drift also satisfies

$$\frac{1}{N} \sum_{i=1}^N \sum_{t=1}^T \mathbb{E}_i[\psi_{it,K}(\hat{\alpha}_i) - \psi_{it,K}(\alpha_i)] \rightarrow B. \quad (5.7)$$

Appendices E.5 and E.6 prove the separate channel limits and the effects of finite lag windows and clipping. Covariance bias requires both a nonzero long-run v - ζ covariance and a nonzero average boundary contrast in h_0 . Curvature bias depends on the average product of the two coordinate derivatives. The terms can have either relative sign.

For $\gamma_c(j) = \mathbb{E}[c_{i0}c_{ij}]$, $c \in \{v, \varepsilon\}$, the limiting variance in this benchmark is

$$\sigma^2 = \frac{\sum_{|j| \leq L} \gamma_v(j) \gamma_\varepsilon(j)}{\gamma_v(0)^2}. \quad (5.8)$$

It is strictly positive under the benchmark. The variance calculation and independent-unit central limit theorem are proved in Appendix E.4.

The sufficient dimension window is

$$\frac{d_W}{\tau_m + \tau_h} < q < \min \left\{ 1, \frac{2d_W}{\max\{d_W, 3\} + 4} \right\}, \quad K \asymp T^q. \quad (5.9)$$

Theorem 5.3 (Asymptotic bias and feasible correction). *Suppose Assumption 5.1 holds and $K \asymp T^q$ satisfies (5.9). Then*

$$\begin{aligned}\sqrt{NT}(\hat{\theta}_K - \theta_0) &= \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T \phi_{it,K} + \sqrt{\kappa_0} \frac{B}{\gamma_v(0)} + o_p(1), \\ \sqrt{NT}(\tilde{\theta}_K - \theta_0) &= \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T \phi_{it,K} + o_p(1),\end{aligned}\tag{5.10}$$

where the common leading sum converges to $\mathcal{N}(0, \sigma^2)$. Moreover, $\hat{\sigma}_K^2 \xrightarrow{p} \sigma^2$, and

$$\frac{\sqrt{NT}(\tilde{\theta}_K - \theta_0)}{\hat{\sigma}_K} \rightsquigarrow \mathcal{N}(0, 1).$$

The correction changes centering while preserving the leading influence and variance. Centered inference using the uncorrected root is invalid when $B \neq 0$. Appendix F.1 proves both expansions.

Because $K \asymp \lambda_K^{-d_W}$, the lower bound on q makes $TK^{-(\tau_m + \tau_h)/d_W} \rightarrow 0$, so the finite-sieve target is negligible at the inference scale. The upper bounds control nuisance estimation, the quadratic profile error, and fitted coordinate derivatives. The window is nonempty exactly when $\tau_m + \tau_h > \max\{d_W, (\max\{d_W, 3\} + 4)/2\}$. Appendix F.1 verifies the rates and their geometric counterparts.

If, additionally, the ζ , v , and ε processes are mutually independent and each is independent and identically distributed over time, then $B_K^{(1)} = 0$ and $\sigma^2 = \text{Var}(\varepsilon_{it}) / \text{Var}(v_{it})$. Curvature can remain. The specialization is proved in Appendix F.2.

5.3 Why profiling leaves a quadratic remainder

The proof separates score expansion from estimation of the shared functions. Define the estimation errors at generated coordinates by $e_{m,it} = (\hat{m}_K - m_K)(X_{it}, \hat{\alpha}_i)$ and $e_{\ell,it} = (\hat{\ell}_K - \ell_K)(X_{it}, \hat{\alpha}_i)$. On the event where the profile normal equations hold,

$$\frac{1}{NT} \sum_{i,t} \psi_{it,K}(\hat{\alpha}_i) = \hat{J}_K(\hat{\theta}_K - \theta_K) + \frac{1}{NT} \sum_{i,t} e_{m,it}(e_{\ell,it} - \theta_K e_{m,it}).$$

The normal equations eliminate terms linear in these nuisance estimation errors. The product term remains, with a negative sign after solving for the slope. These estimation errors differ from the projection errors in (2.6). Lemma A.2 gives the exact coefficient proof. The score expansion, clipping error, correction consistency, and variance consistency then control the remaining terms at the $\sqrt{NT}$ scale. Each argument uses the finite-range structure of oracle arrays before bounding errors from estimating the coordinate and nuisance functions.

6 Smooth Scalar Moment Extension

A local coordinate need not be a sample mean. This section allows a known smooth scalar moment to identify it, while retaining the same partially linear equations, grid projections, and target score. The moment can depend on both regression disturbances. Its estimation error can therefore have a nonzero second-order mean, which adds a third bias channel. The resulting feasible theorem uses a smaller sieve window and a different finite-sample reporting rule from the sample-mean theorem.

6.1 The First-Step Expansion and Three Bias Channels

Let $g(W, a)$ be a known scalar moment with $\mathbb{E}_i[g(W_{i0}, \alpha_i)] = 0$. A regular coordinate solves

$$\frac{1}{T} \sum_{t=1}^T g(W_{it}, a) = 0. \quad (6.1)$$

The equation is just identified and contains no estimated nuisance. The observations W_{it} include the variables needed to evaluate g ; a ratio coordinate, for example, uses (U_{it}, R_{it}) rather than Z_{it} alone. The observable root rule below differs from the oracle local branch used for the fixed-sieve calculation in Appendix C.4.

Write $G_i = \mathbb{E}_i[\partial_a g(W_{i0}, \alpha_i)]$, $H_i = \mathbb{E}_i[\partial_a^2 g(W_{i0}, \alpha_i)]$, and $\xi_{it} = -G_i^{-1} g(W_{it}, \alpha_i)$. Thus $\mathbb{E}_i[\xi_{it}] = 0$. For the supplied fixed dependence bound L , define $\Omega_i = \sum_{|j| \leq L} \mathbb{E}_i[\xi_{i0} \xi_{ij}]$. The first-step bias coefficient is

$$\beta_i = -G_i^{-1} \left\{ \sum_{|j| \leq L} \mathbb{E}_i[\{\partial_a g(W_{i0}, \alpha_i) - G_i\} \xi_{ij}] + \frac{1}{2} H_i \Omega_i \right\}. \quad (6.2)$$

The first term reflects the random sample Jacobian; the second reflects curvature of the first-step moment. Both vanish for $g(W, a) = Z - a$.

For $g((U, R), a) = U - aR$, with R bounded away from zero, the raw coordinate is $\check{\alpha}_i = \frac{\sum_t U_{it}}{\sum_t R_{it}}$ and $\xi_{it} = (U_{it} - \alpha_i R_{it}) / \mathbb{E}_i[R_{i0}]$. Its bias coefficient is

$$\beta_i = - \frac{\sum_{|j| \leq L} \text{Cov}_i(R_{i0}, U_{ij} - \alpha_i R_{ij})}{\mathbb{E}_i[R_{i0}]^2}.$$

Thus an affine moment can have a biased root through its random Jacobian.

Under the benchmark below, Lemma G.1 gives

$$\hat{\alpha}_i - \alpha_i = \frac{1}{T} \sum_t \xi_{it} + \frac{\beta_i}{T} + \tilde{q}_{iT} + r_{iT}, \quad (6.3)$$

where $\tilde{q}_{iT} = q_{iT} - \mathbb{E}_i[q_{iT}]$ and hence $\mathbb{E}_i[\tilde{q}_{iT}] = 0$. The bounds are $\|\tilde{q}_{iT}\|_{L_i^r} = O(T^{-1})$ and $\|r_{iT}\|_{L_i^r} = O(T^{-3/2})$ for $2 \leq r \leq 8$. Here q_{iT} is the quadratic expansion using the population Hessian H_i , defined in that lemma. The sample-Hessian expansion and its exact finite-window mean are treated separately in Appendix C.4. The centered quadratic term is not negligible within a unit; its average across independent units is negligible at the $\sqrt{NT}$ scale.

Use the population score $\psi_{it,K}(a)$ and its derivative means $\mu_{iK} = \mathbb{E}_i[\psi_{it,K}^{(1)}]$ and $\nu_{iK} = \mathbb{E}_i[\psi_{it,K}^{(2)}]$. The three score-drift channels are

$$\begin{aligned} B_K^{(1)} &= \frac{1}{N} \sum_i \sum_{|j| \leq L} \mathbb{E}_i[\psi_{i0,K}^{(1)} \xi_{ij}], \\ B_K^{(2)} &= \frac{1}{2N} \sum_i \nu_{iK} \Omega_i, \quad B_K^{(3)} = \frac{1}{N} \sum_i \mu_{iK} \beta_i, \quad B_K = B_K^{(1)} + B_K^{(2)} + B_K^{(3)}. \end{aligned} \quad (6.4)$$

These channels arise from derivative-influence covariance, score curvature, and propagation of first-step bias. Along the growing-sieve sequences below, Lemma G.2 shows that B_K differs from T times the expected panel-average score change by $O(T^{-1/2})$. The change compares coordinates $\hat{\alpha}_i$ and α_i .

The fixed-sieve calculation under local identification is given in Appendix C.4. It also applies to scalar downstream scores with bounded first three coordinate derivatives and the stated finite-range dependence.

The third channel is small in the partially linear benchmark. Using the projection errors $r_{m,K}, r_{h,K}$ defined in Section 2.2, differentiation at fixed data gives

$$\mu_{iK} = \mathbb{E}_X [-(\partial_a m_K) r_{h,K} - r_{m,K} \partial_a h_K + 2(\theta_K - \theta_0) r_{m,K} \partial_a m_K], \quad (6.5)$$

Here $\mathbb{E}_X$ integrates over the uniform covariate law, with functions evaluated at (X, α_i) . The projection bounds give $\sup_i |\mu_{iK}| \leq CK^{-\min\{\tau_m, \tau_h\}/d_W}$ under either growing-sieve benchmark. Since the first-step bias coefficients are uniformly bounded, $|B_K^{(3)}| \leq CK^{-\min\{\tau_m, \tau_h\}/d_W} \rightarrow 0$. The feasible correction retains this channel: it can matter at finite K and for other downstream scores. Appendix C.5 also shows why it must be retained after a nonlinear re-expression of the same coordinate, even when the original coordinate is a sample mean.

6.2 Observable Roots and Primitive Conditions

For this benchmark, write $V = \mathbb{E}[v_{i0}^2]$ and use σ^2 for its limiting variance:

$$\sigma^2 = V^{-2} \sum_{|j| \leq L} \mathbb{E}[v_{i0} \varepsilon_{i0} v_{ij} \varepsilon_{ij}]. \quad (6.6)$$

The variance need not factor into separate autocovariances of v and ε , because the present benchmark allows these processes to be dependent. The conditions below replace the measurement equation and the sample-mean benchmark for this section.

Assumption 6.1 (Smooth scalar moment benchmark). *All constants are uniform in i, N, T .*

(i) Sampling and model. *The unit histories are independent. The deterministic coordinates satisfy (5.1), with $b_T = C_b \sqrt{\log N/T}$ and $C_i = \sigma(\alpha_i)$. The covariates X_{it} are independent and uniform on $[0, 1]^{d_X}$, and independent of the entire (v, ε) histories. Equations (2.2)–(2.3) hold. The bounded, strictly stationary (v, ε) histories have a common law that does not vary with N, T , with $\mathbb{E}[v_{i0}] = \mathbb{E}[\varepsilon_{i0}] = \mathbb{E}[v_{i0} \varepsilon_{i0}] = 0$, $V > 0$, and $\sigma^2 > 0$. The joint process containing $X_{it}, v_{it}, \varepsilon_{it}$ and the maps $a \mapsto \partial_a^r g(W_{it}, a)$, $r = 0, 1, 2, 3$, is strictly stationary and L -dependent for a known fixed integer $L \geq 0$. For date sets whose minimum separation exceeds L , their generated σ -fields are independent. The moment may depend on X, v, ε .*

(ii) Moment and reported coordinate. *There is a known compact interval $\mathcal{A} = [-\delta, 1 + \delta]$, $\delta > 0$, such that $\mathbb{E}_i[g(W_{i0}, \alpha_i)] = 0$ and $\inf_{i, a \in \mathcal{A}} |\mathbb{E}_i[\partial_a g(W_{i0}, a)]| \geq c_G > 0$. For each unit the population derivative has a constant sign on $\mathcal{A}$. The moment and its first three derivatives are uniformly bounded there, and the third derivative is uniformly Lipschitz. Let $\check{\alpha}_i$ be the unique zero of (6.1) in $\mathcal{A}$ when exactly one zero exists, and set $\check{\alpha}_i = 1/2$ otherwise. Report $\hat{\alpha}_i = \Pi_{[b_T, 1-b_T]}(\check{\alpha}_i)$. Choose C_b so that $cC_b^2 > 16$, where c is the uniform root-tail constant in Lemma G.1.*

(iii) Spline and smoothness. *Use the normalized tensor-product B-splines of Assumption 5.1(iv), with fixed $d_W = d_X + 1$ and degree $p \geq 4$. The functions m_0, h_0 have respective smoothness $3 < \tau_m, \tau_h \leq p + 1$ on W^+ , in the extension sense of Lemma D.4: their extensions to a fixed open rectangle containing W^+ have bounded Hölder norms.*

(iv) Growth. *$N/T \rightarrow \kappa_0 \in (0, \infty)$, $K \rightarrow \infty$, $K \asymp \lambda_K^{-d_W}$, and*

$$\frac{K^{1+4/d_W}}{T} \rightarrow 0, \quad TK^{-(\tau_m + \tau_h)/d_W} \rightarrow 0. \quad (6.7)$$

The reported coordinate is selected from the observable sample moment equation on the known interval $\mathcal{A}$. Global population monotonicity ensures identification throughout that interval. Independence of X and the joint moment condition $\mathbb{E}[v_{i0} \varepsilon_{i0}] = 0$ establish the grid-projection and target-drift identities under part (i). The growth conditions are equivalent to $K \lambda_K^{-4}/T \rightarrow 0$ and $T \lambda_K^{\tau_m + \tau_h} \rightarrow 0$.

6.3 The Feasible Correction and Standard Error

Throughout the remainder of this section, hatted and tilded quantities refer to the smooth-moment procedure specified here. Its notation agrees with the sample-mean procedure, but its thresholds and rejection defaults differ. Use the candidate profile fits in (A.1) with both thresholds equal to $a_T = c_{\text{tol}} T^{-1/4}$, where $c_{\text{tol}} > 0$ is known and fixed; $c_{\text{tol}} = 1$ suffices. Accept when

$$\mathcal{O}_T = \{\lambda_{\min}(\hat{Q}_K) \geq a_T, \check{J}_K \geq a_T\}. \quad (6.8)$$

Here $\hat{Q}_K$ is the generated-coordinate Gram matrix and $\check{J}_K$ is the candidate residual second moment. On acceptance, use the candidate profile fits and slope $\hat{\theta}_K$, and set $\hat{J}_K = \check{J}_K$. The fitted score and its derivatives are $\hat{\psi}_{it,K} = \psi(W_{it}; \hat{\theta}_K, \hat{\eta}_K, \hat{\alpha}_i)$ and $\hat{\psi}_{it,K}^{(r)} = \partial_a^r \psi(W_{it}; \hat{\theta}_K, \hat{\eta}_K, \hat{\alpha}_i)$ for $r = 1, 2$. Write $\hat{\mu}_{iK} = T^{-1} \sum_t \hat{\psi}_{it,K}^{(1)}$ and $\hat{\nu}_{iK} = T^{-1} \sum_t \hat{\psi}_{it,K}^{(2)}$.

Estimate the first-step derivatives by $\hat{G}_{it} = \partial_a g(W_{it}, \hat{\alpha}_i)$. Their sample mean is $\hat{G}_i = T^{-1} \sum_t \hat{G}_{it}$. The second-derivative average is $\hat{H}_i = T^{-1} \sum_t \partial_a^2 g(W_{it}, \hat{\alpha}_i)$. Estimate G_i^{-1} by $\widehat{G}_i^{-1} = 1/\hat{G}_i$ when $|\hat{G}_i| \geq a_T$, and set $\widehat{G}_i^{-1} = 0$ otherwise. The explicitly centered first-step influence is

$$\hat{\xi}_{it} = -\widehat{G}_i^{-1} \left\{ g(W_{it}, \hat{\alpha}_i) - \frac{1}{T} \sum_s g(W_{is}, \hat{\alpha}_i) \right\}. \quad (6.9)$$

This definition applies even when root selection or clipping binds. With $I_T(j) = \{t : 1 \leq t \leq T, 1 \leq t+j \leq T\}$, estimate the first-step variance and bias by

$$\begin{aligned} \hat{\Omega}_i &= \sum_{|j| \leq L} \frac{1}{T} \sum_{t \in I_T(j)} \hat{\xi}_{it} \hat{\xi}_{i,t+j}, \\ \hat{\beta}_i &= -\widehat{G}_i^{-1} \left\{ \sum_{|j| \leq L} \frac{1}{T} \sum_{t \in I_T(j)} (\hat{G}_{it} - \hat{G}_i) \hat{\xi}_{i,t+j} + \frac{1}{2} \hat{H}_i \hat{\Omega}_i \right\}. \end{aligned} \quad (6.10)$$

Every lag statistic is divided by T , so its finite-window expectation contains $1 - |j|/T$. The three-channel correction is

$$\begin{aligned} \hat{B}_K &= \frac{1}{N} \sum_i \left\{ \sum_{|j| \leq L} \frac{1}{T} \sum_{t \in I_T(j)} \hat{\psi}_{it,K}^{(1)} \hat{\xi}_{i,t+j} + \frac{1}{2} \hat{\nu}_{iK} \hat{\Omega}_i + \hat{\mu}_{iK} \hat{\beta}_i \right\}, \\ \tilde{\theta}_K &= \hat{\theta}_K - \frac{\hat{B}_K}{T \hat{J}_K}. \end{aligned} \quad (6.11)$$

For variance estimation, let $\hat{\phi}_{it,K} = \hat{J}_K^{-1} \{\hat{\psi}_{it,K} + \hat{\mu}_{iK} \hat{\xi}_{it}\}$ and $\bar{\hat{\phi}}_{iK} = T^{-1} \sum_t \hat{\phi}_{it,K}$. Evaluate the lag statistic $\hat{q}_K$ of Section 4.2 using this smooth-moment influence, and report $\hat{\sigma}_K^2 = \max\{a_T, \hat{q}_K\}$. The coordinate contribution is retained before demeaning, and the floor keeps the reported standard error well defined when the finite-sample lag statistic is nonpositive. On $\mathcal{O}_T^c$, report $\hat{\theta}_K = \tilde{\theta}_K = 0$, $\hat{J}_K = \hat{\sigma}_K^2 = 1$, and $\hat{B}_K = 0$; set second-stage coefficients and uncalculated second-stage arrays to zero. The first-step statistics keep their own definitions. All profile normal equations are asserted only on $\mathcal{O}_T$. The standard error is $\hat{\sigma}_K / \sqrt{NT}$, with $\hat{\sigma}_K = \sqrt{\hat{\sigma}_K^2}$.

6.4 Feasible Inference

The normalized oracle influence remains unit-centered:

$$\phi_{it,K} = J_K^{-1} \{\psi_{it,K} - \mathbb{E}_i[\psi_{it,K}] + \mu_{iK} \xi_{it}\}. \quad (6.12)$$

Theorem 6.2 (Feasible inference for a smooth scalar moment coordinate). *Under Assumption 6.1, the procedure with acceptance event (6.8) satisfies $\mathbb{P}(\mathcal{O}_T) \rightarrow 1$, $\hat{B}_K - B_K = o_p(1)$, and*

$$\sqrt{NT}(\tilde{\theta}_K - \theta_0) = \frac{1}{\sqrt{NT}} \sum_{i,t} \phi_{it,K} + o_p(1) \rightsquigarrow \mathcal{N}(0, \sigma^2). \quad (6.13)$$

Moreover, $\hat{\sigma}_K^2 \xrightarrow{p} \sigma^2$ and $\sqrt{NT}(\tilde{\theta}_K - \theta_0)/\hat{\sigma}_K \rightsquigarrow \mathcal{N}(0, 1)$. The uncorrected estimator has the representation

$$\sqrt{NT}(\hat{\theta}_K - \theta_0) = \frac{1}{\sqrt{NT}} \sum_{i,t} \phi_{it,K} + \sqrt{N/T} J_K^{-1} B_K + o_p(1). \quad (6.14)$$

If additionally $B_K \rightarrow B_\infty$, its limit is $\mathcal{N}(\sqrt{\kappa_0} B_\infty / V, \sigma^2)$, with σ^2 given by (6.6). The centered conclusion requires no convergence of the bias coefficient.

Appendix G proves the theorem from the moment equation, the random-coordinate Gram bound, and the nuisance normal equations. At the $\sqrt{NT}$ scale, the quadratic remainder from profile estimation in Section 5.3 is $O_p(K^{1+4/d_W}/T) = o_p(1)$ under (6.7). Thus the proof controls shared nuisance estimation separately from the local coordinate expansion.

To compare the two sufficient sieve windows, write $d = d_W$, $s = \tau_m + \tau_h$, and $K \asymp T^q$. Table 1 expresses both results on the same dimension scale.

Table 1: Two sufficient sieve windows for $K \asymp T^q$

Coordinate and additional structure	Sufficient exponent q
Sample mean; exact linear expansion and mixed-third cancellation	$d/s < q < \min\{1, 2d/(\max\{d, 3\} + 4)\}$
Smooth scalar moment; regular unique root	$d/s < q < d/(d + 4)$

A sequence satisfying (6.7) exists exactly when $s > d + 4$. In mesh terms the first restriction is $T\lambda_K^{d+4} \rightarrow \infty$. If $s \leq d + 4$, then $T\lambda_K^s \geq T\lambda_K^{d+4}$ eventually, contradicting the second restriction. If $s > d + 4$, the interval in Table 1 supplies a sequence. For example, $d = 2$, $\tau_m = \tau_h = 4$, and $\lambda_K \asymp T^{-1/7}$ make both remainders of order $T^{-1/7}$. The fixed-degree smoothness bound also requires $d < 2p - 2$ for an admissible pair of smoothness orders to exist.

For $g(W, a) = Z - a$, $G_i = -1$, $H_i = 0$, $\xi_{it} = \zeta_{it}$, and $\beta_i = 0$. The feasible first-step bias is identically zero, and the correction reduces to two channels, up to the controlled clipping and demeaning differences. The sample-mean theorem retains the wider window (5.9) under its additional structure. Theorem 6.2 supplies a separate sufficient route with no mixed-third-moment cancellation.

7 Simulation Evidence

The simulations ask whether the two bias channels can conceal each other, whether their feasible correction restores centered inference, and whether the smooth-moment procedure works with a biased first step. The fixed-sieve experiment separates the mechanisms. The growing-sieve experiment adds nonlinear nuisance functions and measures coverage. A ratio-coordinate experiment then implements the separate smooth scalar moment procedure.

Common panel design. The first two experiments use the model in Section 2, with $\theta_0 = 1$, independent units, and X_{it} independent and uniformly distributed on $[0, 1]$. For $c \in \{\zeta, v, \varepsilon\}$,

define $q_{it}^c = (u_{it}^c + 0.5u_{i,t-1}^c)/\sqrt{1.25}$. The innovation pairs (u_{it}^ζ, u_{it}^v) are independent and identically distributed over (i, t) , with Rademacher marginals and correlation $\rho_{v\zeta}$. The Rademacher sequence u_{it}^ε is independent of the entire (u^ζ, u^v) process. Set $\zeta_{it} = 0.1q_{it}^\zeta$, $v_{it} = q_{it}^v$, and $\varepsilon_{it} = q_{it}^\varepsilon$. This construction has known dependence range one. The deterministic coordinates are equally spaced on $[2b_{NT}, 1 - 2b_{NT}]$, where $b_{NT} = \min\{0.25\sqrt{\log(N)/T}, 0.20\}$. Full-history estimators use the reported coordinate $\hat{\alpha}_i = \Pi_{[b_{NT}, 1-b_{NT}]}(\check{\alpha}_i)$.

7.1 Separation and Cancellation of the Bias Channels

The fixed-sieve experiment, labeled Layer A in the figure and replication outputs, uses nuisance functions lying in the spline span:

$$\begin{aligned} m_0(x, a) &= 0.5(x - 0.5) + \beta_m(a - 0.5), \\ h_0(x, a) &= 0.5(x - 0.5)^2 + \beta_h(a - 0.5). \end{aligned}$$

Five scenarios pair $(\beta_m, \beta_h, \rho_{v\zeta})$ with the analytic coefficients $(B_K^{(1)}, B_K^{(2)})$: null, $(5, 0, 0.5)$ with $(0, 0)$; derivative only, $(0, 5, 0.5)$ with $(-0.45, 0)$; curvature only, $(5, 5, 0)$ with $(0, 0.45)$; reinforcement, $(-5, 5, 0.5)$ with $(-0.45, -0.45)$; and cancellation, $(5, 5, 0.5)$ with $(-0.45, 0.45)$.

The cancellation scenario is the clearest demonstration that total bias alone does not diagnose reuse. In the high-precision $N = T = 200$ cell, the estimated derivative and curvature coefficients are -0.440 and 0.440 , leaving the uncorrected $\sqrt{NT}$ -scaled bias at 0.003 . Buffered splitting removes the cancellation: its scaled bias is 0.893 , compared with the doubled-curvature prediction 0.900 . In the reinforcement cell, the uncorrected bias is -0.896 , compared with the prediction -0.900 , and full correction reduces it to -0.030 . Across all five high-precision scenarios, the largest gap between the uncorrected bias and its two-channel prediction is 1.84 Monte Carlo standard errors. The largest corrected bias is 1.38 Monte Carlo standard errors from zero, and the buffered-split comparisons are within 1.55 Monte Carlo standard errors of their predictions.

This experiment uses five open-clamped, Lebesgue- L^2 -normalized degree-four B-splines per direction, giving $K = 25$. At $T = 50, 100, 200$, the respective N grids are $(12, 50, 200)$, $(25, 100, 400)$, and $(50, 200, 800)$, corresponding to nominal aspect ratios $1/4, 1$, and 4 . Each broad-grid cell uses 500 replications, and the $N = T = 200$ cells use $2,000$. Figure 1 reports the broad grid. No replication in either set activates the root or variance fallback, and the maximum clipping rate in the $2,000$ -replication cells is 0.002% .

The comparisons include the oracle-coordinate and uncorrected estimators, the oracle bias correction, each partial correction, the full correction, buffered splitting, and the split-panel jackknife (SPJ). The two split estimators are point-estimate diagnostics; we report no standard errors or coverage for them. If $\hat{\theta}_K^{(1)}$ and $\hat{\theta}_K^{(2)}$ are the profile estimates from the first and second halves, the SPJ estimate is $2\hat{\theta}_K - (\hat{\theta}_K^{(1)} + \hat{\theta}_K^{(2)})/2$. Each half reestimates the coordinate and both nuisance regressions using the full-sample spline basis and clipping interval. These halves are contiguous and have no intervening buffer.

7.2 Growing-Sieve Inference

The second experiment, labeled Layer B in the table and replication outputs, uses nonlinear nuisance functions:

$$\begin{aligned} m_0(x, a) &= 0.6 \sin(2\pi x) + 2.5 \sin(\pi a) + 1.5xa, \\ h_0(x, a) &= 0.5 \cos(2\pi x) + 4a + 2 \sin(\pi a) + 1.25 \sin(2\pi x) \sin(\pi a). \end{aligned}$$

layer_a_mechanisms.pdf

Figure 1: Layer A mechanism evidence. The broad grid uses 500 replications per cell. Panels A and C use $N = T = 200$; bars show Monte Carlo means with 95% Monte Carlo intervals and crosses show finite-sieve predictions. Panel B tracks coverage for the reported estimator in the reinforcement design at the aspect ratio nearest one; vertical bars are Wilson intervals. Panel D varies the nominal aspect ratio at $T = 200$ and compares the uncorrected $\sqrt{NT}$ -scaled bias with its predicted noncentral shift.

Here $\rho_{v\zeta} = -0.5$, so the bias channels reinforce. The (N, T) grid is unchanged. Each cell uses 1,000 replications, with $(6, 7, 8)$ normalized basis functions per direction at $T = (50, 100, 200)$, giving $K = (36, 49, 64)$. We compare the oracle-coordinate, uncorrected, two partial-correction, full two-channel, and SPJ estimators. The finite-sieve target θ_K and long-run channel coefficients are computed by 32-node Gauss–Legendre integration over x and discrete averaging over the deterministic α_i grid. Errors are measured relative to both θ_0 and θ_K . Both bias channels and the cluster variance use the uniform lag-one kernel implied by the design.

Table 2 reports the $N = T = 200$ cell. The uncorrected $\sqrt{NT}$ -scaled bias is 0.744, compared with the finite-sieve prediction 0.778, a gap of 0.94 Monte Carlo standard errors. Correcting both channels reduces the scaled bias to 0.008, which is 0.21 Monte Carlo standard errors from zero. The contribution of $\theta_K - \theta_0$ is only 0.0002 on this scale. Each partial correction leaves a substantial residual bias.

The full correction also improves inference. Its SE-to-SD ratio is 0.993, and the studentized statistic has mean 0.007, standard deviation 1.007, and empirical 2.5% and 97.5% quantiles -1.805 and 1.916 . Coverage is 95.9%, with Wilson interval $[94.5\%, 97.0\%]$, compared with 89.5% for the

Table 2: Layer B growing-sieve inference at $N = T = 200$

Method	$\sqrt{NT}$ -scaled bias	SE/SD	Stud. SD	Coverage [Wilson]
Oracle α	0.023 (0.036)	0.987	1.013	0.954 [0.939, 0.965]
Uncorrected	0.744 (0.037)	0.996	1.004	0.895 [0.874, 0.913]
Derivative correction	0.396 (0.037)	0.995	1.005	0.933 [0.916, 0.947]
Curvature correction	0.356 (0.037)	0.994	1.006	0.935 [0.918, 0.949]
Two-channel correction	0.008 (0.037)	0.993	1.007	0.959 [0.945, 0.970]
SPJ	0.047 (0.037)	—	—	—

Notes: 1,000 replications; $K = 64$. Parentheses are Monte Carlo standard errors; brackets are Wilson intervals for coverage. The split-panel jackknife is a point-estimate benchmark.

uncorrected estimator. No replication in this cell activates the root or variance fallback, and clipping never occurs.

Bias correction is effective with fewer clusters as well, although coverage is less accurate. At $T = 200$ and $N = 50$ ($N/T = 1/4$), the corrected scaled bias is 0.002 and coverage is 93.4%, with Wilson interval [91.7%, 94.8%]. The SE-to-SD ratio is 0.952, and the studentized standard deviation is 1.051. Small estimated standard errors therefore accompany the remaining undercoverage. These summaries do not separate variance-estimation error from Gaussian-approximation error with 50 independent clusters. Appendix B, Table B.1, reports all nine design cells.

Both panel experiments use the sample-mean benchmark, including its mixed third-moment restriction. The ratio experiment below uses the smooth scalar moment procedure and permits a nonzero corresponding mixed third moment. All experiments retain deterministic coordinates and a supplied dependence bound.

7.3 A Ratio Coordinate

The final experiment implements the complete procedure in Section 6. There are no X covariates, so $d_W = 1$. Let S_{it} , Z_{it}^* , and E_{it} be independent Rademacher variables across all indices and components, and set

$$\begin{aligned} R_{it} &= 1 + \frac{1}{4}S_{it}, & U_{it} &= \alpha_i R_{it} + 0.1\{\frac{1}{4}S_{it} + R_{it}Z_{it}^*\}, \\ D_{it} &= 10(\alpha_i - \frac{1}{2}) + S_{it}, & Y_{it} &= D_{it} + 10(\alpha_i - \frac{1}{2}) + E_{it}. \end{aligned}$$

The target is $\theta_0 = 1$. The known first-step moment is $g((U, R), a) = U - aR$ on $[-2, 3]$. Because $R_{it} \geq 3/4$, its unique sample root is $\check{\alpha}_i = \frac{\sum_t U_{it}}{\sum_t R_{it}}$. The coordinates are equally spaced on $[2b_T, 1 - 2b_T]$, with $b_T = \sqrt{\log N/T}$, and the reported coordinate clips this root to $[b_T, 1 - b_T]$. The numerator error has range $1/4$. Hoeffding's inequality and $T^{-1} \sum_t R_{it} \geq 3/4$ give $\mathbb{P}_i(|\check{\alpha}_i - \alpha_i| > x) \leq 2 \exp(-18Tx^2)$. Thus the clipping constant meets $cC_b^2 > 16$ in Assumption 6.1.

