Between Models

Counterfactual Frontiers across Causal-Response Experiments

Chae-Yeon Xon
Yonsei University

September 2026

Abstract

When do several structural models support the same policy conclusion? I compare the response distance needed to represent every declared family with the distance to the first policy-sign reversal. A positive gap gives a range in which family coverage and sign agreement coexist. The framework separates agreement on response fibers, the effect of additional evidence, and uncertainty in estimated responses. In a monetary application, the baseline four-family support has a positive gap. A lag-augmented local-projection refit gives negative gaps under three covariance specifications. Common-support comparisons trace the roles of response weighting and shock-path profiling. The positive target crossings are small, and a calendar re-estimation exercise produces gap values on both sides of zero. Each comparison is conditional on its response protocol, evaluated structural domain, and policy target.

JEL Codes: C12, C13, C32, C52, E52

Keywords: structural counterfactuals, partial identification, factorization, weak identification, generated responses, monetary policy

Yonsei University. Email: eunixon@yonsei.ac.kr. This draft is circulated for discussion; comments and suggestions are welcome.

1. Introduction

An analyst comparing monetary models may find that a delayed tightening lowers output and inflation in every model close to the estimated shock responses. How much of this agreement survives a change in the evidence used to fit the models? A response estimate, its uncertainty, and the restrictions on the fitted shock paths together determine which representations are close. I study the policy conclusion conditional on that response experiment. In the application, the shared intervention is a dated additive policy wedge announced at date zero.[1] Each model retains its feedback rule, so the induced interest-rate path can differ across models. Agreement is evaluated for that common intervention and a fixed outcome horizon. The reported object is the gap between family coverage and conclusion reversal under this protocol.

Two distances organize the comparison. The family-entry radius is the smallest response neighborhood containing at least one representation from every declared family.[2] The sign-reversal radius is the smallest neighborhood containing a nonnegative target against the negative-sign benchmark. When entry comes first, the interval between them contains all families and a common negative sign. A family can be represented by a well-fitting point while other compatible points in that family determine the policy boundary. Both distances therefore range over model representations. For experiment ξ, their gap is

$$\Delta(\xi) = c_{\text{sign}}(\xi) - c_{\text{entry}}(\xi)$$

The distinction matters when evidence is added. The new observations increase each model's distance by its residual discrepancy, conditional on the observations already retained. They can move the best representation of a difficult-to-fit family more than the first sign-reversing representation. The sign-agreement interval then contracts even as each response restriction becomes more informative. The distance decomposition and a finite example show how this contraction occurs. A common invertible recoding of responses and their covariance, by contrast, preserves the interval.

In the monetary application, the four-family baseline support has entry radius 3.5058 and first evaluated reversal 4.3567, giving a gap of 0.8509. Re-estimating responses on a 275-quarter calendar by lag-augmented local projections and refitting structural parameters and shock paths gives negative gaps under three covariance bandwidths. On the fifty-coordinate, bandwidth-28 support, reversal occurs at 17.1449 and all-family entry at 23.2840. Common-support comparisons trace the roles of profiling and response weighting. These are comparisons among specified implementations of the response experiment.

A sign crossing answers whether any compatible representation changes direction. Its size answers whether that change reaches an economically specified magnitude. The first bandwidth-28

[1] The wedge is applied at the same announcement date and in the same units across model families. Each family's feedback rule then determines its induced interest-rate path.

[2] Family entry records coverage by one representation from each declared family. Other representations in an entered family can lie beyond the same radius.

reversal is an output response of 0.00001217 in discounted cumulative units. On the evaluated local-projection support, positive output and inflation targets remain below 0.001. I report boundaries for larger target changes alongside the zero threshold. The exercise keeps the announced intervention, horizon, aggregation weights, and equilibrium selection fixed as the response protocol changes.

The identification analysis states what agreement means before a radius is introduced. A policy target is identified on an empirical response fiber if and only if every representation on that fiber gives the same target. The global result factors the target through the response map; the local characterization replaces fiber constancy with constant rank and neighborhood-wide kernel inclusion; the cross-family condition requires the resulting factor maps to agree on overlapping response images. The coverage-and-reversal frontier then measures the policy image when the exact response restriction is relaxed, indexed by the experiment that disciplines both thresholds.

Three nearby literatures clarify this object. Monetary-policy sufficiency and equivalence results relate observed responses to policy counterfactuals under specified structural restrictions (Beraja 2023; McKay and Wolf 2023; Caravello et al. 2026). Here the response restriction admits several families, and the reported frontier follows the first disagreement within that declared envelope. Breakdown analysis measures the departures that overturn a conclusion (Masten and Poirier 2020; Kang and Vasserman 2025). The additional entry threshold asks whether the same neighborhood represents every family being compared. Testing-risk approaches give response distance an interpretation in terms of statistical distinguishability (Hansen and Sargent 2001, 2008, 2010; Bonhomme and Weidner 2022).

Cerreia-Vioglio et al. (2026) provide decision criteria that incorporate uncertainty about structured models and their misspecification. The sign frontier here is a property of a target set; using it to choose a policy also requires the decision-maker's objective.

Giacomini et al. (2022) combine point- and set-identified models and characterize when observed data update their model probabilities. Their framework tracks uncertainty over identifying assumptions. The factorization result here asks whether model-specific policy targets agree on overlapping empirical-response images under one declared experiment.

The evaluated domain also enters the conclusion. A finite set of retained model solutions, a convex completion of joint response–target pairs, and a native nonlinear parameter region support different claims.[3] Partial-identification analysis supplies the set-valued language (Tamer 2010); structural counterfactual work develops bounds and inference on explicit model classes (Kalouptsidi et al. 2026). In this paper, the convex-completion argument applies to the evaluated pairs. Native-domain certification additionally uses verified approximation and numerical bounds. The monetary frontier comparisons use the finite supports and their stated convex completions. The appendix supplies local finite-horizon certificates around four compatible parameter points and states the additional requirements for coverage of a full parameter domain.

[3]The convex completion mixes the evaluated joint response–target pairs. Additional nonlinear model solutions require native-domain analysis.

Sampling analysis follows the estimated objects. The projection theorem profiles a studentized moment criterion over a structural nuisance union under uniform approximation, covariance, and compatibility conditions. The monetary resampling exercises retain a finite structural support. A 1,999-draw calendar calculation re-estimates responses and reprofiles shock paths; its central 95-percent gap range is $[-10.2187, 2.2214]$.[4] This is a descriptive resampling range. Minima, maxima, and active normalization constraints make the frontier nonsmooth, so coverage requires a separate justification of the resampling law (Fang and Santos 2019). The projection theorem and this numerical distribution are reported with their own conditions.

2. The identification problem

2.1. Structural envelopes and response protocols

Let $f \in \{1, \dots, F\}$ index model families. Family f has an economically feasible parameter domain Θ_f^{econ} and model instances $M_f(\theta)$. The *structural envelope* is the disjoint union

$$\mathfrak{E} = \bigsqcup_{f=1}^{F} \mathfrak{M}_f, \qquad \mathfrak{M}_f = \{M_f(\theta) : \theta \in \Theta_f^{\text{econ}}\}. \tag{1}$$

The disjoint union keeps family identity even when two implementations produce the same reduced-form response. The domain Θ_f^{econ} contains the equilibrium, probability, determinacy, borrowing, and calibration restrictions maintained for family f.