We use normalized degree-four B-splines with $\lceil T^{1/7} \rceil$ equal positive knot spans, known lag bound $L = 0$, and threshold constant $c_{\text{tol}} = 10^{-8}$. At $N = T = 500$ and 1000 , this gives $K = 7$; each cell uses 1,000 replications. The nuisance functions lie exactly in these spaces, so $\theta_K = \theta_0$. The two sample sizes hold K fixed and introduce no sieve approximation error.

The first-step bias is nonzero, but its population score contribution is zero. Direct enumeration gives $\beta_i = -0.00625$, $\Omega_i = 0.01125$, and $\mathbb{E}[S_{it}\xi_{it}^2] = 0.005$. The score derivatives are $\psi_{it,K}^{(1)} = -10(E_{it} + S_{it})$ and $\psi_{it,K}^{(2)} = 200$, so $\mu_{iK} = 0$. Consequently, $B_K^{(1)} = -0.25$, $B_K^{(2)} = 1.125$, and $B_K^{(3)} = 0$. The predicted scaled root shift is 0.875, with $J_K = \sigma^2 = 1$. This design separates first-step bias from its propagation to the target score, while allowing a nonzero mixed third moment.

Table 3: Feasible inference with a ratio coordinate

$N = T$	Method	$\sqrt{NT}$ -scaled bias	SE/SD	Coverage [Wilson]
500	Uncorrected	0.830 (0.032)	0.992	0.863 [0.840, 0.883]
	Full correction	−0.029 (0.032)	0.990	0.950 [0.935, 0.962]
1000	Uncorrected	0.859 (0.032)	1.000	0.862 [0.839, 0.882]
	Full correction	−0.010 (0.032)	1.000	0.963 [0.949, 0.973]

Notes: 1,000 replications per cell, seed 20260913, $K = 7$. Parentheses report Monte Carlo standard errors of scaled bias; brackets report 95% Wilson intervals for coverage of nominal 95% intervals. SE/SD is the mean estimated standard error divided by the empirical standard deviation of the corresponding estimator. All statistics include the specified rejection rules. The first-step Jacobian, bias channels, and variance are estimated from the simulated observations.

Table 3 compares uncorrected and fully corrected inference using the same feasible standard error, which includes the first-step influence. At $N = T = 1000$, correction reduces scaled bias from 0.859 to −0.010, which is 0.32 Monte Carlo standard errors from zero. The estimated channel coefficients average −0.249 for $B_K^{(1)}$, 1.118 for $B_K^{(2)}$, and −0.00014 for $B_K^{(3)}$. The average estimate of β_i is −0.006248. Coverage rises from 86.2% to 96.3%, and the corrected Wilson interval contains 95%. Neither sample size activates root fallback, clipping, second-stage fallback, or the variance floor. This experiment checks the feasible correction and variance estimator for a first step with a random Jacobian; its first-step bias has zero population score multiplier. The growing-sieve comparison with approximation error is supplied by the preceding experiment.

8 Conclusion

A coordinate estimated from its own unit history can change both the centering and the sampling variation of an orthogonal panel score. Its order- T^{-1} drift can remain first order at the panel rate $\sqrt{NT}$. Derivative-influence covariance and curvature explain the two surviving channels in the partially linear model; the propagation of first-step bias vanishes as the nuisance sieve grows. The channels can cancel, so an apparently unbiased uncorrected estimator can still respond sharply to splitting. For a sample mean and a fixed sieve, buffered splitting removes the covariance channel and retains curvature, with twice the leading curvature coefficient under an equal split.

Estimating these channels from the full history restores centered inference. The sample-mean and smooth scalar moment results use separate procedures and sufficient sieve windows. The ratio experiment shows that a biased first step can have zero population contribution to the profiled score, while the other two channels still require correction. Extending the same analysis to random coordinate designs, unknown dependence ranges, or overidentified local moments would require corresponding control of coordinate estimation, lag truncation, and the quadratic profile remainder.

References

- [BJNW26] Stéphane Bonhomme, Koen Jochmans, Whitney K. Newey, and Martin Weidner, *Higher-order Neyman orthogonality in moment-condition models*, 2026, arXiv:2605.10842.
- [BJW26] Stéphane Bonhomme, Koen Jochmans, and Martin Weidner, *A Neyman-orthogonalization approach to the incidental parameter problem in likelihood models*, 2026, arXiv:2412.10304; originally circulated in 2024; revised August 2026.

- [BKRW93] Peter J. Bickel, Chris A. J. Klaassen, Ya'acov Ritov, and Jon A. Wellner, *Efficient and adaptive estimation for semiparametric models*, Johns Hopkins Series in the Mathematical Sciences, Johns Hopkins University Press, Baltimore, 1993.
- [CCD⁺18] Victor Chernozhukov, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins, *Double/debiased machine learning for treatment and structural parameters*, The Econometrics Journal **21** (2018), no. 1, C1–C68.
- [CCJ13] Matias D. Cattaneo, Richard K. Crump, and Michael Jansson, *Generalized jackknife estimators of weighted average derivatives*, Journal of the American Statistical Association **108** (2013), no. 504, 1243–1256.
- [CEI⁺22] Victor Chernozhukov, Juan Carlos Escanciano, Hidehiko Ichimura, Whitney K. Newey, and James M. Robins, *Locally robust semiparametric estimation*, Econometrica **90** (2022), no. 4, 1501–1535.
- [CJM19] Matias D. Cattaneo, Michael Jansson, and Xinwei Ma, *Two-step estimation and inference with possibly many included covariates*, The Review of Economic Studies **86** (2019), no. 3, 1095–1122.
- [CKMS22] Harold D. Chiang, Kengo Kato, Yukun Ma, and Yuya Sasaki, *Multiway cluster robust double/debiased machine learning*, Journal of Business & Economic Statistics **40** (2022), no. 3, 1046–1056.
- [CP26] Paul S. Clarke and Annalivia Polselli, *Double machine learning for static panel models with fixed effects*, The Econometrics Journal **29** (2026), no. 1, 69–86.
- [dB01] Carl de Boor, *A practical guide to splines*, revised ed., Springer, New York, 2001.
- [EPI26] Juan Carlos Escanciano and Telmo Pérez-Izquierdo, *Automatic locally robust GMM with machine-learning-generated regressors*, 2026, arXiv:2301.10643; version 4, revised March 2026.
- [FK26] Hugo Freeman and Dennis Kristensen, *Inference on linear regressions with two-way unobserved heterogeneity*, 2026, arXiv:2605.06491.
- [FVL13] Iván Fernández-Val and Joonhwhah Lee, *Panel data models with nonadditive unobserved heterogeneity: Estimation and inference*, Quantitative Economics **4** (2013), no. 3, 453–481.
- [FVV11] Iván Fernández-Val and Francis Vella, *Bias corrections for two-step fixed effects panel data estimators*, Journal of Econometrics **163** (2011), no. 2, 144–162.
- [FVW18] Iván Fernández-Val and Martin Weidner, *Fixed effects estimation of large- T panel data models*, Annual Review of Economics **10** (2018), 109–138.
- [HHKN24] Jinyong Hahn, David W. Hughes, Guido Kuersteiner, and Whitney K. Newey, *Efficient bias correction for cross-section and panel data*, Quantitative Economics **15** (2024), no. 3, 783–816.
- [HK11] Jinyong Hahn and Guido Kuersteiner, *Bias reduction for dynamic nonlinear panel models with fixed effects*, Econometric Theory **27** (2011), no. 6, 1152–1191.

- [HN04] Jinyong Hahn and Whitney Newey, *Jackknife and analytical bias reduction for nonlinear panel models*, *Econometrica* **72** (2004), no. 4, 1295–1319.
- [HR13] Jinyong Hahn and Geert Ridder, *Asymptotic variance of semiparametric estimators with generated regressors*, *Econometrica* **81** (2013), no. 1, 315–340.
- [Hua26] Jiaqi Huang, *Fixed effects as generated regressors*, 2026, arXiv:2602.08899.
- [JW24] Koen Jochmans and Martin Weidner, *Inference on a distribution from noisy draws*, *Econometric Theory* **40** (2024), no. 1, 60–97.
- [MRS16] Enno Mammen, Christoph Rothe, and Melanie Schienle, *Semiparametric estimation with generated covariates*, *Econometric Theory* **32** (2016), no. 5, 1140–1177.
- [MSZ18] Lester Mackey, Vasilis Syrgkanis, and Ilias Zadik, *Orthogonal machine learning: Power and limitations*, *Proceedings of the 35th International Conference on Machine Learning* (Jennifer Dy and Andreas Krause, eds.), *Proceedings of Machine Learning Research*, vol. 80, PMLR, 2018, pp. 3375–3383.
- [MT85] Kevin M. Murphy and Robert H. Topel, *Estimation and inference in two-step econometric models*, *Journal of Business & Economic Statistics* **3** (1985), no. 4, 370–379.
- [New84] Whitney K. Newey, *A method of moments interpretation of sequential estimators*, *Economics Letters* **14** (1984), no. 2–3, 201–206.
- [New94] ———, *The asymptotic variance of semiparametric estimators*, *Econometrica* **62** (1994), no. 6, 1349–1382.
- [Ney59] Jerzy Neyman, *Optimal asymptotic tests of composite statistical hypotheses*, *Probability and Statistics: The Harald Cramér Volume* (Ulf Grenander, ed.), Almqvist & Wiksell, Uppsala, 1959, pp. 213–234.
- [Pag84] Adrian Pagan, *Econometric issues in the analysis of regressions with generated regressors*, *International Economic Review* **25** (1984), no. 1, 221–247.
- [RLTvdV08] James M. Robins, Lingling Li, Eric J. Tchetgen Tchetgen, and Aad W. van der Vaart, *Higher order influence functions and minimax estimation of nonlinear functionals*, *Probability and Statistics: Essays in Honor of David A. Freedman* (Deborah Nolan and Terry Speed, eds.), *IMS Collections*, vol. 2, Institute of Mathematical Statistics, 2008, pp. 335–421.
- [Sch07] Larry L. Schumaker, *Spline functions: Basic theory*, third ed., Cambridge University Press, Cambridge, 2007.
- [Tro12] Joel A. Tropp, *User-friendly tail bounds for sums of random matrices*, *Foundations of Computational Mathematics* **12** (2012), no. 4, 389–434.
- [vdV14] Aad W. van der Vaart, *Higher order tangent spaces and influence functions*, *Statistical Science* **29** (2014), no. 4, 679–686.

Appendix roadmap. Appendix A collects the coefficient representation, observable reporting rules, and exact profile identities. Appendix B reports the complete simulation grid. Appendix C develops the fixed-sieve bias mechanism and sampling expansion, including its smooth-coordinate and split-sample extensions. Appendix D proceeds from deterministic spline properties and coordinate clipping to nuisance coefficient rates, Gram-matrix bounds, and estimator localization. Appendix E derives the scalar score expansion, proves consistency of the bias correction and variance estimator, and identifies the primitive bias limits. Appendix F combines these bounds in one proof of the sample-mean growing-sieve theorem. Appendix G proves the observable-root expansion, uniform Gram bound, nuisance rate, feasible correction, and studentization for smooth scalar moment coordinates.

A Coefficient Representation, Implementation, and Supporting Identities

The main text and appendix use explicit unit and time sums. Unless a range is specified, unit sums run from $i = 1$ to N and time sums from $t = 1$ to T ; $\sum_{i,t}$ includes both indices. Write $b_K = \theta_K - \theta_0$ for the finite-sieve target drift. Under the conditional-law convention, persistent unit coordinates and population coefficients are held fixed, while a coordinate estimated from a unit history is integrated jointly with that unit's score.

For the coefficient calculations, use the unnormalized influence $x_{it,K} = J_K \phi_{it,K}$. Hatted quantities keep the reporting rules of their respective sample-mean or smooth-moment procedure. For an observation array Q , write $\|Q\|_{NT,2}^2 = \frac{1}{NT} \sum_{i,t} \|Q_{it}\|^2$, using the Euclidean norm for vector arrays. For proof-local time averages, write $\bar{\psi}_{iK}(a) = T^{-1} \sum_t \psi_{it,K}(a)$ and $\bar{\psi}_{iK}^{(r)}(a) = T^{-1} \sum_t \psi_{it,K}^{(r)}(a)$, with $r = 1, 2, 3$. An omitted argument means evaluation at α_i ; the parenthesized superscript records a derivative order.

The exact clipped-coordinate drift coefficient $\mathcal{B}_{KT}$ is defined in (A.6). Appendix C.2 treats the raw sample mean directly. Appendix C.4 introduces the oracle local root and its sample- and population-Hessian expansions; Appendix G applies them to the observable smooth-moment root.

A.1 Stacked Moments and the Plug-In Estimator

For $\vartheta = (\theta, c^\top)^\top$ with $c = (\gamma^\top, \delta^\top)^\top$, set $u_{D,it}(\gamma, a) = D_{it} - p_K(X_{it}, a)^\top \gamma$ and $u_{Y,it}(\delta, a) = Y_{it} - p_K(X_{it}, a)^\top \delta$. Write $c_K = (\gamma_K^\top, \delta_K^\top)^\top$ and $\vartheta_K = (\theta_K, c_K^\top)^\top$. The scalar coefficient score is $s_{it,K}(\vartheta, a) = \psi(W_{it}; \theta, (p_K^\top \gamma, p_K^\top \delta), a)$; in particular $s_{it,K}(\vartheta_K, a) = \psi_{it,K}(a)$. The full $(1 + 2K)$ -dimensional moment vector is

$$\mathbf{f}_{it,K}(\vartheta, a) = \begin{pmatrix} s_{it,K}(\vartheta, a) \\ p_K(X_{it}, a) u_{D,it}(\gamma, a) \\ p_K(X_{it}, a) u_{Y,it}(\delta, a) \end{pmatrix}.$$

The last two blocks identify the grid projections. At the population coefficients, their normal equations give the nuisance orthogonality

$$\begin{aligned} \partial_\gamma \left\{ \frac{1}{NT} \sum_{i,t} \mathbb{E}_i[s_{it,K}(\vartheta, \alpha_i)] \right\} \Big|_{\vartheta=\vartheta_K} &= \frac{1}{NT} \sum_{i,t} \mathbb{E}_i[p_K(X_{it}, \alpha_i) \{-u_{Y,it,K} + 2\theta_K u_{D,it,K}\}] = 0, \\ \partial_\delta \left\{ \frac{1}{NT} \sum_{i,t} \mathbb{E}_i[s_{it,K}(\vartheta, \alpha_i)] \right\} \Big|_{\vartheta=\vartheta_K} &= -\frac{1}{NT} \sum_{i,t} \mathbb{E}_i[p_K(X_{it}, \alpha_i) u_{D,it,K}] = 0. \end{aligned}$$

Here every basis and residual is evaluated at the true coordinate. The sample counterparts are also the normal equations that produce the exact profile cancellation used in Section 5. For the fixed- K analysis below, let $\widehat{\vartheta}_K$ denote the globally measurable extension defined by the reporting rule in Appendix A.2. On the acceptance event $\mathcal{G}_K$ it is the profiled solution to $\frac{1}{NT} \sum_{i,t} \mathbf{f}_{it,K}(\widehat{\vartheta}_K, \check{\alpha}_i) = 0$; off that event it equals a fixed deterministic vector. Let $\widehat{\theta}_K$ be its first coordinate. Section 4 uses the same construction at the clipped coordinate for the growing-sieve analysis.

For the coefficient derivatives, use $p_{a,it,K}(a) = \partial_a p_K(X_{it}, a)$, $p_{aa,it,K}(a) = \partial_a^2 p_K(X_{it}, a)$, and $q_{it}(\vartheta, a) = u_{Y,it}(\delta, a) - \theta u_{D,it}(\gamma, a)$. Differentiating the scalar score gives

$$\begin{aligned} s_{it,K}^{(1)}(\vartheta, a) &= -\{p_{a,it,K}(a)^\top \gamma\} q_{it}(\vartheta, a) - u_{D,it}(\gamma, a) p_{a,it,K}(a)^\top (\delta - \theta \gamma), \\ s_{it,K}^{(2)}(\vartheta, a) &= -\{p_{aa,it,K}(a)^\top \gamma\} q_{it}(\vartheta, a) - u_{D,it}(\gamma, a) p_{aa,it,K}(a)^\top (\delta - \theta \gamma) \\ &\quad + 2\{p_{a,it,K}(a)^\top \gamma\} \{p_{a,it,K}(a)^\top (\delta - \theta \gamma)\}. \end{aligned}$$

The first and second derivatives generate the derivative-influence and curvature channels, respectively.

A.2 Profile Estimation and Acceptance

Both coordinate constructions use the same candidate profile regression. A check denotes candidate profile coefficients, residuals, and their second moment; their reported versions carry a hat. At generated coordinates $\widehat{\alpha}_i$, set $\widehat{p}_{it,K} = p_K(X_{it}, \widehat{\alpha}_i)$ and $\widehat{Q}_K = \frac{1}{NT} \sum_{i,t} \widehat{p}_{it,K} \widehat{p}_{it,K}^\top$. For positive thresholds $c_{\Sigma,T}, c_{V,T}$, compute

$$\begin{aligned} \check{\gamma}_K &= \widehat{Q}_K^{-1} \frac{1}{NT} \sum_{i,t} \widehat{p}_{it,K} D_{it}, & \check{\delta}_K &= \widehat{Q}_K^{-1} \frac{1}{NT} \sum_{i,t} \widehat{p}_{it,K} Y_{it}, \\ \check{u}_{D,it} &= D_{it} - \widehat{p}_{it,K}^\top \check{\gamma}_K, & \check{u}_{Y,it} &= Y_{it} - \widehat{p}_{it,K}^\top \check{\delta}_K, \\ \check{J}_K &= \frac{1}{NT} \sum_{i,t} (\check{u}_{D,it})^2 \end{aligned} \tag{A.1}$$

when $\lambda_{\min}(\widehat{Q}_K) \geq c_{\Sigma,T}$. On failure of this Gram test, set the candidate coefficients to fixed deterministic defaults, define the candidate residuals from those coefficients, and set $\check{J}_K = 0$. Thus the last equality in (A.1) is asserted when the Gram test passes. Accept the candidate when this Gram test and $\check{J}_K \geq c_{V,T}$ both hold. On acceptance, report the candidate coefficients and $\widehat{\theta}_K = (\check{J}_K)^{-1} \frac{1}{NT} \sum_{i,t} \check{u}_{D,it} \check{u}_{Y,it}$.

The sample-mean procedure uses fixed thresholds $c_{\Sigma,T} = c_\Sigma$ and $c_{V,T} = c_V/2$. For the growing-sieve benchmark choose (A.2); for the raw fixed- K result use the thresholds in Assumption C.2. Its acceptance event is $\mathcal{G}_K$. On rejection, report a fixed deterministic vector $\vartheta_K^\dagger = (\theta_K^\dagger, \gamma_K^{\dagger\top}, \delta_K^{\dagger\top})^\top$. Using the reported coefficients on the whole sample space, define $\widehat{u}_{D,it,K} = D_{it} - \widehat{p}_{it,K}^\top \widehat{\gamma}_K$, $\widehat{u}_{Y,it,K} = Y_{it} - \widehat{p}_{it,K}^\top \widehat{\delta}_K$, and $\widehat{J}_K = \frac{1}{NT} \sum_{i,t} \widehat{u}_{D,it,K}^2$. For the coefficient calculations, let $\mathcal{H}_K = \{\lambda_{\min}(\widehat{Q}_K) \geq c_\Sigma\}$. Then $\mathcal{G}_K = \mathcal{H}_K \cap \{\check{J}_K \geq c_V/2\}$. When the Gram test passes but the second-moment test fails, the checked quantities retain their candidate values; the reported hats use the rejection vector. The correction and variance formulas below apply on $\mathcal{G}_K$ and receive their stated defaults otherwise. Appendix D.5 proves that acceptance has probability approaching one. The smooth moment procedure in Section 6 instead uses declining thresholds and its own rejection defaults.

Tuning convention. For bias correction and variance estimation, retain the integer lags $|j| \leq L$, each with weight one. All other lags have weight zero. The supplied upper bound L is held fixed; the panel simulations use the known MA(1) dependence range.

Let $c_{\text{gr}} > 0$ denote the uniform lower eigenvalue constant for the population grid Gram established in Lemma D.8. Choose the deterministic estimator thresholds so that

$$0 < c_{\Sigma} < c_{\text{gr}}/2, \quad 0 < c_V < \sigma_v^2. \quad (\text{A.2})$$

A.3 Sieve Approximation Bias

Write $r_{m,it,K} = m_0(X_{it}, \alpha_i) - m_K(X_{it}, \alpha_i)$ and $r_{h,it,K} = h_0(X_{it}, \alpha_i) - h_K(X_{it}, \alpha_i)$; when indices are suppressed, r_m, r_h denote these projection errors. The following argument proves (2.6) and its projection-error bound.

Proof. Substitute $u_D = v + r_m$ and $u_Y = \theta_0 u_D + r_h + \varepsilon$ into the target numerator:

$$\begin{aligned} J_K b_K &= \frac{1}{NT} \sum_{i,t} \mathbb{E}_i[u_{D,it,K}(u_{Y,it,K} - \theta_0 u_{D,it,K})] \\ &= \frac{1}{NT} \sum_{i,t} \mathbb{E}_i[(v_{it} + r_{m,it,K})(r_{h,it,K} + \varepsilon_{it})] \\ &= \frac{1}{NT} \sum_{i,t} \mathbb{E}_i[r_{m,it,K} r_{h,it,K}] \\ &\quad + \frac{1}{NT} \sum_{i,t} \mathbb{E}_i[v_{it} r_{h,it,K} + r_{m,it,K} \varepsilon_{it} + v_{it} \varepsilon_{it}]. \end{aligned}$$

Under the conditional-law convention in Section 2.1, the projection targets are held fixed and r_m, r_h are functions of $(X_{it}, \mathcal{C}_i)$. Hence

$$\begin{aligned} \mathbb{E}_i[v_{it} r_{h,it,K}] &= \mathbb{E}_i[r_{h,it,K} \mathbb{E}[v_{it} \mid X_{it}, \mathcal{C}_i]] = 0, \\ \mathbb{E}_i[r_{m,it,K} \varepsilon_{it}] &= \mathbb{E}_i[r_{m,it,K} \mathbb{E}[\varepsilon_{it} \mid X_{it}, \mathcal{C}_i]] = 0. \end{aligned}$$

The remaining cross term vanishes by $\mathbb{E}_i[v_{it} \varepsilon_{it}] = 0$. Dividing by J_K proves (2.6). The same orthogonality gives

$$\begin{aligned} J_K &= \frac{1}{NT} \sum_{i,t} \mathbb{E}_i[v_{it}^2 + 2v_{it} r_{m,it,K} + r_{m,it,K}^2] \\ &= \frac{1}{NT} \sum_{i,t} \mathbb{E}_i[v_{it}^2] + \frac{1}{NT} \sum_{i,t} \mathbb{E}_i[r_{m,it,K}^2], \\ |b_K| &\leq J_K^{-1} \left\{ \frac{1}{NT} \sum_{i,t} \mathbb{E}_i[r_{m,it,K}^2] \right\}^{1/2} \left\{ \frac{1}{NT} \sum_{i,t} \mathbb{E}_i[r_{h,it,K}^2] \right\}^{1/2} \\ &\leq C \lambda_K^{\tau_m + \tau_h}. \end{aligned}$$

Using $K \asymp \lambda_K^{-d_w}$ gives the dimension-rate bound in Section 2.2. $\square$

Lemma A.1 (Centered fixed-sieve sampling expansion). *Suppose K is fixed, Assumption 3.1 holds, and $N, T \rightarrow \infty$ with $N/T \rightarrow \kappa \in (0, \infty)$. Then*

$$\frac{1}{NT} \sum_{i,t} \{\psi_{it,K}(\check{\alpha}_i) - \mathbb{E}_i[\psi_{it,K}(\check{\alpha}_i)]\} = \frac{1}{NT} \sum_{i,t} x_{it,K} + o_p((NT)^{-1/2}). \quad (\text{A.3})$$

A.4 Exact Profile and Correction Identities

Let $e_\gamma = \hat{\gamma}_K - \gamma_K$, $e_\delta = \hat{\delta}_K - \delta_K$, and $e_\pi = e_\delta - \theta_K e_\gamma$.

Lemma A.2 (Exact profile identity). *For any generated-coordinate array, on an event where the profile normal equations hold and $\hat{J}_K > 0$,*

$$\frac{1}{NT} \sum_{i,t} \psi_{it,K}(\hat{\alpha}_i) = \hat{J}_K(\hat{\theta}_K - \theta_K) + e_\gamma^\top \hat{Q}_K e_\pi. \quad (\text{A.4})$$

For the sample-mean procedure this event is $\mathcal{G}_K$; for the smooth moment procedure it is $\mathcal{O}_T$, defined in (6.8). Multiplication by its indicator gives the corresponding unconditional identity.

Proof. We first express the target-coefficient residuals in terms of the estimated residuals and coefficient errors. We then derive the four normal-equation cancellations available on the event in the lemma. Expanding the target score produces six products; those cancellations remove four of them and leave the denominator term and the quadratic profile term in the lemma.

Define the target-coefficient residuals at the generated coordinate, $\tilde{u}_{D,it,K} = D_{it} - \hat{p}_{it,K}^\top \gamma_K$ and $\tilde{u}_{Y,it,K} = Y_{it} - \hat{p}_{it,K}^\top \delta_K$. They satisfy

$$\begin{aligned} \tilde{u}_{D,it,K} &= \hat{u}_{D,it,K} + \hat{p}_{it,K}^\top e_\gamma, \\ \tilde{u}_{Y,it,K} - \theta_K \tilde{u}_{D,it,K} &= \hat{q}_{it,K} + (\hat{\theta}_K - \theta_K) \hat{u}_{D,it,K} + \hat{p}_{it,K}^\top e_\pi, \end{aligned} \quad (\text{A.5})$$

where $\hat{q}_{it,K} = \hat{u}_{Y,it,K} - \hat{\theta}_K \hat{u}_{D,it,K}$.

The profile normal equations imply

$$\frac{1}{NT} \sum_{i,t} \hat{p}_{it,K} \hat{u}_{D,it,K} = 0, \quad \frac{1}{NT} \sum_{i,t} \hat{p}_{it,K} \hat{u}_{Y,it,K} = 0, \quad \frac{1}{NT} \sum_{i,t} \hat{u}_{D,it,K} \hat{q}_{it,K} = 0.$$

The first two equations also imply

$$\frac{1}{NT} \sum_{i,t} \hat{p}_{it,K} \hat{q}_{it,K} = \frac{1}{NT} \sum_{i,t} \hat{p}_{it,K} \hat{u}_{Y,it,K} - \hat{\theta}_K \frac{1}{NT} \sum_{i,t} \hat{p}_{it,K} \hat{u}_{D,it,K} = 0.$$

The target score is $\frac{1}{NT} \sum_{i,t} \tilde{u}_{D,it,K} (\tilde{u}_{Y,it,K} - \theta_K \tilde{u}_{D,it,K})$. Substituting (A.5) and expanding every product gives

$$\begin{aligned} \frac{1}{NT} \sum_{i,t} \psi_{it,K}(\hat{\alpha}_i) &= \frac{1}{NT} \sum_{i,t} \hat{u}_{D,it,K} \hat{q}_{it,K} + (\hat{\theta}_K - \theta_K) \frac{1}{NT} \sum_{i,t} \hat{u}_{D,it,K}^2 + e_\pi^\top \frac{1}{NT} \sum_{i,t} \hat{p}_{it,K} \hat{u}_{D,it,K} \\ &\quad + e_\gamma^\top \frac{1}{NT} \sum_{i,t} \hat{p}_{it,K} \hat{q}_{it,K} + (\hat{\theta}_K - \theta_K) e_\gamma^\top \frac{1}{NT} \sum_{i,t} \hat{p}_{it,K} \hat{u}_{D,it,K} \\ &\quad + e_\gamma^\top \left(\frac{1}{NT} \sum_{i,t} \hat{p}_{it,K} \hat{p}_{it,K}^\top \right) e_\pi. \end{aligned}$$

The first, third, fourth, and fifth terms are zero by the displayed normal equations. The second term equals $\hat{J}_K(\hat{\theta}_K - \theta_K)$, and the sixth equals $e_\gamma^\top \hat{Q}_K e_\pi$. These are exactly the two terms in (A.4). $\square$

For the clipped coordinate, define

$$\mathcal{B}_{KT} = \frac{1}{N} \sum_{i,t} \mathbb{E}_i[\psi_{it,K}(\hat{\alpha}_i) - \psi_{it,K}(\alpha_i)]. \quad (\text{A.6})$$

Appendix E.1 proves the exact score expansion

$$\frac{1}{NT} \sum_{i,t} \psi_{it,K}(\hat{\alpha}_i) = \frac{1}{NT} \sum_{i,t} x_{it,K} + \frac{1}{T} \mathcal{B}_{KT} + r_{\theta,F,K}, \quad (\text{A.7})$$

where $\sqrt{NT}|r_{\theta,F,K}| = o_p(1)$ under the rate window below.

To expose each term, substitute the score expansion (A.7) into the left side of the profile identity (A.4), solve for $\hat{\theta}_K - \theta_K$, and then subtract the correction in (4.1). On $\mathcal{G}_K$, these three operations give

$$\begin{aligned} \tilde{\theta}_K - \theta_K &= \hat{J}_K^{-1} \frac{1}{NT} \sum_{i,t} x_{it,K} + \hat{J}_K^{-1} r_{\theta,F,K} \\ &\quad - \hat{J}_K^{-1} e_\gamma^\top \hat{Q}_K e_\pi + \frac{\hat{J}_K^{-1}}{T} (\mathcal{B}_{KT} - \hat{B}_K). \end{aligned} \quad (\text{A.8})$$

The four terms are sampling variation, the score-expansion remainder, the profile remainder, and correction error. Under the rate window below, the last three terms are $o_p((NT)^{-1/2})$, whereas the sampling term satisfies a cluster central limit theorem.

The exact finite- T coefficient $\mathcal{B}_{KT}$, its long-run approximation B_K , and the feasible estimate $\hat{B}_K$ enter separately. Appendix E.2 establishes their asymptotic equivalence.

A.5 A Finite Moving-Average Benchmark

Example A.3 (Finite moving averages, including MA(1)). Let $\{U_{it} = (u_{it}^\zeta, u_{it}^v)^\top\}$ be a bounded, centered array that is independent and identically distributed over (i, t) , and let $\{u_{it}^\varepsilon\}$ be a bounded, centered array that is independent and identically distributed over (i, t) and independent of the entire U process. For fixed $q_c \leq L$, coefficients $\rho_{c,0} = 1$, and $c \in \{\zeta, v, \varepsilon\}$, set $c_{it} = \sum_{j=0}^{q_c} \rho_{c,j} u_{i,t-j}^c$. If $\mathbb{E}[u_{it}^v (u_{it}^\zeta)^2] = 0$, then (5.2) holds. Moreover,

$$\gamma_{v\zeta}(j) = \mathbb{E}[u_{it}^v u_{it}^\zeta] \sum_{k \in \mathbb{Z}} \rho_{v,k} \rho_{\zeta,k+j},$$

where coefficients outside their filter ranges are zero. Taking $q_\zeta, q_v, q_\varepsilon \leq 1$ recovers the bounded joint MA(1) benchmark.

A.6 When the Derivative-Influence Channel Is Nonzero

For the primitive channel proofs, write $\gamma_{v\zeta}(j) = \mathbb{E}[v_{i0} \zeta_{ij}]$ and $\gamma_\zeta(j) = \mathbb{E}[\zeta_{i0} \zeta_{ij}]$. Because $\mathbb{E}_i[\zeta_{ij}] = 0$, the covariance channel is unchanged if $\psi_{i0,K}^{(1)}$ is replaced by $\psi_{i0,K}^{(1)} - \mu_{iK}$. Define the local loading $\bar{a}_{v,K} = N^{-1} \sum_i \mathbb{E}_X[2b_K \partial_a m_K(X, \alpha_i) - \partial_a h_K(X, \alpha_i)]$ and boundary contrast $\mathbf{b}_h = \int_{[0,1]^{d_X}} \{h_0(x, 1) - h_0(x, 0)\} dx$. Here $\mathbb{E}_X$ integrates over the uniform covariate law. The exact covariance loading and its boundary limit are

$$B_K^{(1)} = \Omega_{v\zeta} \bar{a}_{v,K}, \quad \bar{a}_{v,K} \longrightarrow -\mathbf{b}_h.$$

The proof is in Appendix E.5.

For the curvature proof, locally write $\mathbf{c}_{mh} = \int_{W^+} \partial_a m_0(x, a) \partial_a h_0(x, a) dx da$.

A.7 Supporting Rate Conditions

The channel limits identify the centering error. We next collect the rate conditions used to control nuisance estimation, the feasible correction, and studentization at the $\sqrt{NT}$ scale. Let

$m = \max\{d_W, 3\}$ and $r_K = \lambda_K^{-m/2}/T$. The sufficient remainder conditions are

$$\lambda_K^{-m}/T \rightarrow 0, \quad \lambda_K^{-2}r_K \rightarrow 0, \quad K_x \log(NK_x)/T \rightarrow 0. \quad (\text{A.9})$$

Let $\gamma_c(j) = \mathbb{E}[c_{i0}c_{ij}]$ for $c \in \{v, \varepsilon\}$, let $\Phi_{iT,K} = T^{-1/2} \sum_{t=1}^T \phi_{it,K}$, and define $\sigma^2 = \frac{\sum_{|j| \leq L} \gamma_v(j)\gamma_\varepsilon(j)}{\gamma_v(0)^2}$.

Under (A.9), the appendices establish that acceptance has probability tending to one, together with the nuisance and profile-remainder rates, consistency of both bias channels, and feasible variance consistency. Along growing-sieve sequences with $\lambda_K \rightarrow 0$, they also show that the residual denominator and variance target converge to strictly positive limits and prove the cluster central limit theorem. Appendix F.1 combines these bounds in the proof of Theorem 5.3.

B Additional Simulation Results

Table B.1 reports inference for the corrected estimator in every Layer B design cell. The $\sqrt{NT}$ -scaled bias is small by $T = 200$ at all three aspect ratios. Coverage is weakest in designs with few independent clusters and is otherwise close to nominal. In the $T = 200$, $N = 50$ cell, the corrected $\sqrt{NT}$ -scaled bias is 0.002 but coverage is 93.4%. The SE-to-SD ratio of 0.952 and studentized standard deviation of 1.051 show that small estimated standard errors accompany the undercoverage. These summaries do not identify the separate contributions of variance estimation and the Gaussian approximation with 50 independent clusters.

Table B.1: Layer B corrected inference across the full design grid

T	N/T	N	K	$\sqrt{NT}$ -scaled bias	SE/SD	Stud. SD	Coverage [Wilson]	Defaults (root/var.)
50	0.24	12	36	-0.242 (0.039)	0.915	1.102	0.915 [0.896, 0.931]	0.000/0.000
50	1.00	50	36	-0.050 (0.038)	0.948	1.056	0.939 [0.922, 0.952]	0.000/0.000
50	4.00	200	36	0.050 (0.037)	0.978	1.022	0.940 [0.924, 0.953]	0.000/0.000
100	0.25	25	49	-0.204 (0.038)	0.960	1.043	0.939 [0.922, 0.952]	0.000/0.000
100	1.00	100	49	-0.003 (0.036)	1.009	0.991	0.954 [0.939, 0.965]	0.000/0.000
100	4.00	400	49	0.070 (0.036)	1.004	0.996	0.956 [0.941, 0.967]	0.000/0.000
200	0.25	50	64	0.002 (0.038)	0.952	1.051	0.934 [0.917, 0.948]	0.000/0.000
200	1.00	200	64	0.008 (0.037)	0.993	1.007	0.959 [0.945, 0.970]	0.000/0.000
200	4.00	800	64	0.019 (0.037)	0.993	1.008	0.950 [0.935, 0.962]	0.000/0.000

Notes: Each cell uses 1,000 replications. Parentheses report Monte Carlo standard errors for the $\sqrt{NT}$ -scaled bias. Coverage is unconditional; brackets give Wilson intervals. The final column reports the root- and variance-default rates. The basis dimension is $K = 36, 49, 64$ at $T = 50, 100, 200$, respectively.

C Fixed-Sieve Bias Mechanism and Extensions

C.1 Shared Finite-Range Probability Bounds

All bounds below may be applied conditionally on the unit information, with constants uniform in the units. The dependence range is fixed.

Lemma C.1 (Scalar and vector means, lag products, and window weights). *Let Q_{it} be centered and L -dependent in the sigma-field sense. For scalar or vector Q_{it} with $\|Q_{it}\|_{L_i^4} \leq A$, where $A > 0$,*

the constant in

$$\left\| \frac{1}{T} \sum_t Q_{it} \right\|_{L_i^2} + \left\| \frac{1}{T} \sum_t Q_{it} \right\|_{L_i^4} \leq C_L A T^{-1/2} \quad (\text{C.1})$$

is independent of the vector dimension. If $|Q_{it}| \leq A$ is scalar, then $\mathbb{P}_i(|T^{-1} \sum_t Q_{it}| > x) \leq C_L \exp(-c_L T x^2 / A^2)$, and its conditional L_i^r norm is at most $C_{L,r} A T^{-1/2}$ for every fixed $r \geq 2$. For jointly finite-range scalar arrays U, V with bounded fourth moments,

$$\text{Var}_i \left\{ \frac{1}{T} \sum_{t \in I_T(j)} (U_{it} V_{i,t+j} - \mathbb{E}_i[U_{it} V_{i,t+j}]) \right\} \leq \frac{C_j}{T}, \quad (\text{C.2})$$

for each fixed j . Stationarity gives the exact identity

$$\frac{1}{T} \sum_{t,s=1}^T \mathbb{E}_i[U_{it} V_{is}] = \sum_{|j| < T} (1 - |j|/T) \mathbb{E}_i[U_{i0} V_{ij}]. \quad (\text{C.3})$$

If units are independent, the centered pooled vector mean has conditional L^2 and L^4 norms at most $C_L A / \sqrt{NT}$. These statements apply to fixed-lag oracle products after increasing the range by their fixed lag span; they do not assert finite dependence of fitted arrays.

Proof. The second-moment expansion has at most $T(2L+1)$ nonzero terms $\mathbb{E}_i \langle Q_{is}, Q_{it} \rangle$, each bounded by A^2 . For the fourth moment, expand

$$\mathbb{E}_i \left\| \sum_t Q_{it} \right\|^4 = \sum_{t_1, t_2, t_3, t_4} \mathbb{E}_i [\langle Q_{it_1}, Q_{it_2} \rangle \langle Q_{it_3}, Q_{it_4} \rangle].$$

Join dates separated by at most L . An isolated date is independent of the other three and occurs linearly, so its centered expectation is zero. A surviving graph has at most two components: choose their representative dates in at most T^2 ways, then their remaining offsets in C_L ways. Hölder's inequality bounds each surviving term by A^4 . This proves (C.1). The same argument applied to the independent centered cluster sums proves the pooled bounds. For bounded scalar arrays, partition dates into their $L+1$ residue classes. Each class is an independent sum; the full average is a convex combination of its class averages. Hoeffding's inequality and a union bound give the stated tail, and $\mathbb{E}_i |\bar{Q}_i|^r = r \int_0^\infty x^{r-1} \mathbb{P}_i(|\bar{Q}_i| > x) dx$ gives the moments. A fixed-lag product is $(L+|j|)$ -dependent. Its variance has only $O(T)$ nonzero covariance pairs, proving (C.2). Finally, exactly $T - |j|$ pairs satisfy $s - t = j$, which proves (C.3). $\square$

C.2 Direct Sample-Mean Drift and Sampling Expansion

Population part of Proposition 3.2 and Lemma A.1. Use Assumption 3.1, with K fixed. Here $\Delta_i = \check{\alpha}_i - \alpha_i = T^{-1} \sum_t \zeta_{it}$ exactly. The localization condition and bounded score on $\mathcal{A}$ allow the following global decomposition:

$$\begin{aligned} \bar{\psi}_{iK}(\check{\alpha}_i) - \bar{\psi}_{iK}(\alpha_i) &= \bar{\psi}_{iK}^{(1)} \Delta_i + \frac{1}{2} \bar{\psi}_{iK}^{(2)} \Delta_i^2 + R_{iK}, \\ \|R_{iK}\|_{L_i^r} &\leq C \|\Delta_i\|_{L_i^{3r}}^3 + C \mathbb{P}_i(\check{\alpha}_i \notin I_i)^{1/r} = O(T^{-3/2}), \quad r = 1, 2. \end{aligned}$$

The second line uses Taylor's bound on I_i , a bounded envelope on its complement, and Lemma C.1. Exact stationary window counts give

$$\begin{aligned} T \mathbb{E}_i[\bar{\psi}_{iK}^{(1)} \Delta_i] &= \sum_{|j| \leq L} (1 - |j|/T) \mathbb{E}_i[\psi_{i0,K}^{(1)} \zeta_{ij}], \\ T \mathbb{E}_i[\Delta_i^2] &= \Omega_i + O(T^{-1}), \\ T \left| \mathbb{E}_i[(\bar{\psi}_{iK}^{(2)} - \nu_{iK}) \Delta_i^2] \right| &\leq T \|\bar{\psi}_{iK}^{(2)} - \nu_{iK}\|_{L_i^2} \|\Delta_i\|_{L_i^4}^2 = O(T^{-1/2}). \end{aligned}$$

Averaging these expectations proves (3.2). For centered sampling, put $U_{iK} = \bar{\psi}_{iK}(\check{\alpha}_i) - \bar{\psi}_{iK}(\alpha_i) - \mu_{iK} \Delta_i$. The same expansion gives $\|U_{iK}\|_{L_i^2} = O(T^{-1})$. Thus conditional independence across units implies

$$\mathbb{E} \left[\left| \frac{\sqrt{NT}}{N} \sum_i (U_{iK} - \mathbb{E}_i U_{iK}) \right|^2 \middle| \mathcal{C} \right] = \frac{T}{N} \sum_i \text{Var}_i(U_{iK}) \leq C/T.$$

After subtracting the exact unit expectations, the remaining linear term is $\frac{1}{NT} \sum_{i,t} x_{it,K}$, which proves (A.3). $\square$

Assumption C.2 (Fixed- K sample stability). *The following conditions hold:*

(i) **Acceptance probability and Gram stability.** For the event $\mathcal{G}_K$ constructed in Section D.5, $\mathbb{P}(\mathcal{G}_K^c) \rightarrow 0$. On $\mathcal{G}_K$, $\|\hat{Q}_K\|_{\text{op}} + \|\hat{Q}_K^{-1}\|_{\text{op}} = O_p(1)$.

(ii) **Oracle denominator law.** $\frac{1}{NT} \sum_{i,t} u_{D,it,K}^2 - J_K = o_p(1)$.

(iii) **Bounded targets.** $0 < c_V \leq J_K \leq C_V < \infty$ and $|\theta_K| + \|\gamma_K\| + \|\delta_K\| = O(1)$.

Estimator part of Proposition 3.2. Let R denote D or Y , let b_R denote γ_K or δ_K , and put $F_{it}(a) = p_K(X_{it}, a) \{R_{it} - p_K(X_{it}, a)^\top b_R\}$. Write $\bar{F}_i^{(1)} = T^{-1} \sum_t \partial_a F_{it}(\alpha_i)$. The population projection gives $\frac{1}{NT} \sum_{i,t} \mathbb{E}_i[F_{it}(\alpha_i)] = 0$. Taylor expansion of this fixed-dimensional moment gives

$$\begin{aligned} \frac{1}{NT} \sum_{i,t} F_{it}(\check{\alpha}_i) &= \frac{1}{NT} \sum_{i,t} F_{it}(\alpha_i) + \frac{1}{N} \sum_i \mathbb{E}_i[\bar{F}_i^{(1)}] \Delta_i \\ &\quad + \frac{1}{N} \sum_i (\bar{F}_i^{(1)} - \mathbb{E}_i \bar{F}_i^{(1)}) \Delta_i + O_p(T^{-1}). \end{aligned}$$

The first term is centered unit by unit before applying the shared vector bound; its aggregate population mean is zero by the projection equation. The second term uses centered coordinate means and bounded derivative means. Both are $O_p((NT)^{-1/2})$. Hölder's inequality bounds the average absolute value of the third by C/T . The local Taylor remainder and the bounded exceptional branch give the displayed $O_p(T^{-1})$ term. On $\mathcal{G}_K$, the affine normal equations now give

$$\hat{b}_R - b_R = \hat{Q}_K^{-1} \frac{1}{NT} \sum_{i,t} F_{it}(\check{\alpha}_i), \quad \|\hat{\gamma}_K - \gamma_K\| + \|\hat{\delta}_K - \delta_K\| = O_p(T^{-1}),$$

because $N/T \rightarrow \kappa \in (0, \infty)$ and the Gram is stable. With $d_{\alpha,it,K} = m_K(X_{it}, \check{\alpha}_i) - m_K(X_{it}, \alpha_i)$, the exact square identity (D.36) applies with $e_\gamma = \hat{\gamma}_K - \gamma_K$. Here $\|d_\alpha\|_{NT,2} = O_p(T^{-1/2})$, $|\frac{1}{NT} \sum_{i,t} u_{D,it,K} d_{\alpha,it,K}| \leq \|u_D\|_{NT,2} \|d_\alpha\|_{NT,2}$, and $e_\gamma^\top \hat{Q}_K e_\gamma = O_p(T^{-2})$. Assumption C.2(ii) therefore gives $\hat{J}_K - J_K = o_p(1)$. Also $e_\pi = (\hat{\delta}_K - \delta_K) - \theta_K e_\gamma = O_p(T^{-1})$, so $\sqrt{NT} |e_\gamma^\top \hat{Q}_K e_\pi| = O_p(T^{-1})$.

The population and centered score calculations above, inserted in the exact profile identity, yield

$$\sqrt{NT}(\hat{\theta}_K - \theta_K) = \hat{J}_K^{-1} \frac{1}{\sqrt{NT}} \sum_{i,t} x_{it,K} + \sqrt{N/T} \hat{J}_K^{-1} B_K + o_p(1).$$

The normalized centered sum is $O_p(1)$ by the shared second-moment bound, and $B_K = O(1)$. Replacing $\hat{J}_K^{-1}$ by J_K^{-1} proves (3.3); the acceptance-event transfer in Remark D.18 makes it global. $\square$

C.3 Buffered Within-Unit Splitting

Proof of Proposition 3.3. Use the two evaluation blocks from Section 3.2. Let $\bar{\psi}_{i,j,K}(a)$ be the score average on block j , $\Delta_i^{(-j)} = T_{-j}^{-1} \sum_{s \in \mathcal{I}_{-j,T}} \zeta_{is}$, and $T_{-1} = T_2$, $T_{-2} = T_1$. The closest training and evaluation dates are separated by $2L + 1$. Sigma-field L -dependence therefore makes the evaluation score and its true-coordinate derivatives conditionally independent of the training mean. Taylor expansion gives

$$\begin{aligned} D_{i,j,T} &:= \bar{\psi}_{i,j,K}(\alpha_i + \Delta_i^{(-j)}) - \bar{\psi}_{i,j,K}(\alpha_i) \\ &= \bar{\psi}_{i,j,K}^{(1)} \Delta_i^{(-j)} + \frac{1}{2} \bar{\psi}_{i,j,K}^{(2)} (\Delta_i^{(-j)})^2 + R_{i,j,T}. \end{aligned}$$

The shared mean bounds give $\mathbb{E}_i |R_{i,j,T}| \leq CT_{-j}^{-3/2} + C \exp(-cT_{-j})$: the Taylor path lies in I_i except on the exponentially rare event; the block mean lies in the convex compact $\mathcal{A}$ and the score is bounded on its complement. Independence then gives

$$\begin{aligned} \mathbb{E}_i[\bar{\psi}_{i,j,K}^{(1)} \Delta_i^{(-j)}] &= 0, \\ \mathbb{E}_i[D_{i,j,T}] &= \frac{\nu_{iK}}{2} \mathbb{E}_i[(\Delta_i^{(-j)})^2] + O(T^{-3/2}) = \frac{\nu_{iK} \Omega_i}{2T_{-j}} + O(T^{-3/2}). \end{aligned}$$

The last equality uses (C.3) on the contiguous training block. Since $S_{iK} - S_{iK,0} = \sum_{j=1}^2 (T_j/(T - 2L)) D_{i,j,T}$,

$$T \mathbb{E}_i[S_{iK} - S_{iK,0}] = \frac{\nu_{iK} \Omega_i}{2} \frac{T}{T - 2L} \left(\frac{T_1}{T_2} + \frac{T_2}{T_1} \right) + O(T^{-1/2}) = \nu_{iK} \Omega_i + o(1).$$

Here $T_j/T \rightarrow 1/2$. Averaging over units proves (3.5). For the one-way comparison, allow unequal block sizes with the same fixed buffer. Let the first block train the coordinate and the second evaluate the score. The same calculation gives $T \mathbb{E}_i D_{i,2,T} = (T/T_1) \nu_{iK} \Omega_i / 2 + o(1)$; if $T_1/T \rightarrow \lambda$, its aggregate limit is $\lambda^{-1} B_K^{(2)}$. The covariance channel is exactly zero at every T . $\square$

C.4 Smooth Local M -Estimated Coordinates

This calculation uses local identification and an oracle branch. The observable unique-root procedure of Section 6 is treated separately in Appendix G. Write $g_{iT}(a) = T^{-1} \sum_t g(W_{it}, a)$, with $\mathbb{E}_i g(W_{i0}, \alpha_i) = 0$. At α_i , put $G_{it} = \partial_a g(W_{it}, \alpha_i)$, $H_{it} = \partial_a^2 g(W_{it}, \alpha_i)$, $G_i = \mathbb{E}_i G_{i0}$, $H_i = \mathbb{E}_i H_{i0}$, $M_{iT} = T^{-1} \sum_t G_{it} - G_i$, and $\bar{H}_i = T^{-1} \sum_t H_{it}$. Then $\xi_{it} = -G_i^{-1} g(W_{it}, \alpha_i)$ and $\bar{\xi}_i = T^{-1} \sum_t \xi_{it}$.

Assumption C.3 (Local M -score regularity). *The following conditions hold.*

(i) **Sampling and dependence.** *The unit arrays are independent across i . Conditional on the unit information $\mathcal{C}_i$, for each i the joint array containing $g(W_{it}, a)$, its first three a -derivatives, the fixed- K stacked moment, and its first three a -derivatives is strictly stationary and L -dependent*

for a fixed integer L . Here L -dependence is meant in the sigma-field sense: for finite index sets $\mathcal{T}_1, \mathcal{T}_2 \subset \mathbb{Z}$ satisfying $\min_{s \in \mathcal{T}_1, t \in \mathcal{T}_2} |s - t| > L$, the sigma-fields generated by all listed variables at dates in $\mathcal{T}_1$ and $\mathcal{T}_2$ are independent conditional on $\mathcal{C}_i$.

(ii) Local smoothness. For a common fixed $\delta > 0$, these variables are uniformly bounded on $I_i = [\alpha_i - \delta, \alpha_i + \delta]$, and their highest displayed derivatives are uniformly Lipschitz there.

(iii) Local identification. $0 < c_G \leq |G_i| \leq C_G$ and $\inf_{a \in I_i} |\mathbb{E}_i[\partial_a g(W_{i0}, a)]| \geq c_G$. For each i , the map $a \mapsto \mathbb{E}_i[\partial_a g(W_{i0}, a)]$ does not change sign on I_i .

On the event that g_{iT} has exactly one zero in I_i , select it as $\tilde{\alpha}_i$; otherwise set $\tilde{\alpha}_i = \alpha_i$. Denote this fallback event by $\mathcal{D}_{iT}$. The branch is an oracle device because I_i and the fallback use the true coordinate. The localization and root-expansion statements below use only the moment g conditions; the fixed-sieve downstream-score conditions enter the subsequent scalar drift calculation.

Lemma C.4 (Localization of a regular local root). *Under Assumption C.3, $\mathbb{P}_i(\mathcal{D}_{iT}) \leq C \exp(-cT)$ and, for every fixed $r \geq 2$,*

$$\|\tilde{\alpha}_i - \alpha_i\|_{L_i^r} + \|M_{iT}\|_{L_i^r} + \|\bar{H}_i - H_i\|_{L_i^r} \leq C_r T^{-1/2}. \quad (\text{C.4})$$

The proof also gives a uniform sample-Jacobian event with exponentially small failure probability on the specified identification interval.

Proof. Apply Lemma C.1 to the moment and its first two centered derivatives at the true coordinate. For the Jacobian on I_i , use a grid of spacing T^{-2} , containing at most CT^2 points. At every point the scalar concentration bound gives failure probability $C \exp(-cT)$ at tolerance $c_G/4$. The bounded second derivative gives interpolation error at most CT^{-2} . A union bound therefore yields

$$\sup_{a \in I_i} |\partial_a g_{iT}(a) - \mathbb{E}_i \partial_a g(W_{i0}, a)| < c_G/2 \quad (\text{C.5})$$

except with probability $C \exp(-c'T)$. Intersect this event with $|g_{iT}(\alpha_i)| < c_G \delta/2$. Writing $\varsigma_i = \text{sgn}(G_i)$, its signed endpoint values satisfy

$$\begin{aligned} \varsigma_i g_{iT}(\alpha_i - \delta) &\leq |g_{iT}(\alpha_i)| - c_G \delta/2 < 0, \\ \varsigma_i g_{iT}(\alpha_i + \delta) &\geq -|g_{iT}(\alpha_i)| + c_G \delta/2 > 0. \end{aligned}$$

The signed sample derivative is at least $c_G/2$, so continuity and monotonicity give exactly one zero and $|\tilde{\alpha}_i - \alpha_i| \leq 2c_G^{-1}|g_{iT}(\alpha_i)|$. On the complement the oracle error is bounded by δ . The scalar tail and its integrated moments prove all stated rates. The same grid argument works on any common compact interval with the specified derivative bounds and identification condition. A bounded observable selector agreeing with this branch except with probability $C \exp(-cT)$ inherits its moments: the selector difference has L_i^r norm at most $C \exp(-cT/r)$. $\square$

Proposition C.5 (Shared second-order root expansion). *Define*

$$\begin{aligned} q_{iT,1} &= -G_i^{-1} M_{iT} \bar{\xi}_i - \frac{G_i^{-1}}{2} \bar{H}_i (\bar{\xi}_i)^2, \\ q_{iT,2} &= -G_i^{-1} M_{iT} \bar{\xi}_i - \frac{G_i^{-1}}{2} H_i (\bar{\xi}_i)^2. \end{aligned} \quad (\text{C.6})$$

The first expression uses the sample Hessian; the second uses its population mean. For every fixed $r \geq 2$,

$$\|q_{iT,j}\|_{L_i^r} \leq C_r T^{-1}, \quad \|\tilde{\alpha}_i - \alpha_i - \bar{\xi}_i - q_{iT,j}\|_{L_i^r} \leq C_r T^{-3/2}, \quad j = 1, 2. \quad (\text{C.7})$$

With $\beta_{iT,j} = T \mathbb{E}_i q_{iT,j}$ and $\tilde{q}_{iT,j} = q_{iT,j} - \mathbb{E}_i q_{iT,j}$, the exact split is $q_{iT,j} = \beta_{iT,j}/T + \tilde{q}_{iT,j}$. Moreover, $\|q_{iT,1} - q_{iT,2}\|_{L_i^r} = O(T^{-3/2})$.

Proof. On the selected-root event, Taylor's theorem gives

$$\begin{aligned} 0 &= -G_i \bar{\xi}_i + (G_i + M_{iT})(\tilde{\alpha}_i - \alpha_i) + \frac{\bar{H}_i}{2}(\tilde{\alpha}_i - \alpha_i)^2 + R_{g,i}, \\ |R_{g,i}| &\leq C|\tilde{\alpha}_i - \alpha_i|^3, \\ \tilde{\alpha}_i - \alpha_i - \bar{\xi}_i &= -G_i^{-1}M_{iT}(\tilde{\alpha}_i - \alpha_i) \\ &\quad - \frac{G_i^{-1}\bar{H}_i}{2}(\tilde{\alpha}_i - \alpha_i)^2 - G_i^{-1}R_{g,i}. \end{aligned}$$

Hölder's inequality and the root moments first give $\|\tilde{\alpha}_i - \alpha_i - \bar{\xi}_i\|_{L_i^r} = O(T^{-1})$. Subtract the definition of $q_{iT,1}$ from the last identity:

$$\begin{aligned} G_i(\tilde{\alpha}_i - \alpha_i - \bar{\xi}_i - q_{iT,1}) &= -M_{iT}(\tilde{\alpha}_i - \alpha_i - \bar{\xi}_i) \\ &\quad - \frac{\bar{H}_i}{2}\{(\tilde{\alpha}_i - \alpha_i)^2 - (\bar{\xi}_i)^2\} - R_{g,i}. \end{aligned}$$

The square difference is $(\tilde{\alpha}_i - \alpha_i - \bar{\xi}_i)(\tilde{\alpha}_i - \alpha_i + \bar{\xi}_i)$. Its L_i^r norm, and that of the other terms, is $O(T^{-3/2})$. On the fallback event all quantities are bounded and its probability is exponentially small. Finally, $q_{iT,1} - q_{iT,2} = -(G_i^{-1}/2)(\bar{H}_i - H_i)(\bar{\xi}_i)^2$ has the same order by the derivative-mean and fourth-moment bounds. This proves both expansions and their exact centered splits. $\square$

Fixed-sieve scalar score drift. Use the first two channels in (6.4), with the influence ξ defined by the local moment. Under the same local conditions, the calculation holds along the oracle branch with its exact first-step mean retained:

$$\frac{T}{N} \sum_i \mathbb{E}_i[\bar{\psi}_{iK}(\tilde{\alpha}_i) - \bar{\psi}_{iK}(\alpha_i)] = B_K^{(1)} + B_K^{(2)} + \frac{1}{N} \sum_i \mu_{iK} \beta_{iT,1} + O(T^{-1/2}).$$

To verify this, substitute $\tilde{\alpha}_i - \alpha_i = \bar{\xi}_i + q_{iT,1} + O_{L_i^r}(T^{-3/2})$ into the score Taylor expansion. The leading expectations are

$$\begin{aligned} T \mathbb{E}_i[\bar{\psi}_{iK}^{(1)} \bar{\xi}_i] &= \sum_{|j| \leq L} \mathbb{E}_i[\psi_{i0,K}^{(1)} \xi_{ij}] + O(T^{-1}), \\ T \mathbb{E}_i[\bar{\psi}_{iK}^{(1)} q_{iT,1}] &= \mu_{iK} \beta_{iT,1} + T \mathbb{E}_i[(\bar{\psi}_{iK}^{(1)} - \mu_{iK}) \tilde{q}_{iT,1}], \\ \frac{T}{2} \mathbb{E}_i[\bar{\psi}_{iK}^{(2)} (\bar{\xi}_i)^2] &= \frac{\nu_{iK} \Omega_i}{2} + \frac{T}{2} \mathbb{E}_i[(\bar{\psi}_{iK}^{(2)} - \nu_{iK}) (\bar{\xi}_i)^2] + O(T^{-1}). \end{aligned}$$

The two remaining centered products are $O(T^{-1/2})$ by Hölder's inequality; the cubic score remainder has the same scaled expectation. For a sample mean, $G_i = -1$, $M_{iT} = H_i = \bar{H}_i = 0$, so both quadratic root terms vanish. Exact finite-window first-step means are converted to β_i in Appendix G.

C.5 Smooth Changes of the Local Coordinate

The three channels in (6.4) describe the same total drift under a smooth re-expression of a given coordinate and score. Let $\alpha_i^* = F(\alpha_i)$ and $\hat{\alpha}_i^* = F(\hat{\alpha}_i)$, where F is a C^3 diffeomorphism on a compact neighborhood containing the coordinate paths, has bounded derivatives, and satisfies $\inf |F'| > 0$. Write $F_i' = F'(\alpha_i)$ and $F_i'' = F''(\alpha_i)$. Taylor's theorem and (6.3) give

$$\begin{aligned} \hat{\alpha}_i^* - \alpha_i^* &= F_i'(\hat{\alpha}_i - \alpha_i) + \frac{1}{2} F_i''(\hat{\alpha}_i - \alpha_i)^2 + O_{L_i^2}(T^{-3/2}), \\ \xi_{it}^* &= F_i' \xi_{it}, \quad \Omega_i^* = (F_i')^2 \Omega_i, \quad \beta_i^* = F_i' \beta_i + \frac{1}{2} F_i'' \Omega_i. \end{aligned}$$

For the last identity, taking expectations in the coordinate expansion gives $T \mathbb{E}_i[\hat{\alpha}_i - \alpha_i] = \beta_i + O(T^{-1/2})$. Its moment bounds and the finite-window covariance identity also give $T \mathbb{E}_i[(\hat{\alpha}_i - \alpha_i)^2] = \Omega_i + O(T^{-1/2})$.

Express the same scalar score as $\psi_{it,K}^*(a^*) = \psi_{it,K}(F^{-1}(a^*))$. The chain rule, evaluated at the true coordinate, yields

$$\begin{aligned}\partial_{a^*} \psi_{it,K}^* &= \frac{\psi_{it,K}^{(1)}}{F_i'}, \\ \partial_{a^*}^2 \psi_{it,K}^* &= \frac{\psi_{it,K}^{(2)}}{(F_i')^2} - \frac{\psi_{it,K}^{(1)} F_i''}{(F_i')^3}.\end{aligned}$$

The derivative–influence term is unchanged at every lag:

$$\mathbb{E}_i[(\partial_{a^*} \psi_{i0,K}^*) \xi_{ij}^*] = \mathbb{E}_i \left[\frac{\psi_{i0,K}^{(1)}}{F_i'} F_i' \xi_{ij} \right] = \mathbb{E}_i[\psi_{i0,K}^{(1)} \xi_{ij}].$$

For the remaining two terms, insert the transformed derivative means, first-step bias, and variance before canceling:

$$\begin{aligned}\frac{\mu_{iK}}{F_i'} \beta_i^* + \frac{1}{2} \left\{ \frac{\nu_{iK}}{(F_i')^2} - \frac{\mu_{iK} F_i''}{(F_i')^3} \right\} \Omega_i^* &= \mu_{iK} \beta_i + \frac{\mu_{iK} F_i''}{2 F_i'} \Omega_i + \frac{1}{2} \nu_{iK} \Omega_i - \frac{\mu_{iK} F_i''}{2 F_i'} \Omega_i \\ &= \mu_{iK} \beta_i + \frac{1}{2} \nu_{iK} \Omega_i.\end{aligned}$$

Averaging across units therefore preserves $B_K^{(1)}$ and the sum $B_K^{(2)} + B_K^{(3)}$, hence their total B_K . A nonlinear transformation of a sample mean generally has a nonzero first-step coefficient even though the original β_i is zero. This calculation keeps the score and generated coordinate fixed as statistical objects. Choosing a new spline space after transformation changes the finite-sieve projection and is a different operation.

D Technical Foundations for Growing-Sieve Estimation

D.1 Boundary Tensor-Spline Properties

The results below establish the deterministic spline properties used by the main panel theorem. We use standard B-spline notation and basic approximation facts from [dB01, Sch07]. The construction below gives uniform bounds for interior and boundary basis functions. Let $W^+ = [0, 1]^{d_X} \times [0, 1]$, let $d_W = d_X + 1$ be fixed, and write the last coordinate as a . Set $b_T = C_b \sqrt{\frac{\log N}{T}}$ and let the deterministic coordinates satisfy the quasi-uniform low-discrepancy condition (5.1). Let p_K be the Lebesgue- L^2 normalized tensor-product B-spline basis of fixed degree $p \geq 4$, with simple interior knots, endpoint knot multiplicity $p + 1$, and quasi-uniform positive knot spans of width λ_K . The associated spline space is $\mathcal{S}_K = \{p_K^\top b : b \in \mathbb{R}^K\}$ of the sieve in Section 2.2. Write $p_K(x, a) = p_{K_x}^x(x) \otimes p_{K_a}^a(a)$, where $K = K_x K_a$, $K_x \asymp \lambda_K^{-d_X}$, $K_a \asymp \lambda_K^{-1}$, and $K \asymp \lambda_K^{-d_W}$. When $d_X = 0$, take $K_x = 1$ and $p_{K_x}^x \equiv 1$. Throughout this subsection, $N \lambda_K \rightarrow \infty$ and $b_T / \lambda_K \rightarrow 0$.

Use $\mu_c(dw) = dw$ for Lebesgue measure and $\mu_g(dw) = dx \otimes \{N^{-1} \sum_{i=1}^N \delta_{\alpha_i}(da)\}$ for the panel-grid measure. Define $G_c = \int p_K p_K^\top d\mu_c$ and $G_g = \int p_K p_K^\top d\mu_g$; both are $K \times K$ matrices. Whenever the inverse exists, let $\Pi_{K,0} f = p_K^\top G_c^{-1} \int p_K f d\mu_c$ and $\Pi_{K,N} f = p_K^\top G_g^{-1} \int p_K f d\mu_g$. We compare the continuous and discrete projection operators below.