Each family supplies three maps:

$$S_{A,f} : \Theta_f^{\text{econ}} \to \mathbb{R}^{K_A}, \qquad S_{B,f} : \Theta_f^{\text{econ}} \to \mathbb{R}^{K_B}, \qquad T_{p,f} : \Theta_f^{\text{econ}} \to \mathbb{R}^{q}. \tag{2}$$

The training response S_A corresponds to the causal evidence used to fit or screen the model. The held-out response S_B is reserved for evaluation or residual calibration. The policy map T_p returns the counterfactual target under policy p. A scalar welfare or cumulative-response target has $q = 1$; a path-valued target has $q > 1$.

ASSUMPTION 1 (Common response protocol). *Across all families, S_A, S_B, and T_p use the same outcome definitions, units, shock normalization, policy experiment, horizon, timing convention, and aggregation rule. The implementation declares an equilibrium selection for every evaluated model. Policy-rate innovations are reported in annualized percentage points and enter the quarterly model interface after division by four. Thus 25 basis points equals 0.25 annualized percentage point, and coefficients that combine responses to this basis intervention are dimensionless.*

[4] The calculation redraws the response experiment and repeats shock-path profiling on the retained structural support.

Empirical equivalence is defined under the common intervention in Assumption 1. Families may still have different state variables and parameter dimensions because the envelope is a disjoint union. The common codomains in (2) place their responses in the same comparison space.

For a model written as $F_f(z, p; \theta) = 0$ with observation map $\Phi_f(z, \theta, p)$, local response derivatives are

$$D_\theta \Phi_f = \Phi_{f,\theta} - \Phi_{f,z} F_{f,z}^{-1} F_{f,\theta}, \qquad D_p \Phi_f = \Phi_{f,p} - \Phi_{f,z} F_{f,z}^{-1} F_{f,p}, \tag{3}$$

whenever the selected equilibrium has nonsingular $F_{f,z}$. These derivatives enter both the factorization analysis and generated-response inference.

2.2. Empirical fibers and policy sets

DEFINITION 1 (Empirical fiber and policy identified set). *For $s \in \mathbb{R}^{K_A}$, define*

$$\mathfrak{F}_A(s; \mathfrak{E}) = \{M \in \mathfrak{E} : S_A(M) = s\}, \qquad \mathfrak{I}_p(s; \mathfrak{E}) = \mathrm{cl}\{T_p(M) : M \in \mathfrak{F}_A(s; \mathfrak{E})\}. \tag{4}$$

The target is point identified relative to $\mathfrak{E}$ at s when the fiber is nonempty and $\mathfrak{I}_p(s; \mathfrak{E})$ is a singleton.

The closure includes limiting model sequences at the boundary of a family domain. An empty training fiber receives empirical-compatibility failure status. Identification status is assigned to a nonempty fiber.

DEFINITION 2 (Policy-feature identified set). *For a measurable feature $g : \mathbb{R}^q \to \mathcal{G}$, define*

$$\mathfrak{I}_{g,p}(s; \mathfrak{E}) = \mathrm{cl}\{g\{T_p(M)\} : M \in \mathfrak{F}_A(s; \mathfrak{E})\}. \tag{5}$$

The feature is identified relative to $\mathfrak{E}$ at s when the fiber is nonempty and $\mathfrak{I}_{g,p}(s; \mathfrak{E})$ is a singleton. For componentwise sign identification, $g(t) = (\operatorname{sgn} t_1, \ldots, \operatorname{sgn} t_q)$.

Definition 2 separates equality of target values from equality of a policy feature. The monetary application studies the componentwise sign and reports the value image to measure dispersion in magnitudes.

PROPOSITION 1 (Three logically distinct failures). *For a declared envelope $\mathfrak{E}$, training response $s_A^\dagger$, held-out response $s_B^\dagger$, and policy p, consider:*

1. *coverage failure:*
$$\mathfrak{F}_A(s_A^\dagger; \mathfrak{E}) = \varnothing;$$
2. *policy-identification failure conditional on coverage:*
$$\mathfrak{F}_A(s_A^\dagger; \mathfrak{E}) \neq \varnothing, \qquad \operatorname{diam} \mathfrak{I}_p(s_A^\dagger; \mathfrak{E}) > 0;$$

3. *held-out relation failure:*

$$s_B^\dagger \notin \{S_B(M) : M \in \mathfrak{F}_A(s_A^\dagger; \mathfrak{E})\}.$$

Coverage failure and conditional policy-identification failure are mutually exclusive. Held-out relation failure accompanies coverage failure and can also occur with either policy identification or policy-identification failure when the training fiber is nonempty.

The first asks whether the declared model envelope contains the empirical object. The second asks whether the observed block is sufficient for the policy target inside that envelope. The third asks whether the same fitted relation predicts a reserved observable block. In finite samples, projected-moment confidence procedures in Section 5 and Online Appendix C cover the population object in Definition 1.

3. Identification and factorization

Population identification requires the policy target to be constant on empirical fibers. At a model instance, kernel inclusion gives the corresponding first-order condition. Under constant rank, this differential restriction integrates over a neighborhood, and agreement on overlaps joins the maps from different families.

3.1. Global factorization on the envelope

THEOREM 1 (Global fiber-factorization characterization). *For a stated envelope $\mathfrak{E}$, the following are equivalent:*

1. *$T_p(M) = T_p(M')$ whenever $S_A(M) = S_A(M')$;*
2. *there is a unique set map $\lambda_p : S_A(\mathfrak{E}) \to \mathbb{R}^q$ such that*

$$T_p = \lambda_p \circ S_A. \tag{6}$$

Under either condition, $\mathcal{I}_p(s; \mathfrak{E})$ is a singleton for every $s \in S_A(\mathfrak{E})$.

The factor assigns to each observed response the target value shared by its preimages.[5]

3.2. Pointwise and local factorization within a family

Fix a family f and an interior point θ_0. Write

$$D_f = DS_{A,f}(\theta_0), \qquad C_{f,p} = DT_{p,f}(\theta_0).$$

The empirically invisible tangent directions are $\ker D_f$. Their policy image is $C_{f,p} \ker D_f$.

[5]The quotient is formed by observational equivalence under the declared training response. For a broader econometric treatment of identified sets generated by observational equivalence, see Tamer (2010).

PROPOSITION 2 (Pointwise differential factorization). *The following are equivalent:*

1. $\ker D_f \subseteq \ker C_{f,p}$;
2. *there is a unique linear map* $\Lambda_{f,p,\theta_0} : \operatorname{Im} D_f \to \mathbb{R}^q$ *such that*

$$C_{f,p} = \Lambda_{f,p,\theta_0} D_f. \tag{7}$$

The factor is constructed by $\Lambda_{f,p,\theta_0}(D_f h) = C_{f,p} h$. Kernel inclusion makes its value independent of the chosen preimage and gives linearity and uniqueness. Hence

$$T_{p,f}(\theta_0 + h) = T_{p,f}(\theta_0) + \Lambda_{f,p,\theta_0} D_f h + o(\|h\|). \tag{8}$$

Kernel failure supplies a model perturbation h with $D_f h = 0$ and $C_{f,p} h \neq 0$.