Local Dual Functionals at the Boundary

Lemma D.1 (Local dual functionals). *There is a tensor-product quasi-interpolant $\mathcal{I}_K$ that reproduces $\mathcal{S}_K$ and satisfies $\|\mathcal{I}_K f\|_\infty \leq C \|f\|_\infty$, with a constant independent of K . The bound holds uniformly for interior and boundary basis functions.*

Proof. For raw one-dimensional open-clamped B-splines B_j , $j = -p, \dots, M-1$, let $I_r = (\tau_r, \tau_{r+1})$ be the positive cells. Select $I_{r(j)}$ with $r(j) = \max\{j, 0\}$; this belongs to the support of B_j , including the repeated-knot boundary functions. On I_r the active set is $A_r = \{r-p, \dots, r\}$; its $p+1$ restrictions are linearly independent. After affine rescaling, their raw local Gram is $H_r^B = h_r G_{c_r}(\vartheta)$, where $h_r \asymp \lambda_K$, $c_r = ((p-r)_+, \{p-(M-1-r)\}_+)$ records the finite set of endpoint patterns, and the positive span ratios ϑ lie in a compact set. Continuity and local linear independence therefore give $\|(H_r^B)^{-1}\| \leq C \lambda_K^{-1}$ uniformly, including cells near both boundaries. Zero endpoint spans belong to the pattern, not the ratio coordinates. For $s_j = \|B_j\|_{L^2} \asymp \lambda_K^{1/2}$ and $\phi_j = B_j/s_j$, define

$$\begin{aligned} d_j^B(a) &= \mathbf{1}\{a \in I_{r(j)}\} \sum_{k \in A_{r(j)}} [(H_{r(j)}^B)^{-1}]_{jk} B_k(a), \\ \lambda_j^\phi(f) &= s_j \int_{I_{r(j)}} d_j^B(a) f(a) da, & \lambda_j^\phi(\phi_k) &= \delta_{jk}, \\ |\lambda_j^\phi(f)| &\leq C \lambda_K^{1/2} \|f\|_{\infty, I_{r(j)}}. \end{aligned}$$

The last bound uses $\|d_j^B\|_\infty \leq C \lambda_K^{-1}$ and a cell of length $O(\lambda_K)$. Tensorize these integral functionals and normalized bases as Λ_j and Φ_j . Their kernels are supported on the selected tensor cells and satisfy $|\Lambda_j(f)| \leq C \lambda_K^{d_W/2} \|f\|_\infty$ and $\|\Phi_j\|_\infty \leq C \lambda_K^{-d_W/2}$. Since only $(p+1)^{d_W}$ terms are active at any point,

$$|\mathcal{I}_K f(w)| = \left| \sum_j \Lambda_j(f) \Phi_j(w) \right| \leq C \lambda_K^{d_W/2} \|f\|_\infty \lambda_K^{-d_W/2} = C \|f\|_\infty.$$

Tensor biorthogonality proves reproduction of $\mathcal{S}_K$. □

Continuous Riesz Bounds and Coordinate-Grid Quadrature

Lemma D.2 (Continuous Riesz and derivative inverse bounds). *There are constants $0 < c < C < \infty$, independent of K, N, T , such that*

$$c \|b\|_2^2 \leq \int_{W^+} \{b^\top p_K(w)\}^2 dw \leq C \|b\|_2^2. \quad (\text{D.1})$$

For $s_b = p_K^\top b$ and $j = 1, 2, 3, 4$,

$$\|\partial_a^j s_b\|_{L^2(W^+)}^2 \leq C \lambda_K^{-2j} \|b\|_2^2. \quad (\text{D.2})$$

For $j = 4$, the derivative is understood cellwise, equivalently as the almost-everywhere weak derivative. The fourth derivative need not agree across knots.

Proof. At most $(p+1)^{d_W}$ normalized tensor functions overlap, and each has continuous L^2 norm one. Hence

$$\|s_b\|_{L^2}^2 \leq (p+1)^{d_W} \sum_j |b_j|^2 = (p+1)^{d_W} \|b\|_2^2.$$

For the lower bound, the selected-cell dual gives $b_j = \Lambda_j(s_b)$. Polynomial norm equivalence on its selected tensor cell J_j , after affine rescaling, gives

$$\begin{aligned} |b_j|^2 &\leq C\lambda_K^{d_W} \|s_b\|_{\infty, J_j}^2 \leq C\|s_b\|_{L^2(J_j)}^2, \\ \|b\|_2^2 &\leq C(p+1)^{d_W} \sum_r \|s_b\|_{L^2(I_r)}^2 = C(p+1)^{d_W} \|s_b\|_{L^2}^2. \end{aligned}$$

Each tensor cell has at most $(p+1)^{d_W}$ preimages under the selection rule, which justifies the second row. This proves (D.1). On a tensor cell with a -width h_a , rescaling to the unit cube and the fixed-degree polynomial inverse inequality give

$$\begin{aligned} \|\partial_a^j s_b\|_{L^2(I_r)}^2 &\leq Ch_a^{-2j} \|s_b\|_{L^2(I_r)}^2, \\ \|\partial_a^j s_b\|_{L^2(W+)}^2 &\leq C\lambda_K^{-2j} \|s_b\|_{L^2(W+)}^2 \leq C\lambda_K^{-2j} \|b\|_2^2. \end{aligned}$$

For $j \leq 3$ the derivatives agree across knots; for $j = 4$ the cellwise bounds give the weak derivative because $p \geq 4$ and the third derivative is continuous. The partition of unity represents a constant spline with zero derivatives, so no lower derivative bound is available. $\square$

For a fixed integer m_q , let $\mathcal{Q}_{m_q, K}$ be the class of nonnegative continuous functions on $[0, 1]$ whose restriction to each knot span I_r is a polynomial of degree at most m_q .

Lemma D.3 (Quasi-uniform coordinate quadrature). *For every $q \in \mathcal{Q}_{m_q, K}$,*

$$\left| \frac{1}{N} \sum_{i=1}^N q(\alpha_i) - \int_0^1 q(a) da \right| \leq C \left\{ \frac{1}{N\lambda_K} + \frac{b_T}{\lambda_K} \right\} \int_0^1 q(a) da. \quad (\text{D.3})$$

Consequently, with $\varepsilon_{N, T, K} = C\{(N\lambda_K)^{-1} + b_T/\lambda_K\}$, the continuous and grid Grams satisfy

$$-\varepsilon_{N, T, K} G_c \preceq G_g - G_c \preceq \varepsilon_{N, T, K} G_c. \quad (\text{D.4})$$

The result allows arbitrary grid-knot alignment and permits empty knot cells.

Proof. For nonnegative q of fixed degree on a cell I_r of width $h_r \asymp \lambda_K$, affine rescaling and polynomial norm equivalence give

$$\|q\|_{\infty, I_r} + \int_{I_r} |q'(a)| da \leq Ch_r^{-1} \int_{I_r} q(a) da. \quad (\text{D.5})$$

Nonnegativity removes cancellation in the L^1 norm. Continuity leaves no knot-jump variation, so

$$\text{TV}(q) = \sum_r \int_{I_r} |q'| \leq C\lambda_K^{-1} \int_0^1 q. \quad (\text{D.6})$$

For $I_T = [2b_T, 1 - 2b_T]$ and its two omitted strips B_T , $|B_T \cap I_r| \leq \min\{h_r, 4b_T\}$ implies

$$\int_{B_T} q \leq \sum_r |B_T \cap I_r| \|q\|_{\infty, I_r} \leq C \frac{b_T}{\lambda_K} \int_0^1 q. \quad (\text{D.7})$$

This includes strips cutting through a knot cell. The signed measure $\nu = N^{-1} \sum_i \delta_{\alpha_i} - \text{Unif}(I_T)$ has total mass zero, cumulative function $F_\nu = F_N - F_T^I$, and $F_\nu(0) = F_\nu(1) = 0$. Stieltjes integration by parts and the discrepancy assumption give

$$\begin{aligned} \int q d\nu &= [qF_\nu]_0^1 - \int F_\nu dq = - \int F_\nu dq, \\ \left| \frac{1}{N} \sum_i q(\alpha_i) - \frac{1}{1 - 4b_T} \int_{I_T} q \right| &\leq \|F_\nu\|_\infty \text{TV}(q) \leq \frac{C}{N\lambda_K} \int_0^1 q. \end{aligned}$$

The remaining difference is exactly

$$\frac{1}{1-4b_T} \int_{I_T} q - \int_0^1 q = \frac{4b_T}{1-4b_T} \int_{I_T} q - \int_{B_T} q.$$

For large T , $1-4b_T \geq 1/2$; (D.7) bounds its absolute value by $C(b_T/\lambda_K) \int_0^1 q$. Combining the two differences proves (D.3). The bound allows the coordinate grid to leave some knot cells empty. For any coefficient vector c , the function $q_c(a) = \int \{c^\top p_K(x, a)\}^2 dx$ is continuous, nonnegative, and cellwise polynomial of degree at most $2p$. Its grid and continuous integrals are $c^\top G_g c$ and $c^\top G_c c$. Thus (D.3) gives $|c^\top (G_g - G_c) c| \leq \varepsilon_{N,T,K} c^\top G_c c$ for every c , which is precisely (D.4). $\square$

Projection Stability and Derivative Approximation

Lemma D.4 (Projection stability and derivative approximation). *For all sufficiently large N, T , both $\Pi_{K,0}$ and $\Pi_{K,N}$ are well defined and reproduce $\mathcal{S}_K$. For $\mu \in \{\mu_c, \mu_g\}$ and $r = 0, 1, 2, 3$,*

$$\|\partial_a^r \Pi_\mu f\|_\infty \leq C \lambda_K^{-r} \|f\|_\infty, \quad (\text{D.8})$$

where $\Pi_{\mu_c} = \Pi_{K,0}$ and $\Pi_{\mu_g} = \Pi_{K,N}$. If $f \in C^{\tau_f}(W^+)$, $3 < \tau_f \leq p+1$, then, separately for $\Pi \in \{\Pi_{K,0}, \Pi_{K,N}\}$,

$$\|\partial_a^r (f - \Pi f)\|_\infty \leq C \lambda_K^{\tau_f - r}, \quad r = 0, 1, 2, 3. \quad (\text{D.9})$$

Here C^{τ_f} denotes the usual isotropic Hölder class: if $\tau_f = m + \beta$ with $0 < \beta < 1$, derivatives of order m are β -Hölder, while the integer case has the usual C^m meaning. Membership on W^+ is defined through extension to a fixed open rectangle containing W^+ . The constant in (D.9) may depend on the C^{τ_f} -norm of this fixed extension, but not on K, N, T . Writing $\partial_{a,\text{pw}}^4$ for the fourth derivative on knot-cell interiors, the stability conclusion also gives

$$\|\partial_{a,\text{pw}}^4 \Pi_\mu f\|_{\infty,\text{pw}} \leq C \lambda_K^{-4} \|f\|_\infty. \quad (\text{D.10})$$

Consequently, the continuous third derivative is globally Lipschitz in the a -coordinate:

$$\sup_x \sup_{a \neq a'} \frac{|\partial_a^3 \Pi_\mu f(x, a') - \partial_a^3 \Pi_\mu f(x, a)|}{|a' - a|} \leq C \lambda_K^{-4} \|f\|_\infty. \quad (\text{D.11})$$

The suprema in (D.10) are over cell interiors, and the points in (D.11) are restricted to W^+ .

Proof. Local moments and inverse Grams. Each normalized tensor basis function has support R_j of volume $O(\lambda_K^{d_W})$. The coordinate spacing condition gives $\#\{i : \alpha_i \in R_{a,j}\} \leq C(N\lambda_K + 1)$. Since $N\lambda_K \rightarrow \infty$, both measures satisfy $\mu(R_j) \leq C\lambda_K^{d_W}$, including boundary supports. Affine scaling gives $\|\partial_a^r p_{K,j}\|_\infty \leq C\lambda_K^{-d_W/2-r}$ for $r \leq 3$, and cellwise for $r = 4$. Therefore

$$\left| \int p_{K,j} f d\mu \right| \leq \|p_{K,j}\|_\infty \mu(R_j) \|f\|_\infty \leq C \lambda_K^{d_W/2} \|f\|_\infty.$$

The Riesz and relative quadrature bounds give $cI_K \preceq G_\mu \preceq CI_K$ uniformly for μ_c, μ_g . Set $a_G = 2/(c+C)$, $R_\mu = I_K - a_G G_\mu$, and $\rho_G = (C-c)/(C+c) < 1$. The spectra give $\|R_\mu\|_{\text{op}} \leq \rho_G$, hence $G_\mu^{-1} = a_G \sum_{n \geq 0} R_\mu^n$. Index the basis by its rectangular tensor lattice. Supports can overlap only when indices differ by at most $L_p = p+1$ in each coordinate. The corresponding graph distance is $d_L(\mathbf{j}, \mathbf{k}) = \lceil \|\mathbf{j} - \mathbf{k}\|_\infty / L_p \rceil$: each edge moves at most L_p , and coordinatewise steps toward the target stay inside the rectangular lattice. A nonzero entry of R_μ^n is a walk of length n , so it vanishes

for $n < d_L$. Consequently,

$$\begin{aligned} |(G_\mu^{-1})_{j\mathbf{k}}| &\leq a_G \sum_{n \geq d_L} \|R_\mu^n\|_{\text{op}} \leq \frac{a_G}{1 - \rho_G} \rho_G^{d_L(j, \mathbf{k})}, \\ \sum_{\mathbf{k}} |(G_\mu^{-1})_{j\mathbf{k}}| &\leq C \sum_{s \geq 0} (1 + s)^{d_W - 1} \rho_G^s \leq C. \end{aligned}$$

For $s \geq 1$, a graph shell contains at most $(2L_p s + 1)^{d_W} - \{2L_p(s - 1) + 1\}^{d_W} \leq C(1 + s)^{d_W - 1}$ indices; the $s = 0$ shell has one index. Boundary truncation can only reduce these numbers. We have proved the uniform row-sum bound

$$\sup_j \sum_{\mathbf{k}} |(G_\mu^{-1})_{j\mathbf{k}}| \leq C. \quad (\text{D.12})$$

Stability and approximation. For $b_\mu(f) = G_\mu^{-1} \int p_K f \, d\mu$, the local moment and row-sum bounds give $\|b_\mu(f)\|_\infty \leq C \lambda_K^{d_W/2} \|f\|_\infty$. Bounded overlap then gives, through order three and cellwise at order four,

$$|\partial_a^r \Pi_\mu f(w)| \leq \sum_{j: w \in R_j} |b_{\mu,j}(f)| |\partial_a^r p_{K,j}(w)| \leq C \lambda_K^{-r} \|f\|_\infty.$$

The continuous third derivative telescopes across intervening knots, so

$$|\partial_a^3 \Pi_\mu f(x, a') - \partial_a^3 \Pi_\mu f(x, a)| \leq \int_a^{a'} |\partial_{a,\text{pw}}^4 \Pi_\mu f(x, u)| \, du \leq C \lambda_K^{-4} \|f\|_\infty |a' - a|$$

for $a < a'$. These prove all stability statements. Reproduction follows from $\Pi_\mu(p_K^\top b) = p_K^\top G_\mu^{-1} G_\mu b = p_K^\top b$.

For approximation, let $k_f = \lceil \tau_f \rceil - 1$, so $3 \leq k_f \leq p$, and let P_w be the degree- k_f Taylor polynomial of the fixed extension at w . It belongs to $\mathcal{S}_K$ and is reproduced by $\mathcal{I}_K$. Every selected dual cell J_j whose basis support contains w lies within $C \lambda_K$ of w : the selected and active cell indices differ by at most $p + 1$ per coordinate. There are at most $(p + 2)^{d_W}$ such cells, including at boundaries and knots. The Hölder–Taylor bound on this patch and the local dual bound give

$$\begin{aligned} \|f - P_w\|_{\infty, J_j} &\leq C \lambda_K^{\tau_f}, & |\Lambda_j(f - P_w)| &\leq C \lambda_K^{d_W/2 + \tau_f}, \\ \partial_a^r (f - \mathcal{I}_K f)(w) &= - \sum_{j: w \in \text{supp}(\Phi_j)} \Lambda_j(f - P_w) \partial_a^r \Phi_j(w). \end{aligned}$$

The last equality uses $\partial_a^r (f - P_w)(w) = 0$ and $\mathcal{I}_K P_w = P_w$. Scaling and the fixed overlap therefore prove

$$\|\partial_a^r (f - \mathcal{I}_K f)\|_\infty \leq C \lambda_K^{\tau_f - r}, \quad r = 0, 1, 2, 3. \quad (\text{D.13})$$

Put $e = f - \mathcal{I}_K f$. Reproduction gives $f - \Pi_\mu f = e - \Pi_\mu e$, so

$$\|\partial_a^r (f - \Pi_\mu f)\|_\infty \leq \|\partial_a^r e\|_\infty + C \lambda_K^{-r} \|e\|_\infty \leq C \lambda_K^{\tau_f - r}.$$

This proves the approximation separately for each projection measure. $\square$

For later reference, the benchmark nuisance projections are $m_K = \Pi_{K,N} m_0$, $h_K = \Pi_{K,N} h_0$, and $\ell_K = \theta_0 m_K + h_K$. These are panel-grid projections. The continuous projections $\Pi_{K,0} m_0$ and $\Pi_{K,0} h_0$ remain separate comparison objects.

D.2 Generated-Coordinate Clipping and Path Localization

Let ζ_{it} be the bounded, centered finite-range process in Assumption 5.1, and set $\Delta_i = \frac{1}{T} \sum_{t=1}^T \zeta_{it}$, $\check{\alpha}_i = \alpha_i + \Delta_i$, and $\hat{\alpha}_i = \Pi_{[b_T, 1-b_T]}(\check{\alpha}_i)$. Write $d_i = \hat{\alpha}_i - \alpha_i$ and $c_i = d_i - \Delta_i$. Let c_{fm} denote the weighted-Hoeffding constant for this bounded process. Then $\mathbb{P}(|\Delta_i| > x) \leq 2 \exp(-c_{\text{fm}} T x^2)$. Set $a_0 := c_{\text{fm}} C_b^2 > 1$ and $\lambda_{\zeta, T} := T \mathbb{E}[\Delta_i^2] = \sum_{|j| \leq L} \left(1 - \frac{|j|}{T}\right) \gamma_{\zeta}(j)$.

Lemma D.5 (Clipped-coordinate facts). *For $\mathcal{E}_T = \{\max_{1 \leq i \leq N} |\Delta_i| \leq b_T\}$, $\Pr(\mathcal{E}_T^c) \leq 2N^{1-a_0} \rightarrow 0$. On $\mathcal{E}_T$, $d_i = \Delta_i$, $c_i = 0$, and $\hat{\alpha}_i = \check{\alpha}_i$. Globally, $\hat{\alpha}_i \in [b_T, 1-b_T]$, every path $\alpha_i + s d_i$, $s \in [0, 1]$, stays in this interval, $|d_i| \leq |\Delta_i|$, and $|c_i| \leq |\Delta_i| \mathbf{1}\{|\Delta_i| > b_T\}$. For $q = 1, \dots, 8$ and nonnegative q_1, q_2 with $q_2 \geq 1$ and $q_1 + q_2 \leq 8$,*

$$\begin{aligned} \max \{ \mathbb{E}[|\Delta_i|^q \mathbf{1}\{|\Delta_i| > b_T\}], \mathbb{E}|c_i|^q \} &\leq C_q b_T^q N^{-a_0}, \\ \mathbb{E}[|\Delta_i|^{q_1} |c_i|^{q_2}] &\leq C_{q_1, q_2} b_T^{q_1+q_2} N^{-a_0}. \end{aligned} \quad (\text{D.14})$$

For $q = 2, \dots, 8$, $\mathbb{E}|d_i|^q \leq C_q T^{-q/2}$. Moreover,

$$\mathbb{E} d_i \leq C b_T N^{-a_0}, \quad \left| \mathbb{E} d_i^2 - \frac{\lambda_{\zeta, T}}{T} \right| \leq C b_T^2 N^{-a_0}. \quad (\text{D.15})$$

Proof. The coordinate-design condition places $\alpha_i \in [2b_T, 1-2b_T]$. Hence $|\Delta_i| \leq b_T$ implies $\check{\alpha}_i \in [b_T, 1-b_T]$. Therefore, clipping does not bind. The union bound gives $\Pr(\mathcal{E}_T^c) \leq 2N \exp(-c_{\text{fm}} T b_T^2) = 2N^{1-a_0}$. Projection onto an interval is nonexpansive. Thus $|d_i| \leq |\Delta_i|$. Clipping binds only when $|\Delta_i| > b_T$. Direct inspection at either endpoint gives the stated bound for c_i . The segment between $\alpha_i \in [2b_T, 1-2b_T]$ and $\hat{\alpha}_i \in [b_T, 1-b_T]$ remains inside $[b_T, 1-b_T]$.

For $q \leq 8$, the layer-cake identity gives

$$\begin{aligned} \mathbb{E}[|\Delta_i|^q \mathbf{1}\{|\Delta_i| > b_T\}] &= b_T^q \Pr(|\Delta_i| > b_T) \\ &\quad + q \int_{b_T}^{\infty} x^{q-1} \Pr(|\Delta_i| > x) dx. \end{aligned}$$

The first term is $O(b_T^q N^{-a_0})$. With $y = \sqrt{T}x$, the integral is bounded by $C b_T^{q-2} T^{-1} N^{-a_0} \leq C b_T^q N^{-a_0}$ because $T b_T^2 = C_b^2 \log N$. The bounds for c_i and the mixed moments follow from $|c_i| \leq |\Delta_i| \mathbf{1}\{|\Delta_i| > b_T\}$. The bounded sample-mean inequality gives $\mathbb{E}|\Delta_i|^q \leq C_q T^{-q/2}$. Non-expansiveness transfers it to d_i . Finally, use $d_i = \Delta_i + c_i$, $\mathbb{E} \Delta_i = 0$, and $\mathbb{E} \Delta_i^2 = \lambda_{\zeta, T}/T$ together with the sharp tail and mixed-moment bounds to prove (D.15). $\square$

D.3 Finite-Range Moment Inputs

Use Lemma C.1 from Appendix C.1. All oracle observations in Assumption 5.1 are measurable with respect to a fixed-width finite-range block. Hence every fixed-lag oracle product obeys the lemma, with its own nonzero mean subtracted. For an envelope a_K , the scalar fourth-moment bound is $C a_K T^{-1/2}$. For vector moment blocks the bound uses their vector norm, so its constant introduces no extra factor of K . Fitted arrays are compared with these oracle arrays through coordinate and coefficient perturbations.

D.4 Nuisance Estimation and Gram-Matrix Results

The results below add the stochastic inputs to the clipping and spline bounds established above. We maintain Assumption 5.1. The array $\{X_{it}\}$ is independent and uniformly distributed on $[0, 1]^{d_x}$ over (i, t) and is independent of the deterministic quasi-uniform coordinate design and every disturbance

process. The bounded process (ζ_{it}, v_{it}) is jointly stationary and L -dependent, while the bounded L -dependent outcome-error process is independent of the joint (ζ, v) process and of X_{it} . In particular, ζ_{it} and v_{it} may be correlated. The data satisfy $Z_{it} = \alpha_i + \zeta_{it}$, $D_{it} = m_0(X_{it}, \alpha_i) + v_{it}$, and $Y_{it} = \theta_0 D_{it} + h_0(X_{it}, \alpha_i) + \varepsilon_{it}$, where $0 < \sigma_v^2 = \mathbb{E}[v_{it}^2] < \infty$. Write $\gamma_{v\zeta}(j) = \mathbb{E}[v_{i0}\zeta_{ij}]$ and $\Omega_{v\zeta} = \sum_{|j| \leq L} \gamma_{v\zeta}(j)$, and write $\gamma_\zeta(j) = \mathbb{E}[\zeta_{i0}\zeta_{ij}]$ and $\Omega_\zeta = \sum_{|j| \leq L} \gamma_\zeta(j)$. The functions obey $m_0 \in C^{\tau_m}(W^+)$ and $h_0 \in C^{\tau_h}(W^+)$, where $3 < \tau_m, \tau_h \leq p + 1$.

Use the generated coordinate, clipping event $\mathcal{E}_T$, and clipped moments from Lemma D.5. In addition to the deterministic growth conditions above, maintain

$$\frac{N}{T} \longrightarrow \kappa_0 \in (0, \infty), \quad \frac{K \log K}{NT} \longrightarrow 0. \quad (\text{D.16})$$

Write $m = \max\{d_W, 3\}$ and $r_K = \lambda_K^{-m/2}/T$.

Generated-Coordinate Perturbation and Nuisance Expansion Let $m_K = \Pi_{K,N} m_0$ and $\ell_K = \Pi_{K,N}(\theta_0 m_0 + h_0) = \theta_0 m_K + h_K$. Define $u_{D,it,K} = D_{it} - m_K(X_{it}, \alpha_i)$ and $u_{Y,it,K} = Y_{it} - \ell_K(X_{it}, \alpha_i)$. For $b \in \{\gamma, \delta\}$, use the shorthand $(R_{\gamma,it,K}, \ell_{\gamma,K}) = (u_{D,it,K}, m_K)$ and $(R_{\delta,it,K}, \ell_{\delta,K}) = (u_{Y,it,K}, \ell_K)$, and let $f_{it,K}^b = p_K(X_{it}, \alpha_i) R_{b,it,K}$. Its first three coordinate derivatives are

$$\begin{aligned} A_{it,K}^b &= p_{K,a} R_{b,it,K} - p_K \ell_{b,K,a}, \\ H_{it,K}^b &= p_{K,aa} R_{b,it,K} - 2p_{K,a} \ell_{b,K,a} - p_K \ell_{b,K,aa}, \\ T_{it,K}^b &= p_{K,aaa} R_{b,it,K} - 3p_{K,aa} \ell_{b,K,a} - 3p_{K,a} \ell_{b,K,aa} - p_K \ell_{b,K,aaa}, \end{aligned} \quad (\text{D.17})$$

where every term on the right is evaluated at (X_{it}, α_i) . Replacing α_i by a generic a defines the corresponding path arrays. Stack the gamma and delta blocks as $f_{c,it,K}$, $A_{c,it,K}$, $H_{c,it,K}$, and $T_{c,it,K}$. Write their means as $\mu_{f,c,i,K}$, $\mu_{A,c,i,K}$, $\mu_{H,c,i,K}$, and $\mu_{T,c,i,K}$. The panel-grid projection normal equations give $\frac{1}{N} \sum_i \mu_{f,c,i,K} = 0$.

Lemma D.6 (Fixed-coordinate nuisance derivative arrays). *For $r = 0, 1, 2, 3, 4$, uniformly over fixed a and bounded scalar functions $q(\cdot, a)$, write $p_{K,a^r} = \partial_a^r p_K$, with $p_{K,a^0} = p_K$. Then*

$$\begin{aligned} \mathbb{E}_X \|p_{K,a^r}(X, a)\|_2^2 &\leq CK \lambda_K^{-2r}, \\ \|\mathbb{E}_X [p_{K,a^r}(X, a) q(X, a)]\|_2 &\leq C \lambda_K^{-r-1/2} \|q(\cdot, a)\|_\infty. \end{aligned} \quad (\text{D.18})$$

For $r = 4$, these bounds concern cell interiors and hence the cellwise, almost-everywhere derivative. The nuisance arrays defined above satisfy

$$\begin{aligned} \frac{1}{N} \sum_i \mathbb{E}_i \|A_{c,it,K} - \mu_{A,c,i,K}\|_2^2 &\leq CK \lambda_K^{-2}, \\ \frac{1}{N} \sum_i \mathbb{E}_i \|H_{c,it,K} - \mu_{H,c,i,K}\|_2^2 &\leq CK \lambda_K^{-4}, \\ \frac{1}{N} \sum_i \mathbb{E}_i \|T_{c,it,K} - \mu_{T,c,i,K}\|_2^2 &\leq CK \lambda_K^{-6}, \end{aligned} \quad (\text{D.19})$$

and

$$\sup_i \|\mu_{A,c,i,K}\|_2 \leq C \lambda_K^{-1/2}, \quad \sup_i \|\mu_{H,c,i,K}\|_2 \leq C \lambda_K^{-3/2}, \quad \sup_i \|\mu_{T,c,i,K}\|_2 \leq C \lambda_K^{-5/2}. \quad (\text{D.20})$$

Proof. The tensor representation $p_K(x, a) = p_{K_x}^x(x) \otimes p_{K_a}^a(a)$ gives

$$\begin{aligned} \mathbb{E}_X \|p_{K,a^r}(X, a)\|_2^2 &= \mathbb{E}_X \|p_{K_x}^x(X)\|_2^2 \|(p_{K_a}^a)^{(r)}(a)\|_2^2 \\ &\leq CK_x \lambda_K^{-1-2r} \leq CK \lambda_K^{-2r}. \end{aligned}$$

Here local support and normalized affine scaling give the middle inequality; for $r = 4$ it is read on cell interiors. For the second bound in (D.18), duality and the x -coordinate Riesz upper bound imply, for every unit vector $c \in \mathbb{R}^{K_x}$,

$$\left| \mathbb{E}_X[(c^\top p_{K_x}^x(X))q(X, a)] \right| \leq \|c^\top p_{K_x}^x\|_{L^2} \|q(\cdot, a)\|_{L^2} \leq C \|q(\cdot, a)\|_\infty.$$

Thus

$$\|\mathbb{E}_X[p_{K_x}^x(X)q(X, a)]\|_2 \leq C \|q(\cdot, a)\|_\infty.$$

Combining this inequality with $\|(p_{K_a}^a)^{(r)}(a)\|_2 \leq C\lambda_K^{-r-1/2}$ proves the asserted weighted-mean bound.

For $b \in \{\gamma, \delta\}$, put $q_\gamma = m_0$, $q_\delta = \theta_0 m_0 + h_0$, and $r_{b,K} = q_b - \ell_{b,K}$. Then $R_{\gamma,it,K} = v_{it} + r_{\gamma,K}(X_{it}, \alpha_i)$ and $R_{\delta,it,K} = \theta_0 v_{it} + \varepsilon_{it} + r_{\delta,K}(X_{it}, \alpha_i)$. Let $\tau_* = \min\{\tau_m, \tau_h\} > 3$. Lemma D.4 gives

$$\|r_{b,K}\|_\infty \leq C\lambda_K^{\tau_*}, \quad \|\partial_a^j \ell_{b,K}\|_\infty \leq C, \quad j = 1, 2, 3.$$

The innovations are bounded and centered, so the derivative table yields

$$\begin{aligned} \mu_{A,b,i,K} &= \mathbb{E}_X[p_{K,a} r_{b,K} - p_K \ell_{b,K,a}], \\ \mu_{H,b,i,K} &= \mathbb{E}_X[p_{K,aa} r_{b,K} - 2p_{K,a} \ell_{b,K,a} - p_K \ell_{b,K,aa}], \\ \mu_{T,b,i,K} &= \mathbb{E}_X[p_{K,aaa} r_{b,K} - 3p_{K,aa} \ell_{b,K,a} - 3p_{K,a} \ell_{b,K,aa} - p_K \ell_{b,K,aaa}], \end{aligned}$$

where all functions on the right are evaluated at (X, α_i) . Applying the weighted-mean bound term by term gives

$$\begin{aligned} \|\mu_{A,b,i,K}\|_2 &\leq C\{\lambda_K^{\tau_*-3/2} + \lambda_K^{-1/2}\} \leq C\lambda_K^{-1/2}, \\ \|\mu_{H,b,i,K}\|_2 &\leq C\{\lambda_K^{\tau_*-5/2} + \lambda_K^{-3/2} + \lambda_K^{-1/2}\} \leq C\lambda_K^{-3/2}, \\ \|\mu_{T,b,i,K}\|_2 &\leq C\{\lambda_K^{\tau_*-7/2} + \lambda_K^{-5/2} + \lambda_K^{-3/2} + \lambda_K^{-1/2}\} \leq C\lambda_K^{-5/2}. \end{aligned}$$

Stacking the two blocks proves (D.20).

Finally, boundedness of $R_{b,it,K}$ and of the first three derivatives of $\ell_{b,K}$, together with the first bound in (D.18), gives

$$\mathbb{E}_i \|A_{b,it,K}\|_2^2 \leq CK\lambda_K^{-2}, \quad \mathbb{E}_i \|H_{b,it,K}\|_2^2 \leq CK\lambda_K^{-4}, \quad \mathbb{E}_i \|T_{b,it,K}\|_2^2 \leq CK\lambda_K^{-6}.$$

Centering can only decrease each vector second moment. Stacking the blocks and averaging over i proves (D.19). $\square$

Define the unit averages $\bar{A}_{c,i,K} = T^{-1} \sum_t A_{c,it,K}$ and $\bar{H}_{c,i,K} = T^{-1} \sum_t H_{c,it,K}$, and set $\tilde{A}_{c,i,K} = \bar{A}_{c,i,K} - \mu_{A,c,i,K}$. The integral third-derivative remainder uses

$$\mathcal{T}_{c,iT,K}(d) = \frac{1}{T} \sum_t \int_0^1 (1-s)^2 T_{c,it,K}(\alpha_i + sd) ds.$$

Put $z_{c,it,K} = f_{c,it,K} - \mu_{f,c,i,K} + \mu_{A,c,i,K} \zeta_{it}$. Its panel mean is $\bar{Z}_{c,K} = (NT)^{-1} \sum_{i,t} z_{c,it,K}$. Since $\Delta_i = \frac{1}{T} \sum_t \zeta_{it}$,

$$\bar{Z}_{c,K} = \frac{1}{NT} \sum_{i,t} f_{c,it,K} + \frac{1}{N} \sum_i \mu_{A,c,i,K} \Delta_i. \quad (\text{D.21})$$

Thus the $\mu_A \Delta_i$ term is part of the leading influence and is not placed in a Taylor remainder.

Because $f_{c,it,K}$ and ζ_{is} are generally dependent, the exact clipped nuisance-bias coefficient must be centered under their joint law:

$$B_{c,K} = \frac{T}{N} \sum_i \mathbb{E}_i[d_i \bar{A}_{c,i,K}] + \frac{T}{2N} \sum_i \mathbb{E}_i[d_i^2 \bar{H}_{c,i,K}] + \frac{T}{2N} \sum_i \mathbb{E}_i[d_i^3 \mathcal{T}_{c,iT,K}(d_i)].$$

Define three remainders, each centered at its exact joint-law expectation:

$$\begin{aligned} R_{c,1,K} &= \frac{1}{N} \sum_i [d_i \bar{A}_{c,i,K} - \mu_{A,c,i,K} \Delta_i - \mathbb{E}_i\{d_i \bar{A}_{c,i,K} - \mu_{A,c,i,K} \Delta_i\}], \\ R_{c,2,K} &= \frac{1}{2N} \sum_i [d_i^2 \bar{H}_{c,i,K} - \mathbb{E}_i(d_i^2 \bar{H}_{c,i,K})], \\ R_{c,3,K} &= \frac{1}{2N} \sum_i [d_i^3 \mathcal{T}_{c,iT,K}(d_i) - \mathbb{E}_i\{d_i^3 \mathcal{T}_{c,iT,K}(d_i)\}], \end{aligned}$$

and let $R_{\alpha,c,K}$ be their sum. These exact joint centers are essential: in general $\mathbb{E}_i[d_i \bar{A}_{c,i,K}] \neq \mathbb{E}_i[d_i] \mu_{A,c,i,K}$.

For the target coefficients $c_K = (\gamma_K^\top, \delta_K^\top)^\top$, define the generated-coordinate nuisance moment

$$\hat{F}_{c,NT,K}(c_K) = \left(\frac{1}{NT} \sum_{i,t} \hat{p}_{it,K}(D_{it} - \hat{p}_{it,K}^\top \gamma_K) \right), \quad \hat{p}_{it,K} = p_K(X_{it}, \hat{\alpha}_i).$$

Lemma D.7 (Exact clipped nuisance expansion). *If $\lambda_K^{-m}/T \rightarrow 0$, then*

$$\hat{F}_{c,NT,K}(c_K) = \bar{Z}_{c,K} + \frac{1}{T} B_{c,K} + R_{\alpha,c,K}. \quad (\text{D.22})$$

Moreover,

$$\mathbb{E} \|\bar{Z}_{c,K}\|_2^2 \leq \frac{CK}{NT}. \quad (\text{D.23})$$

$$R_{\alpha,c,K} = o_p(\sqrt{K}/T). \quad (\text{D.24})$$

$$\|B_{c,K}\|_2 = O(\lambda_K^{-3/2}). \quad (\text{D.25})$$

Proof. Exact expansion and centering. For each nuisance block, Taylor's theorem along $a = \alpha_i + sd_i$ gives

$$\begin{aligned} f_{c,it,K}(\hat{\alpha}_i) &= f_{c,it,K} + d_i A_{c,it,K} + \frac{d_i^2}{2} H_{c,it,K} \\ &\quad + \frac{d_i^3}{2} \int_0^1 (1-s)^2 T_{c,it,K}(\alpha_i + sd_i) ds. \end{aligned}$$

After averaging over t , subtract the leading coordinate term $\mu_{A,c,i,K} \Delta_i$ and use $d_i = \Delta_i + c_i$: $d_i \bar{A}_{c,i,K} - \mu_{A,c,i,K} \Delta_i = d_i \tilde{A}_{c,i,K} + \mu_{A,c,i,K} c_i$. For each of these three quantities, add and subtract its exact conditional expectation. Because $\mathbb{E}_i \Delta_i = 0$, $\mathbb{E}_i[d_i \bar{A}_{c,i,K} - \mu_{A,c,i,K} \Delta_i] = \mathbb{E}_i[d_i \tilde{A}_{c,i,K}]$. Thus the uncentered expectations sum to $\frac{1}{T} B_{c,K}$, while the centered parts are exactly $R_{c,1,K}$, $R_{c,2,K}$, and $R_{c,3,K}$. Equation (D.21) supplies the remaining leading term, so the resulting identity is (D.22).

Leading moments. The six fixed-coordinate bounds needed below are (D.19) and (D.20) from Lemma D.6. The $r = 0$ case of (D.18), bounded residuals, and centering give

$$\begin{aligned} \sup_i \mathbb{E}_i \|f_{c,it,K} - \mu_{f,c,i,K}\|_2^2 &\leq CK, \\ \sup_i \mathbb{E}_i \|\mu_{A,c,i,K} \zeta_{it}\|_2^2 &\leq C \lambda_K^{-1} \leq CK. \end{aligned}$$

Consequently, $\sup_i \mathbb{E}_i \|z_{c,it,K}\|_2^2 \leq CK$. To verify (D.23), expand the second moment of $\bar{Z}_{c,K}$. Cross-unit covariance terms vanish by unit independence, and within a unit finite-range dependence leaves only $O(T)$ nonzero pairs. Thus the vector second moment satisfies

$$\mathbb{E} \|\bar{Z}_{c,K}\|_2^2 = \frac{1}{N^2 T^2} \sum_{i=1}^N \sum_{t=1}^T \sum_{s=1}^T \mathbb{E}_i [z_{c,it,K}^\top z_{c,is,K}] \leq \frac{CK}{NT}.$$

This is also the conclusion obtained from Lemma C.1.

Joint derivative and curvature centers. The relevant joint centers need not vanish. Put $a_{c,i,K}(X) = (p_{K,a}(X, \alpha_i)^\top, \theta_0 p_{K,a}(X, \alpha_i)^\top)^\top$. In the derivative table, the coefficients of v_{it} in the gamma and delta blocks are $p_{K,a}$ and $\theta_0 p_{K,a}$, respectively. Terms that depend only on X factor from the centered ζ , and the independent outcome-error terms also have zero expectation. Hence independence of X from the joint innovation process gives

$$B_{\zeta,c,K} := \sum_{|j| \leq L} \frac{1}{N} \sum_i \mathbb{E}_i [A_{c,i0,K} \zeta_{ij}] = \Omega_{v\zeta} \frac{1}{N} \sum_i \mathbb{E}_X a_{c,i,K}(X).$$

Tensor local support gives $\|\mathbb{E}_X p_{K,a}(X, a)\|_2 \leq C\lambda_K^{-3/2}$. Hence,

$$\|B_{\zeta,c,K}\|_2 \leq C\lambda_K^{-3/2}.$$

The coefficient of v_{it} in the gamma and delta Hessian blocks is, respectively, $p_{K,aa}$ and $\theta_0 p_{K,aa}$. Condition (5.2), after a time shift, removes these two $v\zeta\zeta$ cross terms. The independent, centered outcome error removes the corresponding $\varepsilon\zeta\zeta$ terms. Every remaining Hessian term is a function of X times a disturbance factor independent of X , so

$$\mathbb{E}_i [H_{c,it,K} \zeta_{is} \zeta_{ir}] = \mu_{H,c,i,K} \mathbb{E}_i [\zeta_{is} \zeta_{ir}].$$

To count the surviving pairs, write

$$\begin{aligned} T \mathbb{E}_i [\Delta_i \bar{A}_{c,i,K}] &= \frac{1}{T} \sum_{s=1}^T \sum_{t=1}^T \mathbb{E}_i [A_{c,it,K} \zeta_{is}], \\ \frac{T}{2} \mathbb{E}_i [\Delta_i^2 \bar{H}_{c,i,K}] &= \frac{1}{2T^2} \sum_{r=1}^T \sum_{s=1}^T \sum_{t=1}^T \mathbb{E}_i [H_{c,it,K} \zeta_{is} \zeta_{ir}]. \end{aligned}$$

For a lag j , exactly $T - |j|$ pairs remain inside the sample. Using the factorization above therefore yields

$$\begin{aligned} \frac{T}{N} \sum_i \mathbb{E}_i [\Delta_i \bar{A}_{c,i,K}] &= B_{\zeta,c,K} - \frac{1}{TN} \left\{ \sum_{|j| \leq L} |j| \gamma_{v\zeta}(j) \right\} \sum_i \mathbb{E}_X a_{c,i,K}(X), \\ \frac{T}{2N} \sum_i \mathbb{E}_i [\Delta_i^2 \bar{H}_{c,i,K}] &= \frac{\Omega_\zeta}{2N} \sum_i \mu_{H,c,i,K} - \frac{1}{2TN} \left\{ \sum_{|j| \leq L} |j| \gamma_\zeta(j) \right\} \sum_i \mu_{H,c,i,K}. \end{aligned} \tag{D.26}$$

Both finite- T endpoint terms are $O(\lambda_K^{-3/2}/T)$, while the two leading centers are $O(\lambda_K^{-3/2})$. Thus the joint derivative expectation need not vanish, but it has the order required by the nuisance rate: after the score's T^{-1} scaling, $\frac{1}{T} O(\lambda_K^{-3/2}) = O(r_K)$ because $m = \max\{d_W, 3\} \geq 3$.

Replacing Δ_i by d_i in the first two lines costs at most

$$\begin{aligned} \left\| \frac{T}{N} \sum_i \mathbb{E}_i[(d_i - \Delta_i) \bar{A}_{c,i,K}] \right\|_2 &\leq CT \lambda_K^{-3/2} b_T N^{-a_0}, \\ \left\| \frac{T}{2N} \sum_i \mathbb{E}_i[(d_i^2 - \Delta_i^2) \bar{H}_{c,i,K}] \right\|_2 &\leq CT \lambda_K^{-5/2} b_T^2 N^{-a_0}. \end{aligned}$$

After division by $\lambda_K^{-3/2}$, these are bounded by $T b_T N^{-a_0} = o(1)$ and $(b_T/\lambda_K) T b_T N^{-a_0} = o(1)$.

Third-order path and centered fluctuations. For the third-order path, separate its fixed-coordinate mean, centered fixed-coordinate average, and path displacement. On each knot-cell interior, differentiation of the third row in (D.17) gives, block by block,

$$\begin{aligned} \partial_{a,\text{pw}} T_{it,K}^b &= p_{K,aaaa} R_{b,it,K} - 4p_{K,aaa} \ell_{b,K,a} - 6p_{K,aa} \ell_{b,K,aa} \\ &\quad - 4p_{K,a} \ell_{b,K,aaa} - p_K \ell_{b,K,aaaa}. \end{aligned}$$

Here D_{it} and Y_{it} are held fixed along the path, so $\partial_a R_{b,it,K} = -\ell_{b,K,a}$. The first three derivatives of $\ell_{b,K}$ are uniformly bounded by derivative approximation, while (D.10) gives $\|\ell_{b,K,aaaa}\|_{\infty,\text{pw}} \leq C \lambda_K^{-4}$. Let $\mathcal{F}_{\text{dist},T} = \sigma\{(\zeta_{is}, v_{is}, \varepsilon_{is}) : 1 \leq i \leq N, 1 \leq s \leq T\}$. For every fixed realization of this bounded disturbance array, (D.18) applied to the five terms yields

$$\text{ess sup}_a \left\| \mathbb{E}[\partial_{a,\text{pw}} T_{it,K}^b(a) \mid \mathcal{F}_{\text{dist},T}] \right\|_2 \leq C \lambda_K^{-9/2}.$$

The path array $T_{it,K}^b(a)$ is continuous because all spline derivatives through order three agree across knots. Partitioning any path segment at its intervening knots and integrating the preceding cellwise derivative therefore gives the global bound

$$\left\| \mathbb{E}[T_{it,K}^b(a') - T_{it,K}^b(a) \mid \mathcal{F}_{\text{dist},T}] \right\|_2 \leq C |a' - a| \lambda_K^{-9/2}. \quad (\text{D.27})$$

Thus Lemma D.6, (D.27), and Lemma D.5 give

$$T \left\| \frac{1}{N} \sum_i \mathbb{E}_i[d_i^3 \mathcal{T}_{c,iT,K}(d_i)] \right\|_2 \leq C \left\{ \lambda_K^{-5/2} T^{-1/2} + \frac{\sqrt{K} \lambda_K^{-3}}{T} + \frac{\lambda_K^{-9/2}}{T} \right\} + o(\lambda_K^{-3/2}) = o(\lambda_K^{-3/2}). \quad (\text{D.28})$$

The three displayed terms follow by first bounding the fixed mean by $C \lambda_K^{-5/2} \mathbb{E}|d_i|^3$. The centered average has second moment $CK \lambda_K^{-6}/T$. The path displacement is bounded by $C|d_i| \lambda_K^{-9/2}$ by (D.27). Multiplying by $T \mathbb{E}|d_i|^3$, $T(\mathbb{E}|d_i|^6)^{1/2}$, and $T \mathbb{E}|d_i|^4$, respectively, gives the three terms in (D.28). Relative to $\lambda_K^{-3/2}$, their exponents are $\lambda_K^{-1} T^{-1/2}$, $\lambda_K^{-(d_W+3)/2}/T$, and λ_K^{-3}/T . All vanish because $(d_W+3)/2 \leq m$ and $\lambda_K^{-m}/T \rightarrow 0$. Replacing Δ_i^3 by d_i^3 in the centered fixed-coordinate term adds at most $C \sqrt{K} \lambda_K^{-3} T^{1/2} b_T^3 N^{-a_0/2} = o(\lambda_K^{-3/2})$ by the clipped-tail moment and $b_T/\lambda_K \rightarrow 0$. Equations (D.26)–(D.28) prove (D.25).

We finally control the centered fluctuations. Define on the full sample space

$$\begin{aligned} S_{c,1,K} &= \frac{1}{N} \sum_i \left[\Delta_i \tilde{A}_{c,i,K} - \mathbb{E}_i(\Delta_i \tilde{A}_{c,i,K}) \right], \\ S_{c,2,K} &= \frac{1}{2N} \sum_i \left[\Delta_i^2 \bar{H}_{c,i,K} - \mathbb{E}_i(\Delta_i^2 \bar{H}_{c,i,K}) \right]. \end{aligned}$$

On $\mathcal{E}_T$, the random, uncentered terms agree because $d_i = \Delta_i$. Using $c_i = d_i - \Delta_i$, the remaining exact center differences are

$$\begin{aligned} R_{c,1,K} - S_{c,1,K} &= -\frac{1}{N} \sum_i \mathbb{E}_i[c_i \bar{A}_{c,i,K}], \\ R_{c,2,K} - S_{c,2,K} &= -\frac{1}{2N} \sum_i \mathbb{E}_i[(d_i^2 - \Delta_i^2) \bar{H}_{c,i,K}]. \end{aligned}$$

Conditional on the disturbances, the weighted-mean bound in (D.18) gives orders $C\lambda_K^{-3/2}$ and $C\lambda_K^{-5/2}$ for the two averaged arrays. The clipped-tail moments from Lemma D.5 therefore imply, on $\mathcal{E}_T$,

$$\begin{aligned} \|R_{c,1,K} - S_{c,1,K}\|_2 &\leq C\lambda_K^{-3/2} b_T N^{-a_0}, \\ \|R_{c,2,K} - S_{c,2,K}\|_2 &\leq C\lambda_K^{-5/2} b_T^2 N^{-a_0}. \end{aligned}$$

The finite-range product bound in Lemma C.1 and cross-unit independence give

$$\begin{aligned} \mathbb{E} \|S_{c,1,K}\|_2^2 &\leq \frac{CK\lambda_K^{-2}}{NT^2}, \\ \mathbb{E} \|S_{c,2,K}\|_2^2 &\leq \frac{CK\lambda_K^{-4}}{NT^3} + \frac{C\lambda_K^{-3}}{NT^2}, \\ \mathbb{E} \|R_{c,3,K}\|_2^2 &\leq \frac{CK\lambda_K^{-6}}{NT^3}. \end{aligned} \tag{D.29}$$

In the second line, the two terms correspond, respectively, to $\Delta_i^2(\bar{H}_{c,i,K} - \mu_{H,c,i,K})$ and the centered product $\mu_{H,c,i,K}\{\Delta_i^2 - \mathbb{E}_i \Delta_i^2\}$. The third line uses the global local-support envelope $\|\mathcal{T}_{c,iT,K}(d_i)\|_2 \leq C\sqrt{K}\lambda_K^{-3}$ and $\mathbb{E}|d_i|^6 \leq CT^{-3}$. To verify the comparison with $(\sqrt{K}/T)^2 = K/T^2$ term by term, divide the four variance orders in (D.29) by K/T^2 :

$$\begin{aligned} \frac{K\lambda_K^{-2}/(NT^2)}{K/T^2} &= \frac{\lambda_K^{-2}}{N}, & \frac{K\lambda_K^{-4}/(NT^3)}{K/T^2} &= \frac{\lambda_K^{-4}}{NT}, \\ \frac{\lambda_K^{-3}/(NT^2)}{K/T^2} &= \frac{\lambda_K^{-3}}{NK}, & \frac{K\lambda_K^{-6}/(NT^3)}{K/T^2} &= \frac{\lambda_K^{-6}}{NT}. \end{aligned}$$

Each ratio vanishes because $N/T \rightarrow \kappa_0$, $K \asymp \lambda_K^{-d_W}$, $m = \max\{d_W, 3\}$, and $\lambda_K^{-m}/T \rightarrow 0$. The two clipping centers are smaller by $a_0 > 1$ and $b_T/\lambda_K \rightarrow 0$, while $\Pr(\mathcal{E}_T^c) \rightarrow 0$. This proves (D.24). $\square$

Use the candidate coefficients, residuals, and Gram event defined in Appendix A.2. They satisfy the affine normal equations on $\mathcal{H}_K$ before the slope acceptance test is imposed.

Gram-Matrix Concentration Define the panel-grid population Gram, oracle sample Gram, and feasible sample Gram by

$$\begin{aligned} Q_K &= \frac{1}{N} \sum_{i=1}^N \mathbb{E}_X[p_K(X, \alpha_i)p_K(X, \alpha_i)^\top], \\ \tilde{Q}_K &= \frac{1}{NT} \sum_{i,t} p_K(X_{it}, \alpha_i)p_K(X_{it}, \alpha_i)^\top, \\ \hat{Q}_K &= \frac{1}{NT} \sum_{i,t} p_K(X_{it}, \hat{\alpha}_i)p_K(X_{it}, \hat{\alpha}_i)^\top. \end{aligned}$$

Lemma D.8 (Population and oracle-sample Gram). *There are constants $0 < c < C < \infty$ such that, for all sufficiently large N, T, K ,*

$$cI_K \preceq Q_K \preceq CI_K. \quad (\text{D.30})$$

Moreover,

$$\left\| \tilde{Q}_K - Q_K \right\|_{\text{op}} = O_p \left(\sqrt{\frac{K \log K}{NT}} + \frac{K \log K}{NT} \right) = o_p(1). \quad (\text{D.31})$$

For $j = 1, 2, 3$, the continuous derivative Gram $G_{aj,K} = \int p_{K,a^j} p_{K,a^j}^\top$ satisfies $\lambda_{\max}(G_{aj,K}) \leq C \lambda_K^{-2j}$. No positive lower bound is asserted for a derivative Gram on the full coefficient space.

Proof. Deterministic spline geometry controls the population eigenvalues, and matrix concentration transfers those bounds to the oracle sample Gram. The population Gram is G_g from Section D.1. The continuous Riesz inequality (D.1), the Loewner comparison (D.4), and $\varepsilon_{N,T,K} = o(1)$ prove (D.30). The derivative claim follows by taking the supremum over $\|b\|_2 = 1$ in (D.2).

Conditional on the deterministic grid, the oracle summands are independent over (i, t) . Local support and L^2 normalization imply $\|p_K(X_{it}, \alpha_i)\|_2^2 \leq CK$. Set $p_{it}^0 = p_K(X_{it}, \alpha_i)$, $A_{it}^0 = p_{it}^0 (p_{it}^0)^\top$, $M_i^0 = \mathbb{E}_X A_{it}^0$, and $Y_{it}^0 = (A_{it}^0 - M_i^0)/(NT)$. Then $\sum_{i,t} Y_{it}^0 = \tilde{Q}_K - Q_K$, and the Y_{it}^0 are independent, self-adjoint, and mean zero. Since $\|A_{it}^0\|_{\text{op}} \leq CK$ and $\|M_i^0\|_{\text{op}} \leq CK$,

$$\|Y_{it}^0\|_{\text{op}} \leq \frac{CK}{NT}.$$

The crucial variance calculation is made only after centering and summing. Because $(A_{it}^0)^2 = \|p_{it}^0\|_2^2 A_{it}^0 \preceq CK A_{it}^0$,

$$\begin{aligned} \sum_{i,t} \mathbb{E}_X [(Y_{it}^0)^2] &= \frac{1}{(NT)^2} \sum_{i,t} \{\mathbb{E}_X [(A_{it}^0)^2] - (M_i^0)^2\} \\ &\preceq \frac{CK}{(NT)^2} \sum_{i,t} M_i^0 = \frac{CK}{NT} Q_K \preceq \frac{CK}{NT} I_K. \end{aligned}$$

Thus the matrix-variance proxy is $CK/(NT)$. The bound uses the averaged population Gram matrix $N^{-1} \sum_i M_i^0 = Q_K$; individual M_i^0 need not be uniformly bounded. Matrix Bernstein's inequality therefore yields

$$\Pr \left(\left\| \tilde{Q}_K - Q_K \right\|_{\text{op}} > C \left\{ \sqrt{\frac{K \log K}{NT}} + \frac{K \log K}{NT} \right\} \right) \rightarrow 0.$$

This is (D.31); its right-hand side is $o(1)$ by (D.16). $\square$

Lemma D.9 (Generated-coordinate Gram perturbation). *The clipped feasible Gram obeys*

$$\left\| \hat{Q}_K - \tilde{Q}_K \right\|_{\text{op}} = O_p(b_T/\lambda_K) + o_p(1). \quad (\text{D.32})$$

Consequently, $\left\| \hat{Q}_K - Q_K \right\|_{\text{op}} = o_p(1)$.

Proof. The proof separates the deterministic coordinate displacement from sample fluctuation in X . Write $\Delta = (\Delta_1, \dots, \Delta_N)^\top$. On $\mathcal{E}_T$, Lemma D.5 gives $\hat{\alpha}_i = \alpha_i + \Delta_i$ and $|\Delta_i| \leq b_T$. For $s_b(x, a) = b^\top p_K(x, a)$, the fundamental theorem of calculus gives

$$s_b(x, \hat{\alpha}_i) - s_b(x, \alpha_i) = d_i \int_0^1 \partial_a s_b(x, \alpha_i + sd_i) ds.$$

The quadratic-form difference is therefore $s_b(x, \hat{\alpha}_i)^2 - s_b(x, \alpha_i)^2 = \{s_b(x, \hat{\alpha}_i) - s_b(x, \alpha_i)\} \times \{s_b(x, \hat{\alpha}_i) + s_b(x, \alpha_i)\}$. Cauchy–Schwarz along the Taylor path gives, uniformly over $\|b\|_2 = 1$,

$$\begin{aligned} \left| b^\top \{\mathbb{E}(\hat{Q}_K \mid \Delta) - Q_K\} b \right| &\leq b_T \left\{ \frac{1}{N} \sum_i \int_0^1 \mathbb{E}_X[\{\partial_a s_b(X, \alpha_i + s d_i)\}^2] ds \right\}^{1/2} \\ &\quad \times \left\{ \frac{1}{N} \sum_i \mathbb{E}_X[\{s_b(X, \hat{\alpha}_i) + s_b(X, \alpha_i)\}^2] \right\}^{1/2}. \end{aligned}$$

The shifted-grid occupancy argument below, the derivative inverse inequality, and the order-zero Riesz bound make the two squared energies $O(\lambda_K^{-2} \|b\|_2^2)$ and $O(\|b\|_2^2)$, respectively. Hence

$$\sup_{\|b\|_2=1} \left| b^\top \{\mathbb{E}(\hat{Q}_K \mid \Delta) - Q_K\} b \right| \leq C \frac{b_T}{\lambda_K},$$

on $\mathcal{E}_T$. Indeed, if a shifted point lies in an interval J of length $O(\lambda_K)$, then its original coordinate lies in the b_T -enlargement of J . The lower spacing bound in (5.1) therefore limits the number of such units to $O\{N(\lambda_K + b_T) + 1\} = O(N\lambda_K)$. Thus the same finite-overlap integrals used in Section D.1 apply along the entire Taylor path.

Because the generated coordinates depend only on the Z process and this process is independent of X , conditional on the generated coordinates the X_{it} remain independent. Set $\hat{A}_{it} = \hat{p}_{it,K} \hat{p}_{it,K}^\top$, $\widehat{M}_i = \mathbb{E}_X(\hat{A}_{it} \mid \Delta)$, and $\hat{Y}_{it} = (\hat{A}_{it} - \widehat{M}_i)/(NT)$. Conditional on Δ , the $\hat{Y}_{it}$ are independent, self-adjoint, and mean zero, and local support gives $\|\hat{Y}_{it}\|_{\text{op}} \leq CK/(NT)$. On $\mathcal{E}_T$, the mean shift bound above and (D.30) imply

$$\|\mathbb{E}(\hat{Q}_K \mid \Delta)\|_{\text{op}} \leq C + C \frac{b_T}{\lambda_K} \leq C.$$

Moreover, using $\hat{A}_{it}^2 \preceq CK \hat{A}_{it}$ and centering before summing,

$$\begin{aligned} \sum_{i,t} \mathbb{E}(\hat{Y}_{it}^2 \mid \Delta) &= \frac{1}{(NT)^2} \sum_{i,t} \{\mathbb{E}_X(\hat{A}_{it}^2 \mid \Delta) - \widehat{M}_i^2\} \\ &\preceq \frac{CK}{(NT)^2} \sum_{i,t} \widehat{M}_i = \frac{CK}{NT} \mathbb{E}(\hat{Q}_K \mid \Delta). \end{aligned}$$

Thus, uniformly on $\mathcal{E}_T$, the conditional variance proxy is $CK/(NT)$. Conditional matrix Bernstein therefore gives

$$\|\hat{Q}_K - \mathbb{E}(\hat{Q}_K \mid \Delta)\|_{\text{op}} = O_p \left(\sqrt{\frac{K \log K}{NT}} + \frac{K \log K}{NT} \right) = o_p(1)$$

with constants uniform over the realized coordinates in $\mathcal{E}_T$. The three pieces enter through the exact decomposition

$$\begin{aligned} \hat{Q}_K - \tilde{Q}_K &= \{\hat{Q}_K - \mathbb{E}(\hat{Q}_K \mid \Delta)\} - \{\tilde{Q}_K - Q_K\} \\ &\quad + \{\mathbb{E}(\hat{Q}_K \mid \Delta) - Q_K\}. \end{aligned}$$

Integrating the conditional probability bound and adding $\Pr(\mathcal{E}_T^c) = o(1)$ proves (D.32). The final claim follows from Lemma D.8 and $b_T/\lambda_K \rightarrow 0$. $\square$

Lemma D.10 (Eigenvalue bounds for the estimated Gram matrix). *With the constants in (D.30),*

$$\Pr \left(\frac{c}{2} I_K \preceq \hat{Q}_K \preceq 2C I_K \right) \rightarrow 1.$$

Proof. By Lemmas D.8 and D.9, the feasible Gram differs from Q_K by $o_p(1)$ in operator norm. Weyl's inequality and (D.30) prove the result. $\square$

Nuisance Coefficient Rate

Lemma D.11 (Nuisance coefficient rate). *Under the preceding assumptions and $\lambda_K^{-m}/T \rightarrow 0$, $\Pr(\mathcal{H}_K^c) \rightarrow 0$ and, on $\mathcal{H}_K$,*

$$\|\check{\gamma}_K - \gamma_K\|_2 + \|\check{\delta}_K - \delta_K\|_2 = O_p(r_K). \quad (\text{D.33})$$

$$\sqrt{NT} (r_K)^2 = o(1). \quad (\text{D.34})$$

Proof. The proof first converts the two affine normal equations into a single coefficient identity and then bounds its three score components. The event $\mathcal{H}_K$ has probability approaching one by Lemma D.10. On this event, subtracting the target coefficients from the two affine normal equations gives

$$\begin{pmatrix} \check{\gamma}_K - \gamma_K \\ \check{\delta}_K - \delta_K \end{pmatrix} = \begin{pmatrix} \hat{Q}_K^{-1} & 0 \\ 0 & \hat{Q}_K^{-1} \end{pmatrix} \hat{F}_{c,NT,K}(c_K).$$

The inverse block matrix has bounded operator norm on $\mathcal{H}_K$. By Lemma D.7, the three terms on the right satisfy

$$\begin{aligned} \|\bar{Z}_{c,K}\|_2 &= O_p\left(\sqrt{\frac{K}{NT}}\right) = O_p\left(\frac{\sqrt{K}}{T}\right), \\ \frac{1}{T} \|B_{c,K}\|_2 &= O\left(\frac{\lambda_K^{-3/2}}{T}\right), \\ \|R_{\alpha,c,K}\|_2 &= o_p\left(\frac{\sqrt{K}}{T}\right). \end{aligned}$$

Because $K \asymp \lambda_K^{-d_W}$ and $m = \max\{d_W, 3\}$, both $\sqrt{K}/T$ and $\lambda_K^{-3/2}/T$ are bounded by a constant multiple of r_K . This proves (D.33). Finally, $\sqrt{NT} (r_K)^2 \asymp \lambda_K^{-m}/T$, which proves (D.34). $\square$

Derivative Gram Bounds Write $p_{K,a^r} = \partial_a^r p_K$ for the derivative-basis array.

Lemma D.12 (Generated-coordinate derivative Grams). *Under the maintained sample-mean conditions, for $r = 1, 2$,*

$$\lambda_{\max} \left\{ \frac{1}{NT} \sum_{i,t} p_{K,a^r}(X_{it}, \hat{\alpha}_i) p_{K,a^r}(X_{it}, \hat{\alpha}_i)^\top \right\} = O_p(\lambda_K^{-2r}). \quad (\text{D.35})$$

Proof. On $\mathcal{E}_T$, $\max_i |\hat{\alpha}_i - \alpha_i| \leq b_T = o(\lambda_K)$. Quasi-uniformity therefore puts at most $CN\lambda_K$ shifted coordinates in each interval of length $O(\lambda_K)$. The cellwise inverse inequality and continuous Riesz bound imply that the conditional mean of the Gram is at most $C\lambda_K^{-2r} I_K$. Conditioning on the innovation histories leaves the X_{it} independent under the sample-mean benchmark. For $A_{it} = p_{K,a^r}(X_{it}, \hat{\alpha}_i) p_{K,a^r}(X_{it}, \hat{\alpha}_i)^\top$, local support gives $\|A_{it}\| \leq CK\lambda_K^{-2r}$ and $A_{it}^2 \preceq CK\lambda_K^{-2r} A_{it}$.