Neighborhood factorization uses constant rank and neighborhood-wide kernel inclusion. These assumptions produce a quotient chart for the following result.

THEOREM 2 (Local factorization characterization). *Suppose $S_{A,f}$ and $T_{p,f}$ are continuously differentiable on an open neighborhood U of θ_0,* $\operatorname{rank} DS_{A,f}(\theta) = r_f$ *for every $\theta \in U$, and*

$$\ker DS_{A,f}(\theta) \subseteq \ker DT_{p,f}(\theta) \quad \text{for every } \theta \in U. \tag{9}$$

Then there is a neighborhood $U_0 \subseteq U$ and a unique continuously differentiable map $\lambda_{f,p} : S_{A,f}(U_0) \to \mathbb{R}^q$ satisfying

$$T_{p,f}|_{U_0} = \lambda_{f,p} \circ S_{A,f}|_{U_0}. \tag{10}$$

The image is endowed with the local embedded-submanifold structure from the constant-rank theorem.

Constant-rank coordinates write the response as $(x, z) \mapsto (x, 0)$. Kernel inclusion makes the target derivative in z vanish. Connected vertical slices then give a unique local factor through x.

3.3. Agreement across model families

Theorem 2 gives a factor on a local image. Its constant-rank condition constructs the quotient chart, and kernel inclusion removes target variation along its connected vertical slices. Extending the factor to a whole family requires agreement across all parts of each response fiber. The following corollary assumes that global within-family step and states the remaining cross-family condition.

THEOREM 3 (Cross-family identification characterization). *Suppose each family admits a factor map $\lambda_{f,p} : S_{A,f}(\Theta_f^{\mathrm{econ}}) \to \mathbb{R}^q$ satisfying $T_{p,f} = \lambda_{f,p} \circ S_{A,f}$. Then the envelope identifies the policy target on its empirical image if and only if*

$$\lambda_{f,p}(s) = \lambda_{g,p}(s) \quad \text{for every } s \in S_{A,f}(\Theta_f^{\mathrm{econ}}) \cap S_{A,g}(\Theta_g^{\mathrm{econ}}) \tag{11}$$

for all f and g.

Equation (11) states the between-model restriction. A fitted predictor measures approximate agreement on an evaluated design. Family-held-out errors and boundary searches probe that design; agreement on the full family-image overlaps is the population requirement. This distinction fixes the domain to which a numerical finding applies.

3.4. Approximate factorization and policy visibility

For a candidate response map λ_p, define the nonlinear residual

$$R_{f,p}(\theta) = T_{p,f}(\theta) - \lambda_p\{S_{A,f}(\theta)\}. \tag{12}$$

Let $\|\cdot\|_{W_A}$ and $\|\cdot\|_{W_p}$ be declared norms and define the tolerance set

$$\mathcal{I}_p^{\epsilon_A}(s;\mathfrak{E}) = \mathrm{cl}\{T_p(M) : M \in \mathfrak{E},\ \|S_A(M) - s\|_{W_A} \le \epsilon_A\}. \tag{13}$$

LEMMA 1 (Approximate factorization bound). *Suppose $\epsilon_A, \epsilon_C, L_p \ge 0$,*

$$\sup_{M\in\mathfrak{E}} \|T_p(M) - \lambda_p\{S_A(M)\}\|_{W_p} \le \epsilon_C,$$

and λ_p is L_p-Lipschitz from the empirical norm to the policy norm on the relevant image. Then

$$\operatorname{diam} \mathcal{I}_p^{\epsilon_A}(s;\mathfrak{E}) \le 2(\epsilon_C + L_p\epsilon_A). \tag{14}$$

The bound follows from two relation residuals and the Lipschitz image of the distance between two admissible training responses. Finite-dimensional norm continuity carries the pairwise bound to the closure. Nonlinear model solutions determine ϵ_C; in-sample regression fit is reported separately.

For a scalar target with $C_{f,p} \neq 0$ and a symmetric positive definite covariance Σ_A, define the local identification strength

$$\kappa_{f,p}^2 = \inf_{h:C_{f,p}h=1} h' D_f' \Sigma_A^{-1} D_f h. \tag{15}$$

If $\ker D_f \subseteq \ker C_{f,p}$, finite dimensionality and positive definiteness imply $\kappa_{f,p} > 0$. Under kernel failure, a direction with $D_f h = 0$ and $C_{f,p}h \neq 0$ gives $\kappa_{f,p} = 0$. The scalar $\kappa_{f,p}$ records the training-response movement required for a normalized policy movement. The set $C_{f,p} \ker D_f$ records policy movements along the training-response kernel.

4. Testing-risk fibers and sign separation

The economic domains define the structural envelope. Statistical distance from the estimated causal response determines the thickness of its empirical fiber. Model admissibility is therefore governed by Θ_f^{econ}, while sampling resolution is measured in response space.

4.1. Thickened empirical fibers

Let $s_A^\dagger \in \mathbb{R}^{K_A}$ be the estimated training response and let $\Sigma_{A,T}$ be its finite-sample covariance on the retained coordinates. Define

$$d_{f,T}(\theta; s_A^\dagger)^2 = \{S_{A,f}(\theta) - s_A^\dagger\}' \Sigma_{A,T}^{-1} \{S_{A,f}(\theta) - s_A^\dagger\}. \tag{16}$$

For radius c, the family neighborhood, thickened fiber, and policy image are

$$\mathcal{N}_f(s_A^\dagger; c) = \{\theta \in \Theta_f^{\mathrm{econ}} : d_{f,T}(\theta; s_A^\dagger) \le c\}, \tag{17}$$

$$\mathfrak{F}_A(s_A^\dagger; c) = \bigsqcup_f \{f\} \times \mathcal{N}_f(s_A^\dagger; c), \tag{18}$$

$$\mathcal{I}_p(s_A^\dagger; c) = \mathrm{cl}\{T_{p,f}(\theta) : (f, \theta) \in \mathfrak{F}_A(s_A^\dagger; c)\}. \tag{19}$$

The zero-radius construction reduces to the exact fiber when $\Sigma_{A,T}$ is positive definite. A positive radius gives a set-valued response compatibility relation indexed by testing risk.

A change of units has a useful invariance check. For an invertible matrix A and a common offset b, transform both the model and empirical responses by $x \mapsto Ax + b$, and use $\Sigma^* = A\Sigma_{A,T}A'$. If $r = S_{A,f}(\theta) - s_A^\dagger$, then

$$(Ar)'(A\Sigma_{A,T}A')^{-1}(Ar) = r'\Sigma_{A,T}^{-1} r. \tag{20}$$

The inverse identity proves the equality, so invertible coordinate changes preserve both radii for a fixed support and target. Covariance reweighting, response reduction, and a new normalization define a different experiment. Peak normalization also restricts feasible shock paths in the application; Freyberger (2026) studies the role of normalization in structural counterfactuals.

For local geometry, compare two nearby model responses with $\Sigma_{A,T} = \Omega_{A,f}/T$. At a reference θ_f^0, let $D_f = DS_{A,f}(\theta_f^0)$ and define $H_f = D_f' \Omega_{A,f}^{-1} D_f$.