The variance bound follows by centering each independent matrix and using

$$\begin{aligned} \frac{1}{(NT)^2} \sum_{i,t} \mathbb{E}_X[(A_{it} - \mathbb{E}_X A_{it})^2 \mid \Delta] &\preceq \frac{CK\lambda_K^{-2r}}{NT} \frac{1}{NT} \sum_{i,t} \mathbb{E}_X[A_{it} \mid \Delta] \\ &\preceq \frac{CK\lambda_K^{-4r}}{NT} I_K. \end{aligned}$$

Thus the normalized envelope and variance proxy are $CK\lambda_K^{-2r}/(NT)$ and $CK\lambda_K^{-4r}/(NT)$. Matrix Bernstein [Tro12] gives

$$\left\| \frac{1}{NT} \sum_{i,t} \{A_{it} - \mathbb{E}_X[A_{it} \mid \Delta]\} \right\|_{\text{op}} = O_p \left[\lambda_K^{-2r} \left\{ \sqrt{\frac{K \log K}{NT}} + \frac{K \log K}{NT} \right\} \right] = o_p(\lambda_K^{-2r}).$$

The conditional mean bound and $\mathbb{P}(\mathcal{E}_T^c) = o(1)$ prove the claim. This argument is specific to the sample-mean coordinate; the smooth-moment case uses the uniform-shift Gram bound in Appendix G. $\square$

Denominator Control

Lemma D.13 (Consistency of the residual second moment). *On $\mathcal{H}_K$, let $d_{\alpha,it,K} = m_K(X_{it}, \hat{\alpha}_i) - m_K(X_{it}, \alpha_i)$. The affine normal equation gives the exact identity*

$$\begin{aligned} \check{J}_K - J_K &= \left\{ \frac{1}{NT} \sum_{i,t} u_{D,it,K}^2 - J_K \right\} - 2 \frac{1}{NT} \sum_{i,t} u_{D,it,K} d_{\alpha,it,K} + \frac{1}{NT} \sum_{i,t} d_{\alpha,it,K}^2 \\ &\quad - (\check{\gamma}_K - \gamma_K)^\top \hat{Q}_K (\check{\gamma}_K - \gamma_K). \end{aligned} \tag{D.36}$$

Under $\lambda_K^{-m}/T \rightarrow 0$,

$$\check{J}_K - J_K = O_p((NT)^{-1/2} + T^{-1/2} + r_K^2) = o_p(1). \tag{D.37}$$

Proof. Write $\tilde{u}_{D,it,K} = u_{D,it,K} - d_{\alpha,it,K}$ and $\check{u}_{D,it} = \tilde{u}_{D,it,K} - \hat{p}_{it,K}^\top (\check{\gamma}_K - \gamma_K)$. Expanding the square around the fitted residual gives

$$\begin{aligned} \frac{1}{NT} \sum_{i,t} \{(\check{u}_{D,it})^2 - (\tilde{u}_{D,it,K})^2\} &= -2(\check{\gamma}_K - \gamma_K)^\top \frac{1}{NT} \sum_{i,t} \hat{p}_{it,K} \tilde{u}_{D,it} \\ &\quad - (\check{\gamma}_K - \gamma_K)^\top \hat{Q}_K (\check{\gamma}_K - \gamma_K) \\ &= -(\check{\gamma}_K - \gamma_K)^\top \hat{Q}_K (\check{\gamma}_K - \gamma_K). \end{aligned}$$

The cross term vanishes by the normal equation. Expanding $(\tilde{u}_{D,it,K})^2 = (u_{D,it,K} - d_{\alpha,it,K})^2$ proves the stated identity. The oracle fluctuation is $O_p((NT)^{-1/2})$ by Lemma C.1. Bounded $m_{K,a}$ and the clipped-coordinate second moment give $\|d_\alpha\|_{NT,2} = O_p(T^{-1/2})$. Therefore

$$\begin{aligned} \left| \frac{1}{NT} \sum_{i,t} u_{D,it,K} d_{\alpha,it,K} \right| &\leq \|u_D\|_{NT,2} \|d_\alpha\|_{NT,2} = O_p(T^{-1/2}), \\ \frac{1}{NT} \sum_{i,t} d_{\alpha,it,K}^2 &= O_p(T^{-1}), \\ |(\check{\gamma}_K - \gamma_K)^\top \hat{Q}_K (\check{\gamma}_K - \gamma_K)| &\leq \|\hat{Q}_K\|_{\text{op}} \|\check{\gamma}_K - \gamma_K\|^2 = O_p(r_K^2). \end{aligned}$$

The last bound uses Lemmas D.11 and D.10. These bounds prove (D.37). $\square$

Lemma D.14 (Oracle denominator). *The finite-sieve denominator satisfies $J_K = \sigma_v^2 + O(\lambda_K^{2\tau_m})$. Hence there is a constant $c_V > 0$ such that $J_K \geq c_V$ for all sufficiently large N, T, K .*

Proof. The partially linear equation and the grid-projection definition give $u_{D,it,K} = v_{it} + r_{m,it,K}$. Since v_{it} is centered and independent of X_{it} and the deterministic grid, $J_K = \sigma_v^2 + \frac{1}{NT} \sum_{i,t} \mathbb{E}_i[r_{m,it,K}^2]$. Lemma D.4 bounds the last term by $C\lambda_K^{2\tau_m}$. Because $\sigma_v^2 > 0$, the lower bound follows for a fixed $c_V < \sigma_v^2$ and all sufficiently large indices. $\square$

D.5 Estimator Localization and Exact Identities

Use the candidates, fixed thresholds, extensions, and reporting rule in Appendix A.2. The same construction applies to the raw coordinate for the fixed- K result and to the clipped coordinate for the growing-sieve result. Its affine normal equations hold on $\mathcal{H}_K$; on $\mathcal{G}_K$ the slope equation also holds. Consequently the exact identities of Lemma A.2 apply on acceptance.

Assumption D.15 (High-level localization inputs). $\mathbb{P}(\mathcal{H}_K^c) \rightarrow 0$, $J_K \geq c_V > 0$, and $\mathbb{P}(|\check{J}_K - J_K| > \varepsilon, \mathcal{H}_K) \rightarrow 0$ for every $\varepsilon > 0$.

Proposition D.16 (Acceptance probability). *Under Assumption D.15, $\mathbb{P}(\mathcal{G}_K^c) \rightarrow 0$.*

Proof. On $\mathcal{H}_K$, failure of the denominator test implies $|\check{J}_K - J_K| \geq c_V/2$. Therefore

$$\mathbb{P}(\mathcal{G}_K^c) \leq \mathbb{P}(\mathcal{H}_K^c) + \mathbb{P}(|\check{J}_K - J_K| \geq c_V/2, \mathcal{H}_K) \rightarrow 0.$$

$\square$

Corollary D.17 (Primitive localization inputs). *Under the maintained sample-mean benchmark, $\lambda_K^{-m}/T \rightarrow 0$ and the thresholds in (A.2) imply Assumption D.15 and*

$$\mathbb{P}(\mathcal{G}_K^c) \rightarrow 0. \tag{D.38}$$

Proof. Lemma D.10 proves the Gram input. The affine-root rate gives $r_K^2 = o(1)$, so Lemma D.13 proves denominator consistency. Lemma D.14 gives $J_K \geq c_V$ eventually. Proposition D.16 then proves acceptance. $\square$

Remark D.18 (Exact expectation and acceptance-event transfer). *The corrected identity uses the exact coefficient $\mathcal{B}_{KT}$. Its replacement by B_K is justified by the separate expectation bridge in Appendix E.2. For any probability remainder proved on an acceptance event $\mathcal{A}_{NT}$ with $\mathbb{P}(\mathcal{A}_{NT}^c) \rightarrow 0$,*

$$\mathbb{P}(|R_{NT}| > \varepsilon) \leq \mathbb{P}(\mathcal{A}_{NT}^c) + \mathbb{P}(|R_{NT}| > \varepsilon, \mathcal{A}_{NT}).$$

This transfers probability orders and limiting distributions to the reported procedure. It gives no expectation bound; exceptional-event expectations are controlled separately by their envelopes and tail moments.

E Score Expansion, Bias Correction, and Limit Theory

E.1 Sieve Approximation Bias and Scalar Score Expansion

Maintain Assumption 5.1. Write $d_i = \hat{\alpha}_i - \alpha_i$, $\Delta_i = T^{-1} \sum_t \zeta_{it}$, and $c_i = d_i - \Delta_i$. Lemma D.5 gives, uniformly in i ,

$$\mathbb{E}_i |d_i|^r \leq C_r T^{-r/2}, \quad \mathbb{E}_i |c_i|^r \leq C_r b_T^r N^{-a_0}, \quad r \leq 6, \tag{E.1}$$

where $a_0 > 1$ and $b_T = C_b \sqrt{\log N/T}$. The oracle score and its first three coordinate derivatives are uniformly bounded by Lemma D.4. Recall $x_{it,K} = \psi_{it,K} - \mathbb{E}_i[\psi_{i0,K}] + \mu_{iK} \zeta_{it}$.

Proposition E.1 (Exact score expansion with a centered total remainder). *The exact clipped-coordinate drift and the score remainder satisfy*

$$\frac{1}{NT} \sum_{i,t} \psi_{it,K}(\hat{\alpha}_i) = \frac{1}{NT} \sum_{i,t} x_{it,K} + \frac{\mathcal{B}_{KT}}{T} + r_{\theta,F,K}, \quad \sqrt{NT} r_{\theta,F,K} = o_p(1). \quad (\text{E.2})$$

Here $\mathcal{B}_{KT}$ is the exact expectation in (A.6).

Proof. Define one unit remainder, $U_{iK} = \bar{\psi}_{iK}(\hat{\alpha}_i) - \bar{\psi}_{iK}(\alpha_i) - \mu_{iK}\Delta_i$. Taylor's theorem and $d_i = \Delta_i + c_i$ give

$$U_{iK} = \Delta_i(\bar{\psi}_{iK}^{(1)} - \mu_{iK}) + c_i\bar{\psi}_{iK}^{(1)} + \frac{d_i^2}{2}\bar{\psi}_{iK}^{(2)} + R_{iK}, \quad |R_{iK}| \leq C|d_i|^3.$$

The finite-range mean bound in Lemma C.1, Hölder's inequality, and (E.1) imply

$$\begin{aligned} \|U_{iK}\|_{L_i^2} &\leq \|\Delta_i\|_{L_i^4} \|\bar{\psi}_{iK}^{(1)} - \mu_{iK}\|_{L_i^4} + C\|c_i\|_{L_i^2} + C\|d_i\|_{L_i^4}^2 + C\|d_i\|_{L_i^6}^3 \\ &\leq C(T^{-1} + b_T N^{-a_0/2}). \end{aligned}$$

Since $\mathbb{E}_i \Delta_i = 0$, the exact coefficient is $\mathcal{B}_{KT} = (T/N) \sum_i \mathbb{E}_i U_{iK}$. The population normal equation sets the unit-averaged expected score to zero. Consequently, the score identity holds with $r_{\theta,F,K} = N^{-1} \sum_i (U_{iK} - \mathbb{E}_i U_{iK})$. Conditional independence across units yields

$$\mathbb{E}[\sqrt{NT} r_{\theta,F,K}^2 \mid \mathcal{C}] = \frac{T}{N} \sum_i \text{Var}_i(U_{iK}) \leq C(T^{-1} + T b_T^2 N^{-a_0}) = o(1).$$

The last equality uses $T b_T^2 N^{-a_0} = C_b^2 (\log N) N^{-a_0}$. Thus the exact drift is retained and only the centered cross-unit remainder is discarded. The target-drift bound follows separately from (2.6) and Lemma D.4. $\square$

E.2 Bias-Correction Consistency

Use the fixed lags $|j| \leq L$ and the sample-mean reporting rule of Appendix A.2. Put $\mathcal{L}_K = \mathcal{E}_T \cap \mathcal{G}_K$. Lemma D.5 and Corollary D.17 give

$$\mathbb{P}(\mathcal{L}_K^c) \leq \mathbb{P}(\mathcal{E}_T^c) + \mathbb{P}(\mathcal{G}_K^c) \longrightarrow 0. \quad (\text{E.3})$$

All fitted comparisons below are pathwise on this event.

Lemma E.2 (Slope and derivative errors at generated coordinates). *Under (A.9),*

$$|\hat{\theta}_K - \theta_K| = O_p(T^{-1}), \quad \|\hat{c}_K - c_K\| = O_p(r_K). \quad (\text{E.4})$$

The derivative errors satisfy

$$\begin{aligned} e_{A,K} &:= \|\hat{\psi}_{it,K}^{(1)} - \psi_{it,K}^{(1)}\|_{NT,2} = O_p(T^{-1/2} + \lambda_K^{-1} r_K), \\ \frac{1}{NT} \sum_{i,t} |\hat{\psi}_{it,K}^{(2)} - \psi_{it,K}^{(2)}| &= O_p(T^{-1/2} + \lambda_K^{-2} r_K) = o_p(1). \end{aligned} \quad (\text{E.5})$$

Proof. Lemma D.11 gives the coefficient rate. The exact identity (A.4) and Proposition E.1 give

$$\hat{\theta}_K - \theta_K = \hat{J}_K^{-1} \left\{ \frac{1}{NT} \sum_{i,t} x_{it,K} + \frac{\mathcal{B}_{KT}}{T} + r_{\theta,F,K} - e_\gamma^\top \hat{Q}_K e_\pi \right\}.$$

Here $\frac{1}{NT} \sum_{i,t} x_{it,K} = O_p(T^{-1})$, the score remainder is $o_p(T^{-1})$, and the profile product is $O_p(r_K^2) = o_p(T^{-1})$. The expectation calculation below gives $\mathcal{B}_{KT} = O(1)$. Lemmas D.13 and D.14 bound the denominator away from zero, proving the slope rate.

At $a = \hat{\alpha}_i$, for $r = 0, 1, 2$, write $M_{it,K}^{(r)} = \partial_a^r m_K(X_{it}, a)$, $P_{it,K}^{(r)} = \partial_a^r (\ell_K - \theta_K m_K)(X_{it}, a)$, and let $\Delta M^{(r)}$, $\Delta P^{(r)}$ be their fitted-minus-oracle counterparts, using $\hat{\theta}_K$ in the fitted P . The fitted-minus-oracle coefficient of P is $e_\delta - \hat{\theta}_K e_\gamma - (\hat{\theta}_K - \theta_K) \gamma_K$, whose norm is $O_p(r_K)$ by the coefficient and slope rates. This fitted- P difference uses $\hat{\theta}_K$; the profile vector e_π in the root identity uses θ_K . The ordinary and derivative Gram bounds in Lemmas D.10 and D.12 give

$$\|\Delta M^{(r)}\|_{NT,2} + \|\Delta P^{(r)}\|_{NT,2} = O_p(\lambda_K^{-r} r_K), \quad r = 0, 1, 2. \quad (\text{E.6})$$

Define $u_{D,it,K}^0 = D_{it} - M_{it,K}^{(0)}$, $q_{it,K}^0 = Y_{it} - \theta_K D_{it} - P_{it,K}^{(0)}$, and $\Delta q_{it,K} = -(\hat{\theta}_K - \theta_K) D_{it} - \Delta P_{it,K}^{(0)}$. Then $\hat{u}_D = u_D^0 - \Delta M^{(0)}$ and $\hat{q} = q^0 + \Delta q$ exactly. Local support and bounded oracle functions imply

$$\|\Delta q\|_{NT,2} = O_p(r_K), \quad \|\Delta q\|_\infty + \|\Delta M^{(0)}\|_\infty = O_p(\sqrt{K} r_K) = o_p(1). \quad (\text{E.7})$$

Indeed, $\sqrt{K} r_K \leq C \lambda_K^{-m}/T = o(1)$. Thus fitted level residuals are bounded in probability.

Apply the score-derivative formulas in Section 2.2 at the same generated coordinate before moving the oracle argument back to α_i . The bounded oracle paths make this latter error $O_p(T^{-1/2})$ in empirical L^2 for both derivative orders. For $r = 1, 2$, let $D^{(r)}$ denote the fitted-minus-oracle derivative at this same generated coordinate. Direct substitution gives

$$\begin{aligned} D^{(1)} &= -M^{(1)} \Delta q - \Delta M^{(1)}(q^0 + \Delta q) - u_D^0 \Delta P^{(1)} + \Delta M^{(0)}(P^{(1)} + \Delta P^{(1)}), \\ D^{(2)} &= -M^{(2)} \Delta q - \Delta M^{(2)}(q^0 + \Delta q) - u_D^0 \Delta P^{(2)} + \Delta M^{(0)}(P^{(2)} + \Delta P^{(2)}) \\ &\quad + 2M^{(1)} \Delta P^{(1)} + 2P^{(1)} \Delta M^{(1)} + 2\Delta M^{(1)} \Delta P^{(1)}. \end{aligned}$$

Subtraction of the first-derivative formulas and Cauchy–Schwarz give

$$\begin{aligned} e_{A,K} &\leq O_p(T^{-1/2}) + C_{NT} \{\|\Delta q\|_{NT,2} + \|\Delta M^{(0)}\|_{NT,2} + \|\Delta M^{(1)}\|_{NT,2} + \|\Delta P^{(1)}\|_{NT,2}\} \\ &= O_p(T^{-1/2} + \lambda_K^{-1} r_K). \end{aligned}$$

The random factor $C_{NT} = O_p(1)$ follows from the fitted residual bounds. For the second derivative the same subtraction gives linear errors through order two and one product of fitted first-derivative errors. This product satisfies

$$\begin{aligned} \frac{1}{NT} \sum_{i,t} |\Delta M_{it,K}^{(1)} \Delta P_{it,K}^{(1)}| &\leq \|\Delta M^{(1)}\|_{NT,2} \|\Delta P^{(1)}\|_{NT,2} \\ &= O_p(\lambda_K^{-2} r_K^2), \\ \frac{1}{NT} \sum_{i,t} |\hat{\psi}_{it,K}^{(2)} - \psi_{it,K}^{(2)}| &\leq O_p(T^{-1/2}) + C_{NT} \sum_{r=0}^2 \{\|\Delta M^{(r)}\|_{NT,2} + \|\Delta P^{(r)}\|_{NT,2}\} \\ &\quad + \frac{2}{NT} \sum_{i,t} |\Delta M_{it,K}^{(1)} \Delta P_{it,K}^{(1)}| \\ &= O_p(T^{-1/2} + \lambda_K^{-2} r_K^2). \end{aligned}$$

□

Proposition E.3 (Consistency of the bias correction). *Under (A.9),*

$$\hat{B}_K - B_K = O_p(T^{-1/2} + \lambda_K^{-2} r_K) = o_p(1). \quad (\text{E.8})$$

Moreover,

$$\hat{J}_K^{-1} \hat{B}_K - J_K^{-1} \mathcal{B}_{KT} = o_p(1). \quad (\text{E.9})$$

Proof. Write $\Omega_{iT}^0 = \sum_{|j| \leq L} T^{-1} \sum_{t \in I_T(j)} \zeta_{it} \zeta_{i,t+j}$ for the oracle innovation lag sum. Because $\hat{\xi}_{it} = \zeta_{it} - d_i$ and the retained lag set is fixed, $\sup_i |\hat{\Omega}_i| \leq C$ pathwise. The reuse-channel error has the exact decomposition

$$\begin{aligned} \hat{B}_K^{(1)} - B_K^{(1)} &= \frac{1}{NT} \sum_i \sum_{|j| \leq L} \sum_{t \in I_T(j)} (\hat{\psi}_{it,K}^{(1)} - \psi_{it,K}^{(1)}) \hat{\xi}_{i,t+j} \\ &\quad - \frac{1}{N} \sum_i d_i \sum_{|j| \leq L} \frac{1}{T} \sum_{t \in I_T(j)} \psi_{it,K}^{(1)} \\ &\quad + \frac{1}{NT} \sum_i \sum_{|j| \leq L} \sum_{t \in I_T(j)} \{\psi_{it,K}^{(1)} \zeta_{i,t+j} - \mathbb{E}_i[\psi_{it,K}^{(1)} \zeta_{i,t+j}]\} \\ &\quad - \frac{1}{N} \sum_i \sum_{|j| \leq L} \frac{|j|}{T} \mathbb{E}_i[\psi_{i0,K}^{(1)} \zeta_{ij}]. \end{aligned}$$

These terms have orders $O_p(e_{A,K})$, $O_p(T^{-1/2})$, $O_p((NT)^{-1/2})$, and $O(T^{-1})$, respectively: use bounded $\hat{\xi}$, the average root moment, the fixed-lag variance bound in Lemma C.1, and exact stationary pair counts. The second term includes the boundary contribution from a constant score derivative.

The fixed lag set and bounded innovations also give $|\hat{\Omega}_i - \Omega_{iT}^0| \leq C(|d_i| + d_i^2)$. Hence the curvature-channel error satisfies

$$\begin{aligned} |\hat{B}_K^{(2)} - B_K^{(2)}| &\leq \frac{C}{NT} \sum_{i,t} |\hat{\psi}_{it,K}^{(2)} - \psi_{it,K}^{(2)}| + \frac{C}{N} \sum_i (|d_i| + d_i^2) \\ &\quad + \frac{C}{N} \sum_i |\bar{\psi}_{iK}^{(2)} - \nu_{iK}| + \frac{C}{N} \sum_i |\Omega_{iT}^0 - \Omega_\zeta|. \end{aligned}$$

The last two averages are $O_p(T^{-1/2})$: the centered unit means and lag products have L_i^2 order $T^{-1/2}$, and $\mathbb{E}_i \Omega_{iT}^0 - \Omega_\zeta = O(T^{-1})$. Lemma E.2 proves (E.8). The calculation below gives $\mathcal{B}_{KT} - B_K = o(1)$ and $B_K = O(1)$. Finally, the exact normalization identity

$$\hat{J}_K^{-1} \hat{B}_K - J_K^{-1} \mathcal{B}_{KT} = \hat{J}_K^{-1} (\hat{B}_K - B_K) + \hat{J}_K^{-1} (B_K - \mathcal{B}_{KT}) + (\hat{J}_K^{-1} - J_K^{-1}) \mathcal{B}_{KT}$$

and denominator consistency prove (E.9). Remark D.18 transfers the result to the reported procedure. $\square$

Comparison of the exact drift and bias coefficient. Let R_{iK} be the exact third-order Taylor remainder in the proof of Proposition E.1. Condition (5.2), after a time shift, implies

$$\mathbb{E}_i[\psi_{it,K}^{(2)} \zeta_{is} \zeta_{ir}] = \nu_{iK} \mathbb{E}_i[\zeta_{is} \zeta_{ir}]. \quad (\text{E.10})$$

Lemma E.4, (E.10), and finite-range pair counts give

$$\begin{aligned} T \mathbb{E}_i [\Delta_i \bar{\psi}_{iK}^{(1)}] &= \mathbb{E}_X [2b_K m_{K,a} - h_{K,a}] \times \left\{ \Omega_{v\zeta} - \frac{1}{T} \sum_{|j| \leq L} |j| \gamma_{v\zeta}(j) \right\}, \\ \frac{T}{2} \mathbb{E}_i [\Delta_i^2 \bar{\psi}_{iK}^{(2)}] &= \frac{\nu_{iK}}{2} \left\{ \Omega_\zeta - \frac{1}{T} \sum_{|j| \leq L} |j| \gamma_\zeta(j) \right\}. \end{aligned}$$

Consequently,

$$\begin{aligned}
\mathcal{B}_{KT} - B_K^{(1)} - B_K^{(2)} &= -\frac{\bar{a}_{v,K}}{T} \sum_{|j| \leq L} |j| \gamma_{v\zeta}(j) \\
&\quad - \frac{1}{2TN} \left\{ \sum_{|j| \leq L} |j| \gamma_{\zeta}(j) \right\} \sum_i \nu_{iK} \\
&\quad + \frac{T}{N} \sum_i \mathbb{E}_i[(d_i - \Delta_i) \bar{\psi}_{iK}^{(1)}] \\
&\quad + \frac{T}{2N} \sum_i \mathbb{E}_i[(d_i^2 - \Delta_i^2) \bar{\psi}_{iK}^{(2)}] + \frac{T}{N} \sum_i \mathbb{E}_i[R_{iK}].
\end{aligned} \tag{E.11}$$

Since clipping is a contraction about the true coordinate and can bind only when $|\Delta_i| > b_T$,

$$|c_i| \leq |\Delta_i| \mathbf{1}\{|\Delta_i| > b_T\}, \quad |d_i^2 - \Delta_i^2| \leq \Delta_i^2 \mathbf{1}\{|\Delta_i| > b_T\}.$$

The raw tail moments in (D.14) and bounded score paths therefore give

$$\begin{aligned}
\frac{T}{N} \sum_i \left| \mathbb{E}_i[(d_i - \Delta_i) \bar{\psi}_{iK}^{(1)}] \right| &\leq CTb_T N^{-a_0}, \\
\frac{T}{2N} \sum_i \left| \mathbb{E}_i[(d_i^2 - \Delta_i^2) \bar{\psi}_{iK}^{(2)}] \right| &\leq CTb_T^2 N^{-a_0}, \\
\frac{T}{N} \sum_i \mathbb{E}_i |R_{iK}| &\leq CT^{-1/2}.
\end{aligned} \tag{E.12}$$

The two finite-window terms in (E.11) are $O(T^{-1})$. Hence $|\mathcal{B}_{KT} - B_K| \leq C\{T^{-1/2} + Tb_T N^{-a_0} + Tb_T^2 N^{-a_0}\} = o(1)$. The mixed third-moment restriction enters this calculation and the nuisance coefficient bound.

E.3 Joint Finite-Range Bias Calculations

The analysis derives the bias formulas under joint bounded finite-range dependence, allowing correlation between the measurement shock and treatment residual while maintaining an independent outcome-error process. These identities feed into the variance limit below and the theorem proofs in Appendix F.

Long-Run Covariances Stationarity and L -dependence give $\gamma_{v\zeta}(j) = \gamma_{\zeta}(j) = 0$ for $|j| > L$, so their long-run sums are exactly $\Omega_{v\zeta}$ and Ω_{ζ} defined above. Example A.3 gives closed-form formulas for these objects under finite moving-average filters, including the MA(1) design used in the simulations.

Derivative and Curvature Terms Suppress (X_{it}, α_i) on grid-projection functions. At the finite-sieve target, $u_{D,it,K} = v_{it} + r_{m,it,K}$, $q_{it,K} := u_{Y,it,K} - \theta_K u_{D,it,K} = \varepsilon_{it} + r_{h,it,K} - b_K u_{D,it,K}$.

Lemma E.4 (Derivative and Hessian decomposition). *Set $a_{it,K}^{(1)} = 2b_K m_{K,a} - h_{K,a}$ and $a_{it,K}^{(2)} = 2b_K m_{K,aa} - h_{K,aa}$, with the functions evaluated at (X_{it}, α_i) . The coordinate derivative of the theta score satisfies*

$$\begin{aligned}
\psi_{it,K}^{(1)} &= b_{it,K}^{(1)} + a_{it,K}^{(1)} v_{it} - m_{K,a} \varepsilon_{it}, \\
b_{it,K}^{(1)} &= -m_{K,a} r_{h,it,K} - r_{m,it,K} h_{K,a} + 2b_K r_{m,it,K} m_{K,a}.
\end{aligned} \tag{E.13}$$

The second derivative satisfies

$$\begin{aligned}\psi_{it,K}^{(2)} &= b_{it,K}^{(2)} + a_{it,K}^{(2)} v_{it} - m_{K,aa} \varepsilon_{it}, \\ b_{it,K}^{(2)} &= -m_{K,aa} r_{h,it,K} - r_{m,it,K} h_{K,aa} + 2m_{K,a} h_{K,a} - 2b_K m_{K,a}^2 + 2b_K r_{m,it,K} m_{K,aa}.\end{aligned}\tag{E.14}$$

Consequently,

$$\mathbb{E}_i[\psi_{i0,K}^{(1)} \zeta_{ij}] = \mathbb{E}_X[a_{i0,K}^{(1)}] \gamma_{v\zeta}(j).\tag{E.15}$$

$$\mathbb{E}_i[\psi_{it,K}^{(2)} \zeta_{is} \zeta_{ir}] = \nu_{iK} \mathbb{E}_i[\zeta_{is} \zeta_{ir}].\tag{E.16}$$

Proof. The proof first differentiates the score while holding (Y_{it}, D_{it}) fixed, and then substitutes the target residuals. Since $\partial_a u_{D,it,K} = -m_{K,a}$ and $\partial_a q_{it,K} = b_K m_{K,a} - h_{K,a}$, the product rule gives

$$\begin{aligned}\psi_{it,K}^{(1)} &= -m_{K,a} q_{it,K} + u_{D,it,K} (b_K m_{K,a} - h_{K,a}) \\ &= -m_{K,a} \varepsilon_{it} - m_{K,a} r_{h,it,K} + 2b_K u_{D,it,K} m_{K,a} - u_{D,it,K} h_{K,a} \\ &= b_{it,K}^{(1)} + a_{it,K}^{(1)} v_{it} - m_{K,a} \varepsilon_{it},\end{aligned}$$

where the last equality uses $u_{D,it,K} = v_{it} + r_{m,it,K}$ and groups the terms without v_{it} into $b_{it,K}^{(1)}$. This proves the decomposition in (E.13).

Differentiate the first line once more before substituting:

$$\begin{aligned}\psi_{it,K}^{(2)} &= -m_{K,aa} q_{it,K} - 2m_{K,a} (b_K m_{K,a} - h_{K,a}) + u_{D,it,K} (b_K m_{K,aa} - h_{K,aa}) \\ &= -m_{K,aa} \varepsilon_{it} - m_{K,aa} r_{h,it,K} + 2b_K u_{D,it,K} m_{K,aa} \\ &\quad - u_{D,it,K} h_{K,aa} - 2b_K m_{K,a}^2 + 2m_{K,a} h_{K,a} \\ &= b_{it,K}^{(2)} + a_{it,K}^{(2)} v_{it} - m_{K,aa} \varepsilon_{it}.\end{aligned}$$

This proves (E.14) and displays the two product-rule cross terms explicitly.

For (E.15), multiply the first decomposition by ζ_{ij} . Terms in $b_{it,K}^{(1)}$ depend only on X and have a centered ζ factor; the outcome-error term has zero expectation by independence and centering. Thus only the $v\zeta$ term remains, giving (E.15).

Condition (5.2), after shifting all three dates by $-t$, gives $\mathbb{E}_i[v_{it} \zeta_{is} \zeta_{ir}] = 0$ for every (t, s, r) . The independent outcome-error process is centered as well. Only $b_{it,K}^{(2)}$ remains after multiplication by two ζ terms. Independence of X and the disturbance process then factors its expectation, proving (E.16). $\square$

Summing (E.15) over lags and using (E.16) give

$$B_K^{(1)} = \Omega_{v\zeta} \frac{1}{N} \sum_i \mathbb{E}_X[a_{i0,K}^{(1)}] = \Omega_{v\zeta} \bar{a}_{v,K}.\tag{E.17}$$

$$B_K^{(2)} = \frac{\Omega_\zeta}{2N} \sum_i \nu_{iK}.\tag{E.18}$$

The exact finite-window pair counts and clipped-coordinate bridge are already derived in (E.11); their remainder bounds are (E.12). Thus $\mathcal{B}_{KT} - B_K^{(1)} - B_K^{(2)} = o(1)$. The raw sample mean has no nonlinear first-step-bias channel. Its clipped coordinate remains distinct on the exceptional event, whose contribution is controlled by (E.12).

E.4 Variance Estimation and the Cluster Central Limit Theorem

Lemma E.5 (Fixed-lag variance estimation and a cluster central limit theorem). *Consider a sequence with $N \rightarrow \infty$ and $T \rightarrow \infty$, with L fixed. Let χ_{it} be conditionally centered, stationary, L -dependent within units, independent across units, and uniformly bounded in conditional L^4 . For $T > L$, put $\Phi_{iT} = T^{-1/2} \sum_t \chi_{it}$ and*

$$S_{NT}^2 = \frac{1}{N} \sum_i \sum_{|j| \leq L} (1 - |j|/T) \mathbb{E}_i[\chi_{i0} \chi_{ij}]. \quad (\text{E.19})$$

If $S_{NT}^2 \rightarrow s^2 > 0$, then $(NT)^{-1/2} \sum_{i,t} \chi_{it} \rightsquigarrow \mathcal{N}(0, s^2)$. Independently of the Gaussian-limit condition, suppose $\chi_{it}^0 = \chi_{it} + a_i$, with a_i constant over dates, and $\|\hat{\chi} - \chi^0\|_{NT,2} = o_p(1)$. Then the difference between the fixed-lag variance computed from the demeaned fitted array and S_{NT}^2 is $o_p(1)$. No dependence restriction is imposed on the fitted or demeaned arrays.

Proof. Write $\bar{\chi}_i = T^{-1} \sum_t \chi_{it}$ and $\bar{\hat{\chi}}_i = T^{-1} \sum_t \hat{\chi}_{it}$. Demeaning removes a_i exactly and contracts the empirical L^2 norm. Thus, with $r_{it} = (\hat{\chi}_{it} - \bar{\hat{\chi}}_i) - (\chi_{it} - \bar{\chi}_i)$, $e = \|r\|_{NT,2} \leq \|\hat{\chi} - \chi^0\|_{NT,2} = o_p(1)$. For each retained lag, multiplication gives

$$\begin{aligned} & (\hat{\chi}_{it} - \bar{\hat{\chi}}_i)(\hat{\chi}_{i,t+j} - \bar{\hat{\chi}}_i) - (\chi_{it} - \bar{\chi}_i)(\chi_{i,t+j} - \bar{\chi}_i) \\ &= r_{it}(\chi_{i,t+j} - \bar{\chi}_i) + (\chi_{it} - \bar{\chi}_i)r_{i,t+j} + r_{it}r_{i,t+j}. \end{aligned}$$

Averaging over $I_T(j)$ and applying Cauchy–Schwarz bounds the absolute difference between the two lag statistics by

$$2e\|\chi_{it} - \bar{\chi}_i\|_{NT,2} + e^2 = o_p(1).$$

For the oracle array, put $a_{ij} = T^{-1} \sum_{t \in I_T(j)} \chi_{it}$ and $b_{ij} = T^{-1} \sum_{t \in I_T(j)} \chi_{i,t+j}$. The exact difference caused by demeaning is

$$\frac{1}{N} \sum_i \{ -\bar{\chi}_i(a_{ij} + b_{ij}) + (1 - |j|/T)(\bar{\chi}_i)^2 \}.$$

The shared mean bound gives conditional second moments C/T for all three unit means. Cauchy–Schwarz and Markov’s inequality make this difference $O_p(T^{-1})$. The undemeaned oracle lag average has its stationary expectation $(1 - |j|/T)N^{-1} \sum_i \mathbb{E}_i[\chi_{i0} \chi_{ij}]$ and fluctuation $O_p((NT)^{-1/2})$ by (C.2). Summing the fixed set of lags proves variance consistency.

For the Gaussian limit, exact window counts give $N^{-1} \sum_i \text{Var}_i \Phi_{iT} = S_{NT}^2$. Lemma C.1 gives $\sup_i \mathbb{E}_i |\Phi_{iT}|^4 \leq C$. Thus the conditional Lyapunov ratio is at most

$$\frac{N^{-2} \sum_i \mathbb{E}_i |\Phi_{iT}|^4}{(S_{NT}^2)^2} \leq \frac{C}{N(S_{NT}^2)^2} \rightarrow 0.$$

The independent-cluster triangular-array theorem proves the limit. $\square$

Feasible-to-Oracle Comparison The comparison uses the fixed lags $|j| \leq L$.

Write $\bar{\phi}_{iK} = T^{-1} \sum_t \phi_{it,K}$ and $\bar{\hat{\phi}}_{iK} = T^{-1} \sum_t \hat{\phi}_{it,K}$ for the unit sample means. These sample means differ from conditional population means, which center the quadratic first-step terms and the coefficient derivative averages elsewhere in the proofs. The centered oracle score is $\phi_{it,K} = J_K^{-1} x_{it,K}$. For the feasible comparison, define the uncentered oracle version

$$\phi_{it,K}^0 = J_K^{-1} \{ \psi_{it,K}(\alpha_i) + \mu_{iK} \zeta_{it} \} = \phi_{it,K} + J_K^{-1} \mathbb{E}_i[\psi_{i0,K}]. \quad (\text{E.20})$$

The last term is constant within unit. Therefore, demeaning $\phi_{it,K}^0$ gives $\phi_{it,K} - \bar{\phi}_{iK}$ exactly. The fixed-lag oracle HAC statistic is

$$\hat{q}_{K,0} = \frac{1}{NT} \sum_{i=1}^N \sum_{|j| \leq L} \sum_{t \in I_T(j)} (\phi_{it,K} - \bar{\phi}_{iK})(\phi_{i,t+j,K} - \bar{\phi}_{iK}).$$

Define its finite- T target by

$$\sigma_{KNT}^2 = \frac{1}{N} \sum_{i=1}^N \sum_{|j| \leq L} \left(1 - \frac{|j|}{T}\right) \text{Cov}_i(\phi_{i0,K}, \phi_{ij,K}). \quad (\text{E.21})$$

Lemma E.6 (Feasible-to-oracle score comparison after demeaning). *Under Assumption 5.1 and the rate conditions in (A.9),*

$$e_{\phi,K} := \left\| \hat{\phi}_{it,K} - \phi_{it,K}^0 \right\|_{NT,2} = O_p\left(T^{-1/2} + \lambda_K^{-1} r_K + \left| \hat{J}_K - J_K \right| \right) = o_p(1). \quad (\text{E.22})$$

Since the unit constant in (E.20) disappears exactly and within-unit demeaning is an L^2 contraction,

$$\left\| (\hat{\phi}_{it,K} - \bar{\phi}_{iK}) - (\phi_{it,K} - \bar{\phi}_{iK}) \right\|_{NT,2} \leq e_{\phi,K}. \quad (\text{E.23})$$

Proof. Work on $\mathcal{L}_K$. The residual identities used in Lemma E.2 give

$$\hat{\psi}_{it,K} - \psi_{it,K} = \psi_{it,K}(\hat{\alpha}_i) - \psi_{it,K} + u_{D,it,K}^0 \Delta q_{it,K} - q_{it,K}^0 \Delta M_{it,K}^{(0)} - \Delta M_{it,K}^{(0)} \Delta q_{it,K}.$$

The bounded coordinate derivative and fitted-level error bounds imply $\|\hat{\psi} - \psi\|_{NT,2} = O_p(T^{-1/2} + r_K)$. The first-derivative error bound and the oracle mean bound give

$$\left\{ \frac{1}{N} \sum_i (\hat{\mu}_{iK} - \mu_{iK})^2 \right\}^{1/2} \leq e_{A,K} + \left\{ \frac{1}{N} \sum_i (\bar{\psi}_{iK}^{(1)} - \mu_{iK})^2 \right\}^{1/2} = O_p(T^{-1/2} + \lambda_K^{-1} r_K),$$

$$\|\hat{\mu}\hat{\xi} - \mu\zeta\|_{NT,2} \leq C \left\{ \frac{1}{N} \sum_i (\hat{\mu}_{iK} - \mu_{iK})^2 \right\}^{1/2} + C \left\{ \frac{1}{N} \sum_i d_i^2 \right\}^{1/2}.$$

The second inequality uses bounded $\hat{\xi}$ and $\hat{\xi}_{it} - \zeta_{it} = -d_i$; it assumes no independence of fits. Since $J_K \phi^0 = \psi + \mu\zeta$, exact denominator replacement gives

$$\hat{\phi} - \phi^0 = (\hat{J}_K^{-1} - J_K^{-1})(\psi + \mu\zeta) + \hat{J}_K^{-1}\{(\hat{\psi} - \psi) + (\hat{\mu}\hat{\xi} - \mu\zeta)\}.$$

The oracle numerator has bounded empirical second moment, and the two denominators are bounded away from zero. The preceding inequalities prove (E.22). Subtracting unit means is an empirical L^2 contraction; the unit constant in (E.20) vanishes exactly, proving (E.23). Use Remark D.18 for the global probability statement. $\square$

Proposition E.7 (Consistency of the fixed-lag variance estimator). *Under Assumption 5.1 and the rate conditions in (A.9),*

$$\hat{q}_K - \hat{q}_{K,0} = o_p(1). \quad (\text{E.24})$$

$$\hat{q}_{K,0} - \sigma_{KNT}^2 = o_p(1). \quad (\text{E.25})$$

Consequently, $\hat{q}_K - \sigma_{KNT}^2 = o_p(1)$.

Proof. Apply Lemma E.5 with $\chi = \phi$ and $\chi^0 = \phi^0$ from (E.20). The oracle score is centered, stationary and finite-range with bounded fourth moments. Lemma E.6 verifies the fitted L^2 comparison. The lemma's lagwise product bound gives (E.24); its oracle demeaning and window calculation gives (E.25) with target (E.21). Both statements hold before taking a variance limit. $\square$

Variance Target and Limit The preceding consistency result concerns the raw lag statistic $\hat{q}_K$. The result below proves $\sigma_{KNT}^2 \rightarrow \sigma^2 > 0$, so the sample-mean reporting rule changes it with probability tending to zero and $\hat{\sigma}_K^2 - \sigma_{KNT}^2 = o_p(1)$.

Lemma E.8 (Oracle score and variance limit). *Under Assumption 5.1 and $\lambda_K \rightarrow 0$,*

$$\frac{1}{N} \sum_{i=1}^N \sum_{|j| \leq L} |\text{Cov}_i(x_{i0,K}, x_{ij,K}) - \text{Cov}_i(v_{i0}\varepsilon_{i0}, v_{ij}\varepsilon_{ij})| = o(1), \quad (\text{E.26})$$

and $J_K \rightarrow \gamma_v(0)$. Hence

$$\begin{aligned} \sigma_{KNT}^2 &\longrightarrow \sigma^2, \\ \sigma^2 &:= \frac{\sum_{|j| \leq L} \gamma_v(j) \gamma_\varepsilon(j)}{\gamma_v(0)^2}. \end{aligned} \quad (\text{E.27})$$

Proof. At the grid target, $u_D = v + r_m$ and $u_Y - \theta_K u_D = \varepsilon + r_h - b_K u_D$. Consequently, the exact score difference is

$$\begin{aligned} \psi_{it,K}(\alpha_i) - v_{it}\varepsilon_{it} &= r_{m,it,K}\varepsilon_{it} + v_{it}r_{h,it,K} + r_{m,it,K}r_{h,it,K} \\ &\quad - b_K(v_{it} + r_{m,it,K})^2. \end{aligned}$$

Bounded disturbances, the two projection rates, and identity (2.6) make each term on the right negligible in finite-grid L^2 :

$$\frac{1}{N} \sum_{i=1}^N \mathbb{E}_i \left[|\psi_{i0,K}(\alpha_i) - v_{i0}\varepsilon_{i0}|^2 \right] = o(1).$$

The outcome error is centered and independent of v , so $\mathbb{E}_i[v_{i0}\varepsilon_{i0}] = 0$. Jensen's inequality therefore supplies the unit-centering bound

$$\frac{1}{N} \sum_{i=1}^N |\mathbb{E}_i[\psi_{i0,K}]|^2 \leq \frac{1}{N} \sum_{i=1}^N \mathbb{E}_i \left[|\psi_{i0,K}(\alpha_i) - v_{i0}\varepsilon_{i0}|^2 \right] = o(1).$$

The derivative decomposition in Section E.3 also gives $N^{-1} \sum_i |\mu_{iK}|^2 = o(1)$. Since ζ is bounded, the full centered-score difference is

$$x_{it,K} - v_{it}\varepsilon_{it} = \{ \psi_{it,K}(\alpha_i) - v_{it}\varepsilon_{it} \} - \mathbb{E}_i[\psi_{i0,K}] + \mu_{iK}\zeta_{it}.$$

Consequently,

$$\begin{aligned} &\frac{1}{N} \sum_{i=1}^N \mathbb{E}_i \left[|x_{i0,K} - v_{i0}\varepsilon_{i0}|^2 \right] \\ &\leq \frac{3}{N} \sum_{i=1}^N \mathbb{E}_i \left[|\psi_{i0,K}(\alpha_i) - v_{i0}\varepsilon_{i0}|^2 \right] + \frac{3}{N} \sum_{i=1}^N |\mathbb{E}_i[\psi_{i0,K}]|^2 + \frac{C}{N} \sum_{i=1}^N |\mu_{iK}|^2 = o(1). \end{aligned}$$

Because both scores are centered, their covariance difference has the exact two-term decomposition

$$\begin{aligned} &\text{Cov}_i(x_{i0,K}, x_{ij,K}) - \text{Cov}_i(v_{i0}\varepsilon_{i0}, v_{ij}\varepsilon_{ij}) \\ &= \text{Cov}_i(x_{i0,K} - v_{i0}\varepsilon_{i0}, x_{ij,K}) + \text{Cov}_i(v_{i0}\varepsilon_{i0}, x_{ij,K} - v_{ij}\varepsilon_{ij}). \end{aligned}$$

Their second moments are uniformly bounded. Averaging the last display over i and applying Cauchy–Schwarz therefore gives

$$\begin{aligned} & \frac{1}{N} \sum_{i=1}^N |\text{Cov}_i(x_{i0,K}, x_{ij,K}) - \text{Cov}_i(v_{i0}\varepsilon_{i0}, v_{ij}\varepsilon_{ij})| \\ & \leq C \left\{ \frac{1}{N} \sum_{i=1}^N \mathbb{E}_i |x_{i0,K} - v_{i0}\varepsilon_{i0}|^2 \right\}^{1/2} = o(1). \end{aligned}$$

Summing this bound over the fixed lag set proves (E.26). Similarly, the denominator obeys

$$\begin{aligned} J_K &= \gamma_v(0) + \frac{2}{N} \sum_{i=1}^N \mathbb{E}_i [v_{i0} r_{m,i0,K}] + \frac{1}{N} \sum_{i=1}^N \mathbb{E}_i [r_{m,i0,K}^2] \\ &= \gamma_v(0) + o(1). \end{aligned}$$

Assumption 5.1(i)–(ii) makes the cross term zero, and the projection-error bound makes the last term $o(1)$. The ε process is independent of the joint (v, ζ) process, so $\text{Cov}(v_{i0}\varepsilon_{i0}, v_{ij}\varepsilon_{ij}) = \gamma_v(j)\gamma_\varepsilon(j)$. Substitution in (E.21), followed by $T \rightarrow \infty$, gives

$$\begin{aligned} \sigma_{KNT}^2 &= \frac{1}{J_K^2} \frac{1}{N} \sum_{i=1}^N \sum_{|j| \leq L} \left(1 - \frac{|j|}{T}\right) \text{Cov}_i(x_{i0,K}, x_{ij,K}) \\ &\rightarrow \frac{1}{\gamma_v(0)^2} \sum_{|j| \leq L} \gamma_v(j)\gamma_\varepsilon(j), \end{aligned}$$

which proves (E.27). For nondegeneracy, define the spectral-density trigonometric polynomials

$$f_c(\omega) = \frac{1}{2\pi} \sum_{|j| \leq L} \gamma_c(j) e^{-ij\omega}, \quad c \in \{v, \varepsilon\}.$$

They are continuous and nonnegative. Since $\gamma_c(0) > 0$, neither is identically zero. A nonzero trigonometric polynomial has a zero set of Lebesgue measure zero. Since scalar covariance functions satisfy $\gamma_c(-j) = \gamma_c(j)$, Parseval's identity gives

$$\sum_{|j| \leq L} \gamma_v(j)\gamma_\varepsilon(j) = 2\pi \int_{-\pi}^{\pi} f_v(\omega) f_\varepsilon(\omega) d\omega > 0.$$

Hence $\sigma^2 > 0$. □

Cluster Central Limit Theorem

Proposition E.9 (Scalar cluster central limit theorem). *Under Assumption 5.1 and $\lambda_K \rightarrow 0$,*

$$N^{-1/2} \sum_{i=1}^N \Phi_{iT,K} \rightsquigarrow \mathcal{N}(0, \sigma^2). \quad (\text{E.28})$$

Proof. Use Lemma E.5 with the centered oracle ϕ . The bounded score and derivative envelopes, and J_K bounded away from zero, verify its uniform fourth-moment condition. Lemma E.8 supplies $S_{NT}^2 = \sigma_{KNT}^2 \rightarrow \sigma^2 > 0$. □

E.5 Proof of the Active-Channel Result

Proof. Equation (E.17), obtained from Lemma E.4, gives $B_K^{(1)} = \Omega_{v\zeta} \bar{a}_{v,K}$. It remains to establish the limit of $\bar{a}_{v,K}$.

Let $q_K(a) = \int_{[0,1]^{d_X}} h_{K,a}(x, a) dx$. Lemma D.4 gives a uniform bound on q'_K , so q_K has uniformly bounded total variation. Applying integration by parts to the discrepancy bound in (5.1), and accounting for the two omitted boundary strips, yields an $O(N^{-1} + b_T)$ approximation error. Uniform derivative approximation and quasi-uniform coordinate quadrature give

$$\left| \frac{1}{N} \sum_i \mathbb{E}_X h_{K,a}(X, \alpha_i) - \int_{[0,1]^{d_X}} \int_0^1 h_{0,a}(x, a) da dx \right| \leq \|h_{K,a} - h_{0,a}\|_\infty + C(b_T + N^{-1}) = o(1).$$

Also, $|\frac{1}{N} \sum_i \mathbb{E}_X m_{K,a}(X, \alpha_i)| = O(1)$. Identity (2.6) gives $b_K \rightarrow 0$. Therefore

$$\begin{aligned} \bar{a}_{v,K} &= 2b_K \frac{1}{N} \sum_i \mathbb{E}_X m_{K,a}(X, \alpha_i) - \frac{1}{N} \sum_i \mathbb{E}_X h_{K,a}(X, \alpha_i) \\ &\rightarrow - \int_{[0,1]^{d_X}} \{h_0(x, 1) - h_0(x, 0)\} dx = -\mathbf{b}_h. \end{aligned} \tag{E.29}$$

The last equality is the fundamental theorem of calculus in a . Combining (E.17) and (E.29) completes the proof. $\square$

E.6 Proof of the Primitive Bias Limit

Proof. Section E.5 establishes the derivative-influence limit for the covariance channel in Proposition 5.2. For the curvature channel, the exact Hessian decomposition in (E.14) gives

$$\begin{aligned} \frac{1}{N} \sum_i \nu_{iK} &= \frac{1}{N} \sum_i \mathbb{E}_X [-m_{K,aa} r_{h,it,K} - r_{m,it,K} h_{K,aa} + 2m_{K,a} h_{K,a} \\ &\quad - 2b_K m_{K,a}^2 + 2b_K r_{m,it,K} m_{K,aa}]. \end{aligned}$$

The projection bounds in Section D.1 imply uniform convergence of the first coordinate derivatives and uniform boundedness of the displayed second derivatives. In particular, the first, second, and fifth terms obey the explicit product bounds

$$\begin{aligned} \left| \frac{1}{N} \sum_i \mathbb{E}_X [m_{K,aa} r_{h,it,K}] \right| &\leq \|m_{K,aa}\|_\infty \|r_{h,\cdot,K}\|_\infty = O(\lambda_K^{\tau_h}), \\ \left| \frac{1}{N} \sum_i \mathbb{E}_X [r_{m,it,K} h_{K,aa}] \right| &\leq \|r_{m,\cdot,K}\|_\infty \|h_{K,aa}\|_\infty = O(\lambda_K^{\tau_m}), \\ \left| \frac{2b_K}{N} \sum_i \mathbb{E}_X [r_{m,it,K} m_{K,aa}] \right| &\leq 2|b_K| \|r_{m,\cdot,K}\|_\infty \|m_{K,aa}\|_\infty = o(1). \end{aligned}$$

Identity (2.6) gives $b_K \rightarrow 0$, so the fourth term is bounded by $2|b_K| \|m_{K,a}\|_\infty^2 = o(1)$.

It remains to transfer the signed product $m_{K,a} h_{K,a}$ from the coordinate grid to the continuous design. Define the two nonnegative functions $q_{\pm,K}(a) = \int_{[0,1]^{d_X}} \{m_{K,a}(x, a) \pm h_{K,a}(x, a)\}^2 dx$. The first coordinate derivatives of the two projections are continuous and cellwise polynomials of degree at most $p-1$. Hence $q_{s,K} \in \mathcal{Q}_{2(p-1),K}$ for each $s \in \{+, -\}$, so Lemma D.3 applies to both signs.

The polarization identity $4m_{K,a}h_{K,a} = (m_{K,a} + h_{K,a})^2 - (m_{K,a} - h_{K,a})^2$ therefore gives

$$\begin{aligned} & \left| \frac{1}{N} \sum_i \mathbb{E}_X[m_{K,a}(X, \alpha_i)h_{K,a}(X, \alpha_i)] - \int_{W^+} m_{K,a}(w)h_{K,a}(w) dw \right| \\ & \leq \frac{1}{4} \sum_{s \in \{+, -\}} \left| \frac{1}{N} \sum_i q_{s,K}(\alpha_i) - \int_0^1 q_{s,K}(a) da \right| \\ & \leq C \left\{ \frac{1}{N\lambda_K} + \frac{b_T}{\lambda_K} \right\} \int_0^1 \{q_{+,K}(a) + q_{-,K}(a)\} da = o(1). \end{aligned}$$

The last line uses (5.3) and (D.9), which imply $\int_0^1 (q_{+,K} + q_{-,K}) da = O(1)$. The same derivative approximation also gives

$$\begin{aligned} & \left| \int_{W^+} m_{K,a}(w)h_{K,a}(w) dw - \int_{W^+} m_{0,a}(w)h_{0,a}(w) dw \right| \\ & \leq \|m_{K,a} - m_{0,a}\|_\infty \|h_{K,a}\|_\infty + \|m_{0,a}\|_\infty \|h_{K,a} - h_{0,a}\|_\infty = o(1). \end{aligned}$$

Combining these two limits with the four bounds above yields

$$\frac{1}{N} \sum_i \nu_{iK} \rightarrow 2 \int_{W^+} m_{0,a}(w)h_{0,a}(w) dw = 2\mathfrak{c}_{mh}.$$

Combining this limit with (E.18) proves (5.5). The two channel limits and the exact decomposition of B_K then give (5.6). Finally, Equation (E.11) shows that the clipped finite- T coefficient differs from the two analytical channels by $o(1)$, proving (5.7). $\square$

F Proof of the Main Growing-Sieve Theorem

F.1 Rates and the Two Expansions

Proof of Theorem 5.3. Write $m = \max\{d_W, 3\}$ and $\lambda_K \asymp T^{-q/d_W}$. The lower rate-window bound gives $T\lambda_K^{\tau_m + \tau_h} \rightarrow 0$, hence $\sqrt{NT}(\theta_K - \theta_0) = o(1)$ by (2.6). The upper bounds give $q < 1$ and $q(m+4)/(2d_W) < 1$, so $\lambda_K^{-2}r_K \rightarrow 0$ and $\lambda_K^{-m}/T \rightarrow 0$. For $d_W \geq 3$ the latter is $q < 1$; for $d_W < 3$ it follows from $q < 2d_W/(m+4)$. Also $K_x \log(NK_x)/T = o(1)$ because $K_x \asymp T^{qd_X/d_W}$ and $d_X < d_W$. Thus (A.9) holds and $\lambda_K \rightarrow 0$.

Lemma D.11 gives both coefficient errors order r_K . The stable Gram therefore gives

$$\begin{aligned} \sqrt{NT}|e_\gamma^\top \hat{Q}_K e_\pi| & \leq \sqrt{NT} \|\hat{Q}_K\|_{\text{op}} \|e_\gamma\| \|e_\pi\| = O_p(\sqrt{NT} r_K^2) \\ & = O_p(\lambda_K^{-m}/T) = o_p(1). \end{aligned}$$

Proposition E.1 controls the score remainder; the rate calculation already controls the target difference. Corollary D.17 proves high-probability acceptance. Lemmas D.13 and D.14 give $\hat{J}_K - J_K = o_p(1)$ and $J_K \rightarrow \gamma_v(0) > 0$.

On acceptance, combine the exact profile identity with the exact score expansion and subtract the feasible correction:

$$\begin{aligned} \sqrt{NT}(\hat{\theta}_K - \theta_0) & = \hat{J}_K^{-1} \frac{1}{\sqrt{NT}} \sum_{i,t} x_{it,K} + \sqrt{N/T} \hat{J}_K^{-1} \mathcal{B}_{KT} + o_p(1), \\ \sqrt{NT}(\tilde{\theta}_K - \theta_0) & = \hat{J}_K^{-1} \frac{1}{\sqrt{NT}} \sum_{i,t} x_{it,K} + \sqrt{N/T} \hat{J}_K^{-1} (\mathcal{B}_{KT} - \hat{B}_K) + o_p(1). \end{aligned}$$

The centered leading sum is $O_p(1)$, and the exact bridge gives $\mathcal{B}_{KT} = B_K + o(1) = O(1)$. Therefore denominator consistency permits $\hat{J}_K^{-1}$ to be replaced by J_K^{-1} in both leading terms. Proposition E.3 and the expectation bridge make $\mathcal{B}_{KT} - \hat{B}_K = o_p(1)$, removing the second term of the corrected expansion. For the uncorrected expansion, Proposition 5.2 gives $B_K \rightarrow B$. Using $x = J_K \phi$, $N/T \rightarrow \kappa_0$ and $J_K \rightarrow \gamma_v(0)$ proves both linear representations. Lemma E.8 gives the positive variance limit, and Proposition E.9 supplies the Gaussian limit. Lemma E.6 and Proposition E.7 give raw variance consistency. The sample-mean safeguard selects its positive branch with probability approaching one, so square-root continuity and Slutsky's theorem give studentization. Remark D.18 makes the result global. $\square$

F.2 Serial-Independence Specialization

Proof. The inherited conditions ensure that Theorem 5.3 applies. Serial independence makes all nonzero-lag population covariances vanish, while mutual independence gives $\Omega_{v\zeta} = 0$. Hence the active derivative channel vanishes. Equation (5.8) reduces to $\text{Var}(\varepsilon_{it})/\text{Var}(v_{it})$, which proves the stated specialization. $\square$

G Proof of the Smooth Scalar Moment Theorem

Throughout this appendix, Assumption 6.1 holds. We first expand the observable coordinate and the target score. We then control the random-coordinate Gram matrix and derive the nuisance rate from its normal equations. The remaining steps compare the feasible lag statistics to their oracle versions and apply a central limit theorem across units. Throughout this appendix, fitted second-stage objects refer to the procedure in Section 6, with acceptance event $\mathcal{O}_T$ and threshold $a_T = c_{\text{tol}} T^{-1/4}$.

The growth conditions imply

$$\begin{aligned} \sqrt{T} \lambda_K &\rightarrow \infty, & b_T / \lambda_K &\rightarrow 0, & N \lambda_K &\rightarrow \infty, \\ K \log(2K) / (NT) &\rightarrow 0, & K_x \log(2NK_x) / T &\rightarrow 0. \end{aligned}$$

For example, $\lambda_K^{-1} = o(T^{1/(dw+4)})$ bounds each power of λ_K^{-1} in these expressions by a strictly smaller power of T than its denominator; the remaining logarithms are absorbed. The deterministic chain in Lemmas D.1, D.2, D.3, and D.4 therefore applies to the present grid and spline. In particular,

$$\begin{aligned} cI_K &\preceq Q_K \preceq CI_K, & Q_K &= \frac{1}{N} \sum_i \mathbb{E}_X[p_K(X, \alpha_i) p_K(X, \alpha_i)^\top], \\ \sup_{x,a} \|\partial_a^r p_K(x, a)\| &\leq C \sqrt{K} \lambda_K^{-r}, & \|\partial_a^r(l_0 - \Pi_{K,N} l_0)\|_\infty &\leq C \lambda_K^{\tau_l - r}, \quad 0 \leq r \leq 3, \end{aligned} \tag{G.1}$$

where l_0 is m_0 or h_0 and the suprema are on W^+ . Consequently the first three coordinate derivatives of the projected functions, and of the scalar score at the population coefficients, have constant envelopes. The basis-vector envelopes have the displayed $\sqrt{K} \lambda_K^{-r}$ factors.

Independence of X from (v, ε) and $\mathbb{E}[v] = \mathbb{E}[\varepsilon] = \mathbb{E}[v\varepsilon] = 0$ verify the moment conditions of identity (2.6). Consequently, $J_K = V + \frac{1}{NT} \sum_{i,t} \mathbb{E}_i[r_{m,it,K}^2] \geq V > 0$ and $|b_K| \leq C \lambda_K^{\tau_m + \tau_h}$.

G.1 Observable Root and Score Expansion

Lemma G.1 (Observable-root expansion). *Let $\Delta_i = \check{\alpha}_i - \alpha_i$ and $d_i = \hat{\alpha}_i - \alpha_i$. There are uniform constants $C, c > 0$ such that*

$$\begin{aligned} \mathbb{P}_i(|\Delta_i| > x) &\leq C \exp(-cTx^2) + C \exp(-cT), \\ \|d_i\|_{L_i^r} + \|\bar{\xi}_i\|_{L_i^r} &\leq C_r T^{-1/2}, \quad 2 \leq r < \infty. \end{aligned}$$

The expansion (6.3) holds with the stated uniform moments. Also $\max_i |d_i| = O_p(b_T)$. For each fixed $r \geq 2$, $N^{-1} \sum_i |d_i|^r = O_p(T^{-r/2})$.

Proof. Under Assumption 6.1, the uniform-Jacobian grid argument of Lemma C.4 applies on the known compact $\mathcal{A}$. Let

$$\mathcal{J}_i = \left\{ \sup_{a \in \mathcal{A}} |\partial_a g_{iT}(a) - \mathbb{E}_i \partial_a g(W_{i0}, a)| \leq c_G/2 \right\}. \quad (\text{G.2})$$

Its failure probability is $C \exp(-cT)$. Intersect it with $|g_{iT}(\alpha_i)| < c_G \delta/2$, whose complement has the same exponential bound by Lemma C.1. The signed endpoint argument in Lemma C.4 places a unique zero in $I_i = [\alpha_i - \delta, \alpha_i + \delta] \subseteq \mathcal{A}$. The Jacobian event makes the sample moment monotone on all of $\mathcal{A}$, so there can be no other zero. The observable selector therefore coincides with the local branch on this intersection. Both selectors have bounded fallbacks; their difference contributes only exponentially small moments. The shared scalar concentration and root inequality give the stated tail and all fixed moments.

Proposition C.5 now supplies the raw-root expansion with the population-Hessian $q_{iT,2}$ from (C.6). Clipping changes the raw root by at most $|\Delta_i| \mathbf{1}\{|\Delta_i| > b_T\}$; integrating its tail gives

$$\mathbb{E}_i |\hat{\alpha}_i - \check{\alpha}_i|^r \leq C_r b_T^r N^{-cC_b^2} + C_r \exp(-cT).$$

For $2 \leq r \leq 8$, $cC_b^2 > 16$ and $N \asymp T$ make its L_i^r norm $o(T^{-3/2})$. Projection about α_i is a contraction, so the clipped-root moment bounds also follow. Exact window counts (C.3) applied to the two products in $q_{iT,2}$ give $\beta_{iT,2} = T \mathbb{E}_i q_{iT,2} = \beta_i + O(T^{-1})$ with β_i from (6.2). Set $\tilde{q}_{iT,2} = q_{iT,2} - \mathbb{E}_i q_{iT,2}$ and $r_{iT,2} = d_i - \bar{\xi}_i - \tilde{q}_{iT,2} - \beta_i/T$. The main text suppresses the last index 2; the $O(T^{-2})$ difference between the two mean terms is included in $r_{iT,2}$. This proves (6.3). Markov's inequality gives the average moments, and the scalar tail plus a union bound gives the maximum rate. For bounded deterministic w_i , unit independence gives

$$\mathbb{E} \left| \frac{\sqrt{NT}}{N} \sum_i w_i \tilde{q}_{iT,2} \right|^2 = \frac{T}{N} \sum_i w_i^2 \text{Var}_i(q_{iT,2}) \leq C/T.$$

Hence the centered quadratic term is $o_p(1)$ at the $\sqrt{NT}$ scale after averaging across units. $\square$

Lemma G.2 (Scalar-score expansion). *Under the maintained conditions,*

$$\frac{1}{NT} \sum_{i,t} \psi_{it,K}(\hat{\alpha}_i) = \frac{1}{NT} \sum_{i,t} x_{it,K} + \frac{B_K}{T} + O_p\left(\frac{1}{T\sqrt{N}} + T^{-3/2}\right). \quad (\text{G.3})$$

The exact clipped population coefficient differs from B_K by $O(T^{-1/2})$.