PROPOSITION 3 (Local response metric). *For $\theta = \theta_f^0 + h/\sqrt{T}$,*

$$\|S_{A,f}(\theta) - S_{A,f}(\theta_f^0)\|^2_{\Sigma_{A,T}^{-1}} = h' H_f h + o(1). \tag{21}$$

Under a smooth locally invertible reparameterization $\theta = \varphi(\eta)$,

$$H_{f,\eta} = D\varphi(\eta_0)' H_{f,\theta} D\varphi(\eta_0), \qquad \ker H_f = \ker D_f. \tag{22}$$

The local zero-distance directions coincide with the empirical-response kernel. Their policy image $C_{f,p} \ker D_f$ connects testing-risk geometry to differential factorization.

4.2. Testing affinity and the radius path

For probability laws P_0 and P_1, define the symmetric total testing risk

$$\tau(P_0, P_1) = \inf_{0 \le \phi \le 1} \{\mathbb{E}_{P_0}\phi + \mathbb{E}_{P_1}(1-\phi)\} = 1 - \|P_0 - P_1\|_{\mathrm{TV}}. \tag{23}$$

PROPOSITION 4 (Gaussian testing-risk calibration). *For two Gaussian experiments with common covariance and Mahalanobis mean distance c,*

$$\tau(c) = 2\Phi(-c/2), \qquad p_{\mathrm{HS}}(c) = \frac{\tau(c)}{2} = \Phi(-c/2), \qquad c = -2\Phi^{-1}\{\tau(c)/2\}. \tag{24}$$

Detection-error calibration follows Hansen and Sargent (2001, 2008, 2010); its connection to misspecification neighborhoods and test power is developed by Bonhomme and Weidner (2022). The quantity p_{HS} is the equal-prior detection-error probability. The chosen benchmark uses total testing risk $\tau_{\mathrm{det}} = 0.05$, hence $p_{\mathrm{HS}} = 0.025$[6] and

$$c_{0.05} = -2\Phi^{-1}(0.025) = 3.9199. \tag{25}$$

This calibration compares two simple Gaussian laws with a common covariance. It fixes a distinguishability scale for the radius path. Composite-null confidence coverage uses a separate projection argument. The application reports the full path from $\tau_{\mathrm{det}} = 0.10$ through 0.02.

The scalar sign-reversal radius is a response-space breakdown point in the sense of Masten and Poirier (2020), with a minimum-deviation interpretation related to Kang and Vasserman (2025). The family-entry radius measures representation of structural families. Confidence coverage concerns the sampling procedure.

For the radius characterization, use a finite collection of nonempty economic domains $\Theta_f^{\mathrm{econ}} = \overline{\Theta}_f$ and a finite target index set $\mathcal{J}$. Radii are finite real tolerances; an empty reversal set has threshold $+\infty$. Define

$$c_{\mathrm{entry}} = \max_f \inf_{\theta \in \Theta_f^{\mathrm{econ}}} d_{f,T}(\theta; s_A^\dagger), \qquad U_j(c) = \sup_{(f,\theta) \in \mathfrak{F}_A(s_A^\dagger; c)} T_{j,f}(\theta), \tag{26}$$

$$\delta_{f,j}^{\mathrm{sign}} = \inf_{\theta \in \overline{\Theta}_f:\, T_{j,f}(\theta) \ge 0} d_{f,T}(\theta; s_A^\dagger), \qquad c_{\mathrm{sign}} = \min_{f,j} \delta_{f,j}^{\mathrm{sign}}, \tag{27}$$

with the extended-real convention $\inf \varnothing = +\infty$.

[6] With equal priors, total testing risk is the sum of the two error probabilities, whereas p_{HS} is their average. This accounts for the factor of two in the calibration.

PROPOSITION 5 (Coverage-and-reversal frontier). *Suppose every $\overline{\Theta}_f$ is compact and the response and target maps are continuous. Then $U_j(c)$ is weakly increasing in c. If $c \geq c_{\text{entry}}$, every coordinate of every policy target in $\mathfrak{F}_A(s_A^\dagger; c)$ is negative exactly when $c < c_{\text{sign}}$. Hence family entry and a common negative sign hold on*

$$[c_{\text{entry}}, c_{\text{sign}}). \tag{28}$$

PROOF OF PROPOSITION 5. The inclusion $\mathfrak{F}_A(s_A^\dagger; c_1) \subseteq \mathfrak{F}_A(s_A^\dagger; c_2)$ for $c_1 \leq c_2$ makes every $U_j(c)$ weakly increasing. Compactness and continuity make each constrained infimum in (27) attainable whenever its feasible set is nonempty. If $c < c_{\text{sign}}$, a compatible point with $T_{j,f}(\theta) \geq 0$ would give $\delta_{f,j}^{\text{sign}} \leq c$, a contradiction. If $c \geq c_{\text{sign}}$, an attaining sign-reversal point belongs to the radius-c fiber. The definition of c_{entry} supplies a nonempty component in every family for $c \geq c_{\text{entry}}$. □

COROLLARY 1 (Persistent interventions). *If Proposition 5 gives $T_{j,f}(h; \theta) < 0$ for delays $h = 0, \ldots, 16$, then $\sum_{h=0}^{16} \rho^h T_{j,f}(h; \theta) < 0$ for every $\rho \in [0, 1]$.*

4.3. Stability across empirical experiments

An empirical response experiment ξ declares the estimator, sample, response coordinates, normalization, covariance construction, and structural-input restrictions used in the distance. Holding the policy target and structural-family domains fixed, write its radius pair as $\{c_{\text{entry}}(\xi), c_{\text{sign}}(\xi)\}$ and define

$$\Delta(\xi) = c_{\text{sign}}(\xi) - c_{\text{entry}}(\xi). \tag{29}$$

For a prespecified collection Ξ of response experiments, define the lower experiment gap

$$\underline{\Delta}(\Xi) = \inf_{\xi \in \Xi} \Delta(\xi). \tag{30}$$

PROPOSITION 6 (Experiment-stable sign separation). *Under the conditions of Proposition 5 for every $\xi \in \Xi$, experiment ξ admits an all-family common-negative-sign interval exactly when $\Delta(\xi) > 0$. Every experiment in Ξ admits such an interval when $\underline{\Delta}(\Xi) > 0$. An experiment $\xi_0 \in \Xi$ with $\Delta(\xi_0) \leq 0$ rules out experiment-stable sign separation over Ξ.*

The comparison keeps the structural target fixed while changing the statistical experiment used to assess response compatibility. Each radius retains its own testing-risk scale through Proposition 4. The application reports the radius pair and gap for the baseline response convention and for three raw-data LP covariance constructions.

The intervals in this proposition may use different radii. For a finite collection Ξ, a single radius that works in every experiment belongs to

$$\bigcap_{\xi\in\Xi}[c_{\text{entry}}(\xi), c_{\text{sign}}(\xi)) = \left[\max_{\xi\in\Xi} c_{\text{entry}}(\xi),\ \min_{\xi\in\Xi} c_{\text{sign}}(\xi)\right), \tag{31}$$

with an empty interval when the lower endpoint reaches the upper endpoint. Thus a positive gap in each experiment and a common numerical radius are distinct questions. The reported collection criterion concerns the former.

4.4. Which evidence moves the frontier?

For a subset J of response coordinates, write $r_J(M) = S_J(M) - s_J^\dagger$ and $d_J(M)^2 = r_J(M)'\Sigma_{JJ}^{-1} r_J(M)$. Holding the covariance and model support fixed, define the feature-specific frontier

$$c_j^*(J;\mathfrak{E}) = \inf_{M\in\mathfrak{E}:\ T_j(M)\geq 0} d_J(M). \tag{32}$$

Thus $c_{\text{sign}} = \min_j c_j^*(A;\mathfrak{E})$ under the same compact economic domains as (27). A superscript $\mathcal{D}$ denotes restriction to an evaluated support.

LEMMA 2 (Evidence increments). *Partition a response block into J and K, and suppose its covariance is positive definite. With $\Sigma_{K|J} = \Sigma_{KK} - \Sigma_{KJ}\Sigma_{JJ}^{-1}\Sigma_{JK}$,*

$$d_{J\cup K}(M)^2 = d_J(M)^2 + \left\| r_K(M) - \Sigma_{KJ}\Sigma_{JJ}^{-1} r_J(M)\right\|_{\Sigma_{K|J}^{-1}}^2. \tag{33}$$

Consequently $\mathfrak{F}_{J\cup K}(s^\dagger; c) \subseteq \mathfrak{F}_J(s^\dagger; c)$ and $c_j^*(J;\mathfrak{E}) \leq c_j^*(J\cup K;\mathfrak{E})$ for a fixed envelope and target. The same conclusions hold on a fixed evaluated support.

PROOF OF LEMMA 2. Define the invertible matrix L and multiply to obtain

$$L = \begin{pmatrix} I & 0 \\ -\Sigma_{KJ}\Sigma_{JJ}^{-1} & I \end{pmatrix}, \qquad L\Sigma L' = \operatorname{diag}(\Sigma_{JJ}, \Sigma_{K|J}).$$

Invertibility gives $r'\Sigma^{-1}r = (Lr)'(L\Sigma L')^{-1}(Lr)$, which is (33). The second quadratic form is nonnegative. Hence each model's distance weakly increases when K is added. The distance sublevel sets are nested, and taking the infimum over the same sign-reversal set proves frontier monotonicity, including the extended-real empty-set convention. □

The identity is the partitioned Gaussian quadratic-form decomposition. Its use here attributes movement in a nonlocal reversal boundary to response evidence, alongside the local moment-informativeness questions of Andrews et al. (2017, 2020). The second term measures the residual contribution of the added observations conditional on the retained ones. An ablation uses Σ_{JJ}, the

covariance of the selected experiment. Re-estimating the parameter support, shock normalization, or covariance defines a further experiment and receives its own report.

Family entry also weakly increases when a block is added on a fixed support. The difference of the two increasing radii can move in either direction. Consider three scalar-target representations with identity covariance and empirical center zero. Family A has responses $(1,0)$ and $(3,0)$ with targets -1 and 1; family B has response $(2,4)$ and target -1. Using only the first coordinate gives entry 2 and reversal 3. Using both gives entry $\sqrt{20}$ and reversal 3. Every individual distance weakly increases, while the common-sign interval becomes empty. The new coordinate challenges family coverage most strongly. This is an exact finite illustration of the definitions, separate from the monetary estimates.

4.5. Continuous numerical certificates

Let $\mathcal{D}_f(c)$ index evaluated joint response–policy pairs inside the radius and define their convex response completion

$$\mathcal{C}_f(c) = \left\{ \sum_{i \in \mathcal{D}_f(c)} \lambda_i \begin{pmatrix} S_{A,f}(\theta_{f,i}) \\ T_{p,f}(\theta_{f,i}) \end{pmatrix} : \lambda_i \geq 0,\ \sum_i \lambda_i = 1 \right\}. \tag{34}$$

PROPOSITION 7 (Convex-completion certificate). *Every empirical-response component of $\mathcal{C}_f(c)$ lies inside the Mahalanobis ball. For target coordinate j,*

$$\inf_{\mathcal{C}_f(c)} T_{p,j} = \min_{i \in \mathcal{D}_f(c)} T_{p,j}(\theta_{f,i}), \qquad \sup_{\mathcal{C}_f(c)} T_{p,j} = \max_{i \in \mathcal{D}_f(c)} T_{p,j}(\theta_{f,i}). \tag{35}$$

Thus a vertex margin $m_j = -\max_i T_{p,j}(\theta_{f,i}) > 0$ certifies $T_{p,j} \leq -m_j$ over this continuous completion.

The convex completion concerns mixtures of evaluated joint response–policy pairs. A native nonlinear parameter-envelope certificate uses a covering or interval argument.

PROPOSITION 8 (Validated covering certificate). *Let $\mathcal{G}_f(c)$ be an ε_f-net of $\mathcal{N}_f(s_A^\dagger; c)$, let $L_{f,j}$ bound the Lipschitz modulus of $T_{j,f}$, and let $e_{f,j}$ bound numerical evaluation error. Then*

$$U_j(c) \leq \max_f \left\{ \max_{\theta \in \mathcal{G}_f(c)} \widehat{T}_{j,f}(\theta) + L_{f,j}\varepsilon_f + e_{f,j} \right\}. \tag{36}$$

A negative right-hand side certifies the sign on the native nonlinear envelope.

4.6. Expansion and approximation

PROPOSITION 9 (Envelope expansion). *If $\mathfrak{E}_1 \subseteq \mathfrak{E}_2$, then for every s and p,*

$$\mathcal{I}_p(s; \mathfrak{E}_1) \subseteq \mathcal{I}_p(s; \mathfrak{E}_2). \tag{37}$$

For a declared extension envelope $\mathfrak{G} \supseteq \mathfrak{E}$, a nonempty baseline policy set, and symmetric positive definite Σ_A, define

$$\rho_p(\Delta; \mathfrak{G}) = \inf_{M \in \mathfrak{G}:\ \operatorname{dist}\{T_p(M), \mathcal{J}_p(s_A^\dagger; \mathfrak{E})\} \geq \Delta} \|S_A(M) - s_A^\dagger\|_{\Sigma_A^{-1}}. \tag{38}$$

The curve gives the empirical distance to an admitted extension that moves the policy set by at least Δ.

Let D collect empirical-response secants and let C_p collect policy-response secants. A reduced-rank map fitted to $(\mathsf{D}, \mathsf{C}_p)$ represents the evaluated response relation. Its nonlinear approximation error at radius c is

$$\epsilon_{\mathrm{sec}}(c, p) = \max_f \sup_{\theta \in \mathcal{N}_f(s_A^\dagger; c)} \|T_{p,f}(\theta) - \lambda_p\{S_{A,f}(\theta)\}\|_{W_p}. \tag{39}$$

The numerical allowance $\epsilon_{\mathrm{num}}(c, p)$ combines solver, discretization, optimization, and simulation errors. Sampling uncertainty enters through the projected confidence set.

5. Inference with generated model responses

The response maps used in the identification analysis are estimated with uncertainty from empirical moments, model parameters, the intervention path, common external inputs, and simulation. A joint first-order representation carries these sources and their cross-family covariance into a projection procedure. A target-constrained quotient uses the geometry of the tested policy level under the stated tangent and covariance conditions.

The statistical references include Stock and Wright (2000) on weakly identified GMM and Kleibergen (2002) on pivotal tests. Andrews and Guggenberger (2019) studies singular moment covariance, AR and conditional quasi-likelihood-ratio tests, and subvector inference. Projection inference here uses a pointwise test over a nuisance union. The quotient results below are regular profiling refinements under explicit tangent conditions, interpreted alongside Andrews and Cheng (2012); Andrews and Mikusheva (2016). Applying the refinement requires estimated covariance, boundary distances, and optimization diagnostics.

5.1. Inputs to the projection procedure

The response center, calibrated parameters, intervention path, and simulation draws can share sampling variation. Proposition A1 gives the joint first-stage conditions and generated-response expansion. The covariance includes the resulting cross-family and common-input blocks. Cocci and Plagborg-Møller (2025) develop a different information benchmark in which the variances of calibration moments are available and their correlations are unrestricted. The procedure here uses the specified joint covariance. The projection result states its own uniform approximation and nondegeneracy requirements.

5.2. Projection over structural nuisances

Let the nuisance index be $\zeta = (f, \theta)$ in the disjoint union

$$\mathcal{H}(c) = \bigsqcup_{f} \{f\} \times \mathcal{N}_f(c). \tag{40}$$

The symbol ζ indexes the model nuisance, and η denotes the simulation ratio in Assumption A1. Define population moments

$$g_A(\zeta) = s_A^\dagger - S_A(\zeta), \quad g_\tau(\zeta, \tau) = T_p(\zeta) - \tau, \quad g_B(\zeta) = s_B^\dagger - S_B(\zeta). \tag{41}$$

Let g_S select a fixed-dimensional stack of these moments, with dimension d_S. For a null value b, set

$$Q_{S,T}(\zeta, b) = T\hat{g}_S(\zeta, b)'\hat{\Omega}_S(\zeta)^{-1}\hat{g}_S(\zeta, b), \qquad P_{S,T}(b) = \inf_{\zeta \in \mathcal{H}(c)} Q_{S,T}(\zeta, b). \tag{42}$$

ASSUMPTION 2 (Uniform structural-moment approximation). *The joint influence representation and studentized central limit theorem hold uniformly jointly over data-generating laws P in the maintained null class and $\zeta \in \mathcal{H}(c)$. Each of the finitely many family components of $\mathcal{H}(c)$ is compact and the profiled criterion is measurable. The selected covariance matrices are symmetric positive definite with eigenvalues in a fixed compact subset of $(0, \infty)$, their estimators converge uniformly, and d_S is fixed.*

THEOREM 4 (Uniform projection inference). *Consider a composite null class such that, for every P, there is a compatible $\zeta_P^0 \in \mathcal{H}(c)$ with $g_{S,P}(\zeta_P^0, b_P) = 0$. Under Assumption 2,*

$$\limsup_{T \to \infty} \sup_{P} \mathbb{P}_P\{P_{S,T}(b_P) > \chi^2_{d_S, 1-\alpha}\} \le \alpha. \tag{43}$$

The conclusion covers singular empirical-response Jacobians and sequences along which their positive singular values converge to zero.

PROOF OF THEOREM 4. For each null law P, choose a compatible nuisance ζ_P^0 with $g_{S,P}(\zeta_P^0, b_P) = 0$. By definition of the infimum,

$$P_{S,T}(b_P) \le Q_{S,T}(\zeta_P^0, b_P).$$

Assumption 2 implies

$$\hat{\Omega}_S(\zeta_P^0)^{-1/2}\sqrt{T}\hat{g}_S(\zeta_P^0, b_P) \Rightarrow N(0, I_{d_S})$$

uniformly over the null class. Consequently the upper tail of $Q_{S,T}(\zeta_P^0, b_P)$ at $\chi^2_{d_S, 1-\alpha}$ converges uniformly to at most α. Since

$$\{P_{S,T}(b_P) > c\} \subseteq \{Q_{S,T}(\zeta_P^0, b_P) > c\},$$

taking the supremum over P and the limit superior proves (43). □

The theorem requires a compatible true nuisance in the profiled union for every maintained null law. A screen or a selected finite support must preserve that inclusion before the same coverage argument is applied. In this subsection, the population centers in (41) are the law-indexed limits of the estimators; hats denote their sampling versions. The moment dimension d_S remains fixed.

The singularity allowed by the theorem concerns the response Jacobian. The studentizing moment covariance retains the positive eigenvalue bounds in Assumption 2. The generated-response expansion uses its own full-rank first-stage minimum-distance derivative. Each condition belongs to the object it controls.

A numerical profile also needs a bound in the correct direction. Suppose a feasible search value $\overline{P}_{S,T}$ has a certified global optimization allowance ϵ_{opt}:

$$\overline{P}_{S,T} - \epsilon_{\text{opt}} \le P_{S,T} \le \overline{P}_{S,T}. \tag{44}$$

Then $\overline{P}_{S,T} - \epsilon_{\text{opt}} > \chi^2_{d_S,1-\alpha}$ certifies rejection. Searching a finite subset alone gives an upper bound on the unrestricted infimum. The monetary calculation certifies its inner shock-path fits and reports the evaluated outer support separately.

Regular refinements. Interior, constant-rank cases admit quotient and target-level refinements under the stated covariance factorization. The projection procedure remains the reference for the broader nuisance union. Each reported calculation identifies which calibration is used.

5.3. A Gaussian inference diagnostic

The Gaussian experiment illustrates the cost of the rank-uniform reference and the additional information used by a regular refinement. The numerical rows reproduce the supplied simulation output.

The weak-rank Gaussian experiment contains 648 cells and 3.24 million replications. Table 1 reports the 324 cells with $c \le 1$ and positive target loading, together with the 27 fixed-positive-singular-value cells.

TABLE 1. Weak-rank inference: size, power, and confidence-set geometry

Procedure and design	Rejection	Power	Unbounded	Empty	Infinite median
Projection, weak sequence	0.146	0.284	74.318	0.053	243
Target-constrained oracle, weak sequence	4.996	6.185	67.306	2.268	243
Quotient, fixed positive singular value	5.031	75.024	0.000	2.042	0

The weak rows contain 324 cells; the regular row contains 27. Rejection ranges in row order are [0.020, 0.340], [4.220, 5.800], and [4.380, 5.740] percent. Rejection, power against a 0.25 target shift, unbounded-set frequency, and empty-set frequency are percentages. The target-constrained row uses known covariance and analytic profiling.

In the weak cells, the profiled null statistic has a χ^2_3 law. Least-favorable projection uses the $\chi^2_{8,.95}$ critical value 15.507. Its analytic rejection probability is 0.1431 percent, close to the simulated mean 0.1463 percent. Mean power against a 0.25 shift is 0.2844 percent, mean unbounded-set frequency is 74.32 percent, and 243 of 324 cells have infinite median length. These quantities measure the inferential cost of the rank-uniform reference. Dufour (1997) establishes that valid confidence sets for unbounded weakly identified targets can require positive probability of unboundedness. The particular excess size conservatism here comes from using eight reference degrees of freedom for a three-dimensional profiled Gaussian statistic. The reported calculations measure both effects through rejection, power, and set length.

For positive target loading, the target-null level set removes the weak direction and the augmented null Jacobian has singular value $\sqrt{\gamma^2 + c^2/T} \geq \gamma$. The target-constrained quotient oracle has mean rejection 4.996 percent and mean power 6.185 percent in the weak cells; its mean unbounded-set frequency remains 67.31 percent. In the regular supplement, rejection ranges from 4.38 to 5.74 percent, mean power is 75.02 percent, and every reported set is bounded. Application of the target-constrained refinement uses the diagnostics following Proposition A3.

The target-constrained row is an oracle comparison with known covariance and analytic profiling. Its rejection frequency assesses that Gaussian design. Applying it to the monetary frontier would require the tangent, covariance, and optimization conditions of the refinement as well as a sampling argument for the selected nuisance domain.

6. Monetary-policy design

The application asks whether causal responses to current monetary shocks identify the output and inflation effects of delayed or persistent interventions. The empirical response and four-family model evaluations come from Caravello et al. (2026). Their extrapolations are placed in a cross-family policy image indexed by testing risk. The analysis then develops inference for the generated response stack and states numerical certification conditions. Response blocks, extension families, norms, and evidence specifications are set before the corresponding evaluation outcomes are used.

6.1. Training evidence and the common protocol

The baseline quarterly sample is 1969Q1–2006Q4. The reported reconstruction uses the public series and shock archives described below, with a 2026-08-25 FRED vintage for the current-vintage analysis.[7] The aggregate variables are log real output per capita (FRED `A939RX0Q048SBEA`), transformed following Hamilton (2018); annualized log-difference inflation in the GDP deflator (FRED `GDPDEF`); and the federal funds rate (FRED `FEDFUNDS`). The two treatment series are the Romer and Romer (2004) monetary-policy shock, using the Ramey replication series, and the natural-language shock of Aruoba and Drechsel (2024). Monthly shocks are aggregated to quarters, with missing shock observations set to zero in the baseline design.

[7] The dated raw-data reconstruction and the current-vintage comparison are stored as separate response protocols.

The VAR order is AD shock, output, inflation, RR shock, and the policy rate, with two lags, an intercept, and a linear trend. Recursive identification fixes the impact responses of output and inflation to the RR innovation at zero. The covariance assigns variance 10^6 to these two coordinates following the author implementation, and the horizon-8 triangular taper applies before the paper-specific selection and scaling.

Within the baseline 1969Q1–2006Q4 VAR design, let $q \in \{RR, AD\}$ denote the two treatments and let $a_q = 0.25/\max_{0\le j\le 24} |IRF_i^q(j)|$. Every response for treatment q is multiplied by a_q, so its annualized policy-rate path has a 25-basis-point peak over the first 25 quarters. Output is measured in percent deviations; inflation and the policy-rate input are divided by four at the model interface. The training response stacks output and quarterly inflation over 13 quarters:

$$s_A^\dagger = \begin{pmatrix} IRF_y^{RR}(0{:}12) \\ IRF_\pi^{RR}(0{:}12) \\ IRF_y^{AD}(0{:}12) \\ IRF_\pi^{AD}(0{:}12) \end{pmatrix} \in \mathbb{R}^{52}. \tag{45}$$

The 52 outcome coordinates and 26 policy-rate coordinates form a 78-dimensional joint input with covariance obtained from the same transformed VAR draws. The response reconstruction implements VAR and lag-augmented LP, following the population-response connection in Plagborg-Møller and Wolf (2021) and the lag-augmentation design of Montiel Olea and Plagborg-Møller (2021). It refits Hamilton preprocessing and peak normalization on the dated raw-data calendar. Alternative shock constructions and post-2006 samples have separate response and covariance requirements.

The population response and its finite-sample implementation are distinct. Local projections and vector autoregressions can target the same impulse response under their unrestricted-lag equivalence. Here lag length, retained horizons, shock normalization, covariance estimation, and profiling restrictions are specified separately. The comparison attributes differences to those declared choices.

6.2. The structural envelope

The filtering ablation holds the 4,000 baseline evaluations, fitted two-shock input paths, normalization, and policy responses fixed. Twenty-six evidence specifications select shock blocks, outcome blocks, individual shock–outcome blocks, their complements, and joint horizons zero through k for every $k = 0, \ldots, 12$. Each specification uses its principal submatrix of the baseline covariance and the same testing-risk radius. The empty-filter row describes the original estimated support. The experiment measures changes in compatibility within this support; parameter and shock-input re-estimation define a separate sensitivity exercise.

The baseline evaluated support consists of 1,000 retained evaluations per family.[8] Its coordinate hull supplies a bounded search region, with fixed coordinates held at their baseline values. An economic admissibility domain additionally imposes the family's probability, equilibrium, determinacy, borrowing, and calibration restrictions. The reported parameter searches use the coordinate hull and the finite-horizon solver selection and are interpreted on that computational domain.

The four baseline families are summarized in Table 2. Their common interface returns S_A, the candidate held-out blocks S_B, and policy responses T_p under the protocol in Assumption 1.

TABLE 2. Baseline model families

Family	Defining margin	Family-specific calibration content
RANK–RE	Representative agent; rational expectations	Aggregate monetary responses and standard nominal/real rigidity moments
HANK–RE	Household risk and portfolio heterogeneity; rational expectations	Common aggregate moments plus liquid wealth, debt exposure, MPC, and idiosyncratic-risk moments
RANK–CD	Representative agent; cognitive discounting in price and wage setting	RANK moments plus expectation-response moments when included in the evidence design
HANK–CD	Household heterogeneity and cognitive discounting	HANK distributional moments plus expectation-response moments

The RANK and HANK benchmarks follow the monetary-transmission specifications of Christiano et al. (2005) and Kaplan et al. (2018). HANK computation uses the sequence-space methods of Auclert et al. (2021); cognitive discounting follows Gabaix (2020). Parameter domains and equilibrium selections are described in Online Appendix D.

The family comparison varies two mechanisms. Household risk and portfolio positions change how a policy wedge reaches consumption and saving. Cognitive discounting changes how agents respond to future conditions in price and wage setting. Their effects are evaluated through the same declared response maps; the comparison allows each family its own state variables and equilibrium interest-rate path.

The baseline cognitive-discounting calibration is $m = 0.65$. The value $m = 1$ corresponds to rational expectations in this implementation. Under weak separation in the calibration experiment, m enters the reported sensitivity coordinates. Each family defines Θ_f^{econ} from its equilibrium and calibration restrictions. The response-detectability distance centers $S_{A,f}(\theta)$ at the estimated 52-vector $s_A^\dagger$ and uses the baseline finite-sample covariance in (16).

The benchmark sets symmetric total testing risk to five percent. Following (25), the corresponding Hansen–Sargent detection-error probability is 0.025 and the radius is 3.9199. The reported path

[8] The four-family frontier therefore starts from 4,000 joint response–target evaluations before any extension point is added.

is

$$\tau_{\text{det}} \in \{0.10, 0.08, 0.07, 0.06, 0.05, 0.04, 0.03, 0.02\}, \tag{46}$$

$$p_{\text{HS}} = \tau_{\text{det}}/2, \qquad c = -2\Phi^{-1}(p_{\text{HS}}). \tag{47}$$

The radius path screens 1,000 retained baseline evaluations per family under a fixed covariance and response design. Separate coordinate and damped Gauss–Newton searches supply nonlinear coverage witnesses. The native-envelope certificate in Proposition 8 additionally requires closed parameter domains, equilibrium and conditioning gates, target-specific interval bounds, and zero unresolved boxes. The local-box calculation leaves this gate open. The reported frontier comparison uses the evaluated support and its separately defined convex completion. A family holdout fits the response relation on the complement of that family's evaluated points.

The extension exercise associated with (38) adjoins one author-calibration point from the financial model of Christiano et al. (2014) to the 4,000 baseline evaluations. The CMR parameter vector and economic equations retain their author values; 24 news states implement the common dated-wedge protocol. Two 25-quarter shock sequences are fitted to the normalized 150-coordinate empirical response before computing the 52-coordinate training distance. The calculation reported here treats this CMR calibration as an evaluated expansion point.

6.3. Policy targets and disagreement geometry

Policy $p = h$ is a one-quarter additive monetary-policy wedge at horizon h, announced at date zero and measured as 25 annualized basis points:

$$u_t^{(h)} = 0.25\,\mathbb{1}\{t = h\}, \qquad h = 0, 1, \ldots, 16, \tag{48}$$

where the solver receives the quarterly counterpart. Each family's maintained feedback rule remains in force, so the equilibrium interest-rate response is an outcome of the intervention. The common object is the announced wedge sequence. A prescribed equilibrium rate path and a change in feedback-rule coefficients define further policy objects. The corresponding model response is

$$v_f(h;\theta) = \begin{pmatrix} IRF_{y,f}^{(h)}(0{:}12;\theta) \\ IRF_{\pi,f}^{(h)}(0{:}12;\theta) \end{pmatrix} \in \mathbb{R}^{26}. \tag{49}$$

The scalar targets are cumulative discounted output and inflation responses,

$$\tau_Y(h;\theta) = \sum_{t=0}^{12} \beta^t IRF_y^{(h)}(t;\theta), \qquad \tau_\pi(h;\theta) = \sum_{t=0}^{12} \beta^t IRF_\pi^{(h)}(t;\theta), \tag{50}$$

with baseline $\beta = 0.99$ and sensitivity values 0.95 and 1.

For family f with n_f centered nonlinear design evaluations, let $\mathsf{D}_f \in \mathbb{R}^{52\times n_f}$ collect training-response secants and $\mathsf{V}_f(h) \in \mathbb{R}^{26\times n_f}$ collect policy-response secants. Let $\widehat{\Lambda}_{-f,r}(h) \in \mathbb{R}^{26\times 52}$ be a rank-r map fitted on families in $\{1,\dots,F\}\setminus\{f\}$, and let $W_\nu \in \mathbb{R}^{26\times 26}$ be symmetric positive definite. The family-held-out loss is

$$E_f(r,h) = \left\| W_\nu^{1/2}\{\mathsf{V}_f(h) - \widehat{\Lambda}_{-f,r}(h)\mathsf{D}_f\}\right\|_F^2. \tag{51}$$

Families receive equal total weight in $CV(r,h) = \sum_f \omega_f E_f(r,h)$. The selected rank is the smallest rank within one family-clustered standard error of the minimum. Reports label it the finite-design rank and place it beside the population-Jacobian rank and the nonlinear residual in (39).

At every family center and policy horizon, the application computes $D_{A,f} = DS_{A,f}(\theta_f^0)$, $C_{Y,f}(h) = D\tau_{Y,f}(h;\theta_f^0)$, and $C_{\pi,f}(h) = D\tau_{\pi,f}(h;\theta_f^0)$. The output includes $C_{p,f}(h)\ker D_{A,f}$ and $\kappa_{f,p}(h)$ from (15). Cross-family overlap is examined at responses lying in, or statistically compatible with, two family images.

6.4. Additional response evidence

The policy-information block uses the public replication shocks of Jarociński and Karadi (2020) with one- and two-year Treasury-yield responses. The target/path decompositions of Gürkaynak et al. (2005) and Swanson (2021), and the observation protocol of Bauer and Swanson (2023), enter source-availability sensitivities. Survey expectations follow the response design in Del Negro et al. (2023). Together these form the expectations and policy-information block B_E.

The financial block B_F contains commercial-paper funding spreads and Census QFR investment and debt responses by asset-size and industry cells, motivated by Gertler and Karadi (2015) and Ottonello and Winberry (2020). The QFR analysis has an aggregate-cell estimand. The household block B_H uses SCF exposures motivated by Auclert (2019) and CEX consumption responses by homeownership, mortgage-payment status, and liquid-resource group motivated by Cloyne et al. (2020). These blocks provide additional response coordinates for transmission checks and family-held-out prediction.

The evaluation design reserves long horizons, alternating output and inflation coordinates, external policy-information responses, expectation and balance-sheet responses, and one structural family at a time. The external blocks are reported as response estimates with their uncertainty, while family-held-out exercises evaluate predictive coverage and generated-response loss.

6.5. Estimation, numerical certification, and policy exercises

Each family is estimated by common minimum distance,

$$\hat{\theta}_f = \operatorname*{argmin}_{\theta\in\Theta_f^{\mathrm{econ}}} \{\hat{\mu}_f - \mu_f(\theta)\}' \widehat{\Omega}_{\mu,f}^{-1}\{\hat{\mu}_f - \mu_f(\theta)\}. \tag{52}$$

Shared samples, treatment paths, and external moments are stacked before covariance estimation. Numerical accuracy is assessed through equilibrium residuals, horizon refinement, solver-tolerance comparisons, and simulation variation. Independent horizon-200 and horizon-300 solves are compared for the 2,457 compatible vertices.

Persistent paths are constructed from the same 0.25-percentage-point basis:

$$u_t(\rho) = 0.25\rho^t, \quad t = 0, \ldots, 16, \qquad \rho \in \{0, 0.50, 0.75, 0.90\}, \tag{53}$$

and, in the linear response system,

$$v(\rho) = \sum_{h=0}^{16} \rho^h v(h). \tag{54}$$

The coefficient ρ^h is dimensionless because $v(h)$ already corresponds to a wedge whose annualized unit is 0.25 percentage point. The intervention equals zero after quarter sixteen. Equation (54) therefore describes a truncated persistent intervention; the equilibrium response after quarter sixteen follows the maintained model and policy rule.

The Monte Carlo varies sample size, the smallest singular value of $D_{A,f}$, the policy loading along its right singular vector, and the covariance between generated responses and derivatives. It compares plug-in Wald, rank-threshold Wald, projection inference, the regular quotient, and a target-constrained quotient oracle. Reported cells use 5,000 replications and record rejection frequency at a five-percent level, 95 percent confidence-set coverage, median set length, power, unbounded-set frequency, and empty-set frequency.

Policy horizon. The delayed interventions are announced at time zero, and their targets sum responses through quarter twelve. A wedge arriving after quarter twelve therefore enters this target through anticipation during the evaluation window. Persistent paths combine the declared delays zero through sixteen. These conventions define the policy comparison; welfare rankings would also require a welfare criterion.

7. Counterfactual frontiers across causal-response experiments

The results answer three questions in order: when the response neighborhood includes every family, whether it then admits a policy-sign reversal, and how that comparison changes with the empirical protocol. Distances are evaluated on the domain named in each exhibit. Resampling ranges describe changes in the estimated