Proof. Use the score $\psi_{it,K}$, its derivatives, and the means μ_{iK}, ν_{iK} from the main text. These arrays have bounded envelopes by (G.1). Let $r_i^0 = d_i - \bar{\xi}_i - q_{iT,2}$. Taylor's formula and addition

and subtraction of the two population derivatives give the exact decomposition

$$\begin{aligned}\bar{\psi}_{iK}(\hat{\alpha}_i) &= \bar{\psi}_{iK} + \mu_{iK}\bar{\xi}_i + (\bar{\psi}_{iK}^{(1)} - \mu_{iK})\bar{\xi}_i + \mu_{iK}q_{iT,2} + \frac{1}{2}\nu_{iK}(\bar{\xi}_i)^2 + R_i, \\ R_i &= (\bar{\psi}_{iK}^{(1)} - \mu_{iK})q_{iT,2} + \bar{\psi}_{iK}^{(1)}r_i^0 + \frac{1}{2}(\bar{\psi}_{iK}^{(2)} - \nu_{iK})(\bar{\xi}_i)^2 \\ &\quad + \frac{1}{2}\bar{\psi}_{iK}^{(2)}(d_i^2 - (\bar{\xi}_i)^2) + \frac{d_i^3}{2T} \sum_t \int_0^1 (1-s)^2 \partial_a^3 \psi_{it,K}(\alpha_i + sd_i) ds.\end{aligned}$$

The path remains in $[0, 1]$. Finite-dependence moment bounds and Lemma G.1 yield, term by term,

$$\begin{aligned}\mathbb{E}_i |R_i| &\leq \|\bar{\psi}_{iK}^{(1)} - \mu_{iK}\|_{L_i^2} \|q_{iT,2}\|_{L_i^2} + C\|r_i^0\|_{L_i^1} + C\|\bar{\psi}_{iK}^{(2)} - \nu_{iK}\|_{L_i^2} \|\bar{\xi}_i\|_{L_i^4}^2 \\ &\quad + C\|d_i - \bar{\xi}_i\|_{L_i^2} \|d_i + \bar{\xi}_i\|_{L_i^2} + C\mathbb{E}_i |d_i|^3 \leq CT^{-3/2}.\end{aligned}$$

No third-moment cancellation is used here. Each of the three displayed quadratic terms preceding R_i has L_i^2 norm $O(T^{-1})$; for the first this follows from the fourth moments of its two centered averages. Their centered cross-unit average is therefore $O_p(1/(T\sqrt{N}))$. Equation (C.3) gives their expectations as the three terms in (6.4), divided by T , with total error $O(T^{-2})$. Finally, the grid projection implies $N^{-1} \sum_i \mathbb{E}_i[\psi_{it,K}] = 0$. Averaging the exact decomposition proves (G.3). Taking expectations first and multiplying by T gives the population error $O(T^{-1/2})$ separately from the centered fluctuation. $\square$

G.2 Uniform Gram Bounds over Coordinate Shifts

Lemma G.3 (Random-coordinate Gram bound). *For $0 \leq \rho \leq \lambda_K$, let*

$$\begin{aligned}\mathcal{D}_\rho &= \{d \in \mathbb{R}^N : \max_i |d_i| \leq \rho, \alpha_i + d_i \in [0, 1]\}, \\ G(d) &= \frac{1}{NT} \sum_{i,t} p_K(X_{it}, \alpha_i + d_i) p_K(X_{it}, \alpha_i + d_i)^\top.\end{aligned}$$

There is an event $\mathcal{E}_x$, depending only on the covariates, with $\mathbb{P}(\mathcal{E}_x) \rightarrow 1$, on which, simultaneously for every $0 \leq \rho \leq \lambda_K$, $d \in \mathcal{D}_\rho$, $s \in [0, 1]$, and $c \in \mathbb{R}^K$,

$$\begin{aligned}\frac{1}{NT} \sum_{i,t} [c^\top \partial_a^r p_K(X_{it}, \alpha_i + sd_i)]^2 &\leq C\lambda_K^{-2r} \|c\|^2, \quad r = 0, 1, \\ \sup_{d \in \mathcal{D}_\rho} \|G(d) - G(0)\|_{\text{op}} &\leq C\{\rho/\lambda_K + (\rho/\lambda_K)^2\}.\end{aligned}$$

Also

$$\begin{aligned}\|G(0) - Q_K\|_{\text{op}} &= O_p\left(\sqrt{\frac{K \log(2K)}{NT}} + \frac{K \log(2K)}{NT}\right), \\ \mathbb{P}(c_* I_K \preceq \hat{Q}_K \preceq C_* I_K) &\longrightarrow 1\end{aligned}$$

for fixed positive c_*, C_* .

Proof. Write the tensor factors as $p_{K_x}^x(x)$ and $p_{K_a}^a(a)$, as in Assumption 5.1(iv). Consider a pair of overlapping basis functions j, k , and put $U_{it} = p_{K,j}(X_{it}, \alpha_i) p_{K,k}(X_{it}, \alpha_i)$. Local support and normalization give $|U_{it}| \leq CK$ and

$$\mathbb{E}_X [p_{j_x}^x(X)^2 p_{k_x}^x(X)^2] \leq CK_x, \quad \#\{i : \alpha_i \in \text{supp}(p_{j_a}^a p_{k_a}^a)\} \leq CN\lambda_K.$$

The count follows from the minimum spacing and support length $C\lambda_K$. Consequently

$$\frac{1}{NT} \sum_{i,t} \mathbb{E}[U_{it}^2] \leq CK_x K_a^2 \lambda_K \leq CK, \quad \sum_{i,t} \text{Var}(U_{it}) \leq CNTK.$$

This is an average second-moment bound; the second moment for a single unit can instead have order $K_x K_a^2$. Independence of the covariates and Bernstein's inequality imply, for $u \geq 1$,

$$\mathbb{P} \left(\left| \frac{1}{NT} \sum_{i,t} (U_{it} - \mathbb{E} U_{it}) \right| > C \left\{ \sqrt{\frac{Ku}{NT}} + \frac{Ku}{NT} \right\} \right) \leq 2e^{-u}.$$

There are at most CK nonzero entries and at most C entries per row. A union bound followed by $\|A\|_{\text{op}} \leq \max_j \sum_k |A_{jk}|$ proves the asserted oracle rate. The same argument within each unit gives

$$G_{x,i} = \frac{1}{T} \sum_t p_{K_x}^x(X_{it}) p_{K_x}^x(X_{it})^\top, \\ \max_i \|G_{x,i} - \mathbb{E}_X[p_{K_x}^x(X) p_{K_x}^x(X)^\top]\|_{\text{op}} = O_p \left(\sqrt{\frac{K_x \log(2NK_x)}{T}} + \frac{K_x \log(2NK_x)}{T} \right) = o_p(1).$$

The continuous Riesz bound permits a fixed M_x such that $\mathcal{E}_x = \{\max_i \|G_{x,i}\|_{\text{op}} \leq M_x\}$ has probability approaching one.

On this event fix any permitted shift and put $a_i = \alpha_i + sd_i$. Each a_i stays in $[0, 1]$. If J is a full positive knot span, then

$$n_J := \#\{i : a_i \in J\} \leq 1 + \frac{N}{c_\alpha}(|J| + 2\rho) \leq CN\lambda_K.$$

Assign a point at a knot to either adjacent span. This count allows shifted points to coincide. At the boundary, use the full positive span of length comparable to λ_K , rather than its intersection with the clipping interval. Reshape c into $B \in \mathbb{R}^{K_x \times K_a}$ with $\|B\|_F = \|c\|$ and $c^\top p_K(x, a) = p_{K_x}^x(x)^\top B p_{K_a}^a(a)$. For the vector polynomial $f(a) = B(p_{K_a}^a)^{(r)}(a)$, scalar norm equivalence on a full span, applied to every component, gives

$$\sup_{a \in J} \|f(a)\|^2 \leq \sum_{j=1}^{K_x} \sup_{a \in J} |f_j(a)|^2 \leq C\lambda_K^{-1} \sum_{j=1}^{K_x} \int_J |f_j(a)|^2 da, \\ \frac{1}{N} \sum_i \|B(p_{K_a}^a)^{(r)}(a_i)\|^2 \leq \sum_J \frac{n_J}{N} \sup_{a \in J} \|f(a)\|^2 \leq C \int_0^1 \|B(p_{K_a}^a)^{(r)}(a)\|^2 da \leq C\lambda_K^{-2r} \|c\|^2.$$

The last inequality applies the one-dimensional Riesz and inverse bounds to each row of B . Their constants depend on the fixed degree and knot ratios, not on K_x . The tensor quadratic form now yields

$$\frac{1}{NT} \sum_{i,t} [c^\top \partial_a^r p_K(X_{it}, a_i)]^2 = \frac{1}{N} \sum_i [B(p_{K_a}^a)^{(r)}(a_i)]^\top G_{x,i} [B(p_{K_a}^a)^{(r)}(a_i)] \\ \leq \frac{M_x}{N} \sum_i \|B(p_{K_a}^a)^{(r)}(a_i)\|^2 \leq C\lambda_K^{-2r} \|c\|^2.$$

This deterministic argument on $\mathcal{E}_x$ holds simultaneously for every coefficient and every shift satisfying $a_i \in [0, 1]$, including data-dependent shifts.

For $\|c\| = 1$, put $z_{it} = c^\top p_K(X_{it}, \alpha_i)$ and $\delta z_{it} = c^\top \{p_K(X_{it}, \alpha_i + d_i) - p_K(X_{it}, \alpha_i)\}$. The fundamental theorem of calculus and Jensen's inequality give

$$\begin{aligned}\delta z_{it} &= d_i \int_0^1 c^\top \partial_a p_K(X_{it}, \alpha_i + s d_i) ds, \\ \|\delta z\|_{NT,2}^2 &\leq \rho^2 \int_0^1 \frac{1}{NT} \sum_{i,t} [c^\top \partial_a p_K(X_{it}, \alpha_i + s d_i)]^2 ds \leq C(\rho/\lambda_K)^2, \\ |c^\top \{G(d) - G(0)\}c| &= \left| \frac{1}{NT} \sum_{i,t} \{2z_{it}\delta z_{it} + (\delta z_{it})^2\} \right| \\ &\leq 2\|z\|_{NT,2}\|\delta z\|_{NT,2} + \|\delta z\|_{NT,2}^2 \\ &\leq C\{\rho/\lambda_K + (\rho/\lambda_K)^2\}.\end{aligned}$$

Take suprema over c and d . By Lemma G.1, the actual shifts satisfy $\max_i |d_i| = O_p(b_T)$, and $b_T/\lambda_K \rightarrow 0$. Combining the perturbation bound with the oracle rate and (G.1) proves the last assertion. When $d_X = 0$, $p_{K_x}^x \equiv 1$, $G(0) = Q_K$ exactly, and $\mathcal{E}_x$ can be the whole sample space. $\square$

G.3 Nuisance Rate, Acceptance, and Exact Profiling

Lemma G.4 (Nuisance rate and derivative consistency). *Put $r_{c,K} = \sqrt{K}\lambda_K^{-2}/T$. Then $\mathbb{P}(\mathcal{O}_T) \rightarrow 1$, $\hat{J}_K - J_K = o_p(1)$, and*

$$\begin{aligned}\|\hat{\gamma}_K - \gamma_K\| + \|\hat{\delta}_K - \delta_K\| &= O_p(r_{c,K}), \\ \sup_{x,a} |\partial_a^r (\hat{l}_K - l_K)(x, a)| &= O_p(K\lambda_K^{-(r+2)}/T) = o_p(1), \quad r = 0, 1, 2,\end{aligned}$$

where $(\hat{l}_K, l_K)$ is either fitted and population projection pair. On $\mathcal{O}_T$, the exact profile identity (A.4) holds for these objects, with a profile remainder $o_p((NT)^{-1/2})$. Also $\hat{\theta}_K - \theta_K = O_p(T^{-1})$.

Proof. For $(R, l_K) = (D, m_K)$ or (Y, ℓ_K) , let $F_{it,K}(a) = p_K(X_{it}, a)\{R_{it} - l_K(X_{it}, a)\}$. Write $c_{R,K}$ for the corresponding projection coefficient γ_K or δ_K , with candidate $\check{c}_{R,K}$ and reported fit $\hat{c}_{R,K}$. These are the original nuisance normal-equation contributions. Product differentiation gives

$$\begin{aligned}F' &= p_a(R - l) - pl_a, \\ F'' &= p_{aa}(R - l) - 2p_a l_a - pl_{aa}, \\ F''' &= p_{aaa}(R - l) - 3p_{aa} l_a - 3p_a l_{aa} - pl_{aaa}.\end{aligned}$$

By (G.1), the entire vector $F^{(r)}$ has envelope $A_r = C\sqrt{K}\lambda_K^{-r}$, $0 \leq r \leq 3$. Write $\mu_{F,i} = \mathbb{E}_i[F_{it,K}(\alpha_i)]$ and $\mu_{F',i} = \mathbb{E}_i[F'_{it,K}(\alpha_i)]$. Since $\sum_i \mu_{F,i} = 0$ by projection, Taylor's formula gives exactly

$$\begin{aligned}\frac{1}{N} \sum_i \bar{F}_{i,K}(\hat{\alpha}_i) &= \frac{1}{N} \sum_i (\bar{F}_{i,K} - \mu_{F,i}) + \frac{1}{N} \sum_i \mu_{F',i} \bar{\xi}_i \\ &\quad + \frac{1}{N} \sum_i (\bar{F}'_{i,K} - \mu_{F',i}) \bar{\xi}_i + \frac{1}{N} \sum_i \bar{F}'_{i,K}(d_i - \bar{\xi}_i) \\ &\quad + \frac{1}{2N} \sum_i \bar{F}''_{i,K} d_i^2 + \frac{1}{N} \sum_i R_{F,i}, \quad \|R_{F,i}\| \leq CA_3 |d_i|^3.\end{aligned}$$

Derivatives without an argument are evaluated at α_i . Independence across units and Lemma C.1

imply

$$\begin{aligned}\mathbb{E} \left\| \frac{1}{N} \sum_i (\bar{F}_{i,K} - \mu_{F,i}) \right\|^2 &\leq CA_0^2/(NT), \\ \mathbb{E} \left\| \frac{1}{N} \sum_i \mu_{F',i} \bar{\xi}_i \right\|^2 &= \frac{1}{N^2} \sum_i \|\mu_{F',i}\|^2 \mathbb{E}_i(\bar{\xi}_i)^2 \leq CA_1^2/(NT).\end{aligned}$$

The second line uses $\mathbb{E}_i \bar{\xi}_i = 0$, not a zero derivative mean. For the remaining four terms, the expected norms are bounded by

$$\begin{aligned}\frac{1}{N} \sum_i \|\bar{F}'_{i,K} - \mu_{F',i}\|_{L_i^2} \|\bar{\xi}_i\|_{L_i^2} &\leq CA_1/T, \\ \frac{A_1}{N} \sum_i \mathbb{E}_i |d_i - \bar{\xi}_i| &\leq CA_1/T, \\ \frac{A_2}{2N} \sum_i \mathbb{E}_i d_i^2 &\leq CA_2/T, \\ \frac{CA_3}{N} \sum_i \mathbb{E}_i |d_i|^3 &\leq CA_3/T^{3/2}.\end{aligned}$$

Markov's inequality and $N \asymp T$, $\sqrt{T}\lambda_K \rightarrow \infty$ show that all six terms are $O_p(\sqrt{K}\lambda_K^{-2}/T)$. This calculation permits arbitrary within-unit correlation between F and ξ .

On the Gram acceptance event, the candidate coefficient error obeys the exact normal equation

$$\check{c}_{R,K} - c_{R,K} = \hat{Q}_K^{-1} \frac{1}{NT} \sum_{i,t} F_{it,K}(\hat{\alpha}_i).$$

Lemma G.3 bounds its inverse in probability and proves the candidate coefficient rate. It remains to prove acceptance of the candidate denominator before transferring this conclusion to reported objects. For the candidate treatment residual, the Gram energy bound gives

$$\begin{aligned}\|\check{u}_D - u_{D,K}\|_{NT,2} &\leq \|\hat{p}^\top(\check{\gamma}_K - \gamma_K)\|_{NT,2} + C \left(\frac{1}{N} \sum_i d_i^2 \right)^{1/2} \\ &= O_p(r_{c,K} + T^{-1/2}), \\ |\check{J}_K - \frac{1}{NT} \sum_{i,t} u_{D,it,K}^2| &\leq \|\check{u}_D - u_{D,K}\|_{NT,2} \{2\|u_{D,K}\|_{NT,2} + \|\check{u}_D - u_{D,K}\|_{NT,2}\} = o_p(1).\end{aligned}$$

The finite-range variance bound implies $\frac{1}{NT} \sum_{i,t} u_{D,it,K}^2 - J_K = O_p((NT)^{-1/2})$. Also $J_K = V + \frac{1}{NT} \sum_{i,t} \mathbb{E}_i[r_{m,it,K}^2] \rightarrow V > 0$. Since the Gram test passes with probability approaching one and $a_T \rightarrow 0$, the candidate denominator passes with probability approaching one. This proves $\mathbb{P}(\mathcal{O}_T) \rightarrow 1$, and Remark D.18 transfers the rates to the reported coefficients and denominator.

On $\mathcal{O}_T$, the candidate construction satisfies all three profile normal equations. Lemma A.2 therefore gives (A.4) for these coordinates and coefficients. The coefficient rate and (6.7) give

$$\sqrt{NT}|e_\gamma^\top \hat{Q}_K e_\pi| = O_p(T(r_{c,K})^2) = O_p(K\lambda_K^{-4}/T) = o_p(1).$$

Lemma G.2 now proves $\hat{\theta}_K - \theta_K = O_p(T^{-1})$. Finally, for $r = 0, 1, 2$,

$$\sup_{x,a} |\partial_a^r (\hat{l}_K - l_K)(x, a)| \leq C\sqrt{K}\lambda_K^{-r} \|\hat{c}_{R,K} - c_{R,K}\| = O_p(K\lambda_K^{-(r+2)}/T) = o_p(1),$$

as claimed. $\square$

G.4 Feasible First-Step Influence and Bias

For the random-Jacobian covariance, write $C_{G\xi,i} = \sum_{|j| \leq L} \mathbb{E}_i[\{\partial_a g(W_{i0}, \alpha_i) - G_i\} \xi_{ij}]$ and $\widehat{C}_{G\xi,i} = \frac{1}{T} \sum_{|j| \leq L} \sum_{t \in I_T(j)} (\widehat{G}_{it} - \widehat{G}_i) \widehat{\xi}_{i,t+j}$.

Lemma G.5 (Truncated-inverse and lag-statistic moments). *For every fixed finite $r \geq 2$, uniformly in i ,*

$$\begin{aligned} \|\widehat{G}_i - G_i\|_{L_i^r} + \|\widehat{H}_i - H_i\|_{L_i^r} + \|\widehat{G}^{-1}_i - G_i^{-1}\|_{L_i^r} + \left\| \max_{t \leq T} |\widehat{\xi}_{it} - \xi_{it}| \right\|_{L_i^r} &\leq C_r T^{-1/2}, \\ \|\widehat{\Omega}_i - \Omega_i\|_{L_i^2} + \|\widehat{C}_{G\xi,i} - C_{G\xi,i}\|_{L_i^2} + \|\widehat{\beta}_i - \beta_i\|_{L_i^2} &\leq C T^{-1/2}. \end{aligned}$$

Furthermore, $\max_{i,t} |\widehat{\xi}_{it}| = O_p(1)$ and $N^{-1} \sum_i (\widehat{\beta}_i^2 + \widehat{\Omega}_i^2) = O_p(1)$.

Proof. The event $\mathcal{J}_i$ in (G.2) controls the sample Jacobian over the entire interval. On that event, for large T ,

$$|\widehat{G}_i| \geq c_G/2 > a_T, \quad |\widehat{G}^{-1}_i| \leq 2/c_G.$$

On the whole sample space only the bound $|\widehat{G}^{-1}_i| \leq c_{\text{tol}}^{-1} T^{1/4}$ is available. Passing the observable threshold does not itself imply a constant envelope. Bounded derivatives and Lemma G.1 give

$$\begin{aligned} |\widehat{G}_i - G_i| &\leq |M_{iT}| + C|d_i|, \\ |\widehat{H}_i - H_i| &\leq |\overline{H}_i - H_i| + C|d_i|. \end{aligned}$$

On $\mathcal{J}_i$,

$$|\widehat{G}^{-1}_i - G_i^{-1}| = \frac{|\widehat{G}_i - G_i|}{|\widehat{G}_i G_i|} \leq C(|M_{iT}| + |d_i|).$$

The complementary event has probability $C \exp(-cT)$ and an envelope $C(1 + T^{1/4})$. Its L_i^r contribution is at most $C_r(1 + T^{1/4}) \exp(-cT/r)$, proving the inverse rate.

To control the maximum over dates, let $z_{it} = g(W_{it}, \alpha_i)$ and $\widehat{z}_{it} = g(W_{it}, \widehat{\alpha}_i) - T^{-1} \sum_s g(W_{is}, \widehat{\alpha}_i)$. The exact differences are

$$\begin{aligned} \widehat{z}_{it} - z_{it} &= \{g(W_{it}, \widehat{\alpha}_i) - g(W_{it}, \alpha_i)\} - \frac{1}{T} \sum_s \{g(W_{is}, \widehat{\alpha}_i) - g(W_{is}, \alpha_i)\} - \bar{z}_i, \\ \widehat{\xi}_{it} - \xi_{it} &= -(\widehat{G}^{-1}_i - G_i^{-1})z_{it} - \widehat{G}^{-1}_i(\widehat{z}_{it} - z_{it}). \end{aligned}$$

Therefore, on $\mathcal{J}_i$,

$$D_{\xi,i} := \max_t |\widehat{\xi}_{it} - \xi_{it}| \leq C(|M_{iT}| + |d_i| + |\bar{z}_i|).$$

The right side contains only unit-level averages and the coordinate error; no union bound over t is needed. Its L_i^r norm is $O(T^{-1/2})$. On $\mathcal{J}_i^c$, $D_{\xi,i} \leq C(1 + T^{1/4})$, whose moment contribution is exponentially small. This proves the influence rate even when clipping binds. The subtracted mean in (6.9) is zero when selection succeeds and clipping is inactive.

Set $s_{it} = \partial_a g(W_{it}, \alpha_i) - G_i$ and $\widehat{s}_{it} = \widehat{G}_{it} - \widehat{G}_i$. Their maximum difference satisfies $D_{s,i} \leq 2C|d_i| + |M_{iT}|$. For $(U, V) = (\xi, \xi)$ or (s, ξ) , the oracle factors are bounded, stationary and jointly finite-range under Assumption 6.1. Lemma C.1 gives lag-average variance C/T ; stationarity gives its exact expectation $(1 - |j|/T) \mathbb{E}_i[U_{i0} V_{ij}]$. The oracle lag statistics have L_i^2 error $O(T^{-1/2})$ relative

to their population sums. On $\mathcal{J}_i$, replacing one factor at a time gives

$$\left| \frac{1}{T} \sum_{|j| \leq L} \sum_{t \in I_T(j)} \{ \widehat{\xi}_{it} \widehat{\xi}_{i,t+j} - \xi_{it} \xi_{i,t+j} \} \right| \leq CD_{\xi,i},$$

$$\left| \frac{1}{T} \sum_{|j| \leq L} \sum_{t \in I_T(j)} \{ \widehat{s}_{it} \widehat{\xi}_{i,t+j} - s_{it} \xi_{i,t+j} \} \right| \leq C(D_{s,i} + D_{\xi,i}).$$

On the complementary event the whole-space envelopes, with constants allowed to depend on the fixed c_{tol} , are

$$\max_t |\widehat{\xi}_{it}| \leq CT^{1/4}, \quad |\widehat{C}_{G\xi,i}| \leq CT^{1/4}, \quad |\widehat{\Omega}_i| \leq CT^{1/2}, \quad |\widehat{\beta}_i| \leq CT^{3/4}.$$

Their L_i^2 contributions on $\mathcal{J}_i^c$ are these polynomials multiplied by $C \exp(-cT/2)$. On $\mathcal{J}_i$, subtract (6.2) from (6.10) and replace factors to obtain

$$|\widehat{\beta}_i - \beta_i| \leq C(|\widehat{G}^{-1}_i - G_i^{-1}| + |\widehat{C}_{G\xi,i} - C_{G\xi,i}| + |\widehat{H}_i - H_i| + |\widehat{\Omega}_i - \Omega_i|).$$

This completes all moment bounds. Markov's inequality yields the average square bound. Finally $\mathbb{P}(\cap_i \mathcal{J}_i) \geq 1 - CN \exp(-cT) \rightarrow 1$, which gives the maximum bound. Finite dependence has been applied only to the oracle arrays, never to the estimated influences using the full history. $\square$

G.5 Correction, Variance, and the Limit Distribution

Proof of Theorem 6.2. Consistency of all three channels. Lemma G.4, bounded third derivatives of the oracle score, and $\max_i |d_i| = O_p(b_T)$ imply

$$\max_{i,t} |\widehat{\psi}_{it,K}^{(1)} - \psi_{it,K}^{(1)}| + \max_{i,t} |\widehat{\psi}_{it,K}^{(2)} - \psi_{it,K}^{(2)}| = O_p(b_T + K\lambda_K^{-4}/T + T^{-1}) = o_p(1).$$

Here one first replaces coefficients at $\widehat{\alpha}_i$, using the fitted derivatives through order two, then moves the oracle score back to α_i . For the reuse channel, the exact difference is

$$\begin{aligned} \widehat{B}_K^{(1)} - B_K^{(1)} &= \frac{1}{NT} \sum_i \sum_{|j| \leq L} \sum_{t \in I_T(j)} (\widehat{\psi}_{it,K}^{(1)} - \psi_{it,K}^{(1)}) \widehat{\xi}_{i,t+j} \\ &\quad + \frac{1}{NT} \sum_i \sum_{|j| \leq L} \sum_{t \in I_T(j)} \psi_{it,K}^{(1)} (\widehat{\xi}_{i,t+j} - \xi_{i,t+j}) \\ &\quad + \frac{1}{NT} \sum_i \sum_{|j| \leq L} \sum_{t \in I_T(j)} \{ \psi_{it,K}^{(1)} \xi_{i,t+j} - \mathbb{E}_i[\psi_{it,K}^{(1)} \xi_{i,t+j}] \} \\ &\quad - \frac{1}{NT} \sum_i \sum_{|j| \leq L} |j| \mathbb{E}_i[\psi_{i0,K}^{(1)} \xi_{ij}]. \end{aligned}$$

The four terms have respective orders $o_p(1)$, $O_p(T^{-1/2})$, $O_p((NT)^{-1/2})$, and $O(T^{-1})$: use derivative consistency and Lemma G.5 for the first two, finite-range variance and independent units for the third, and the exact window weights for the fourth.

For the first-step channel, expand

$$\begin{aligned} \widehat{\mu}_{iK}\widehat{\beta}_i - \mu_{iK}\beta_i &= (\widehat{\mu}_{iK} - \overline{\psi}_{iK}^{(1)})\widehat{\beta}_i + (\overline{\psi}_{iK}^{(1)} - \mu_{iK})\widehat{\beta}_i + \mu_{iK}(\widehat{\beta}_i - \beta_i), \\ \frac{1}{N} \sum_i |(\overline{\psi}_{iK}^{(1)} - \mu_{iK})\widehat{\beta}_i| &\leq \left\{ \frac{1}{N} \sum_i (\overline{\psi}_{iK}^{(1)} - \mu_{iK})^2 \right\}^{1/2} \left\{ \frac{1}{N} \sum_i \widehat{\beta}_i^2 \right\}^{1/2} = O_p(T^{-1/2}). \end{aligned}$$

The average absolute values of the other two terms are $o_p(1)$ and $O_p(T^{-1/2})$. The same three-factor decomposition with $(\psi^{(1)}, \mu, \beta)$ replaced by $(\psi^{(2)}, \nu, \Omega)$ proves curvature consistency. No independence between these fitted factors is used. Thus $\widehat{B}_K - B_K = o_p(1)$.

Oracle influence and positive variance. At the true coordinate,

$$\begin{aligned} u_{D,K} &= v + r_m, & q_K &= \varepsilon + r_h - b_K(v + r_m), \\ \partial_a \psi_{it,K}(\alpha_i) &= -\partial_a m_K \{\varepsilon + r_h - b_K(v + r_m)\} \\ &\quad - (v + r_m) \{\partial_a h_K - b_K \partial_a m_K\}. \end{aligned}$$

Taking expectation over (X, v, ε) eliminates the terms linear in v or ε and proves (6.5). Together with (G.1), it gives

$$\begin{aligned} \sup_i |\mu_{iK}| &\leq C \lambda_K^{\min\{\tau_m, \tau_h\}}, & \sup_i |\beta_i| &\leq C, & |B_K^{(3)}| &\leq C \lambda_K^{\min\{\tau_m, \tau_h\}} \rightarrow 0, \\ \sup_i \|\psi_{it,K}(\alpha_i) - v_{it}\varepsilon_{it}\|_{L_i^2} &\rightarrow 0, \\ \sup_i \|\phi_{it,K} - V^{-1}v_{it}\varepsilon_{it}\|_{L_i^2} &\rightarrow 0. \end{aligned}$$

The centered influence is defined in (6.12). The last bound also centers the oracle score, because $\mathbb{E}[v_{it}\varepsilon_{it}] = 0$. For each of the finitely many lags, Cauchy-Schwarz transfers this bound to the covariance products. Therefore

$$\sigma_{KNT}^2 := \frac{1}{N} \sum_i \sum_{|j| \leq L} (1 - |j|/T) \mathbb{E}_i[\phi_{i0,K} \phi_{ij,K}] \rightarrow \sigma^2 > 0.$$

Feasible variance and cluster limit. Use the uncentered oracle $\phi_{it,K}^0 = J_K^{-1} \{\psi_{it,K}(\alpha_i) + \mu_{iK}\xi_{it}\}$. Its difference from ϕ is constant within each unit. Lemma G.4 and the first-step moment bounds give $\|\widehat{\psi} - \psi\|_{NT,2} = o_p(1)$, $N^{-1} \sum_i (\widehat{\mu}_{iK} - \mu_{iK})^2 = o_p(1)$, and $\|\widehat{\xi} - \xi\|_{NT,2} = O_p(T^{-1/2})$. Replacing the coordinate factors and the denominator gives

$$\begin{aligned} \|\widehat{\mu}\widehat{\xi} - \mu\xi\|_{NT,2} &\leq \max_{i,t} |\widehat{\xi}_{it}| \left\{ \frac{1}{N} \sum_i (\widehat{\mu}_{iK} - \mu_{iK})^2 \right\}^{1/2} + C \|\widehat{\xi} - \xi\|_{NT,2} = o_p(1), \\ \|\widehat{\phi} - \phi^0\|_{NT,2} &\leq |\widehat{J}_K^{-1} - J_K^{-1}| \|\psi + \mu\xi\|_{NT,2} + |\widehat{J}_K^{-1}| \{\|\widehat{\psi} - \psi\|_{NT,2} + \|\widehat{\mu}\widehat{\xi} - \mu\xi\|_{NT,2}\} = o_p(1). \end{aligned}$$

The centered oracle is stationary and finite-range with bounded fourth moments. The preceding paragraph establishes the variance limit $\sigma_{KNT}^2 \rightarrow \sigma^2 > 0$. Lemma E.5 therefore proves the raw variance consistency and $N^{-1/2} \sum_i \Phi_{iT,K} \rightsquigarrow \mathcal{N}(0, \sigma^2)$. The smooth procedure's positive variance floor is inactive with probability approaching one.

Final profile expansion. Substitute (G.3) into the exact profile identity and subtract (6.11). On $\mathcal{O}_T$ this gives

$$\begin{aligned} \sqrt{NT}(\widetilde{\theta}_K - \theta_0) &= \widehat{J}_K^{-1} \frac{1}{\sqrt{NT}} \sum_{i,t} x_{it,K} + \sqrt{N/T} \widehat{J}_K^{-1} (B_K - \widehat{B}_K) \\ &\quad - \sqrt{NT} \widehat{J}_K^{-1} e_\gamma^\top \widehat{Q}_K e_\pi + \sqrt{NT}(\theta_K - \theta_0) + o_p(1). \end{aligned}$$

The correction error is $o_p(1)$ by the first part of the proof. The profile term is $O_p(K\lambda_K^{-4}/T) = o_p(1)$ by Lemma G.4. The target term is $O(T\lambda_K^{\tau_m+\tau_h}) = o(1)$ by the projection-error product identity. The score remainder after multiplication by $\sqrt{NT}$ is $O_p(N^{-1/2} + T^{-1/2}) = o_p(1)$. The leading centered sum is $O_p(1)$, so $\hat{J}_K^{-1}$ can be replaced by J_K^{-1} . This proves (6.13), and variance consistency gives studentization by Slutsky's theorem. The preceding arguments first proved $\mathbb{P}(\mathcal{O}_T) \rightarrow 1$, so all conclusions transfer to the globally reported procedure. Finally, omit the correction subtraction to obtain (6.14); a fixed noncentral limit additionally requires convergence of B_K , as stated. $\square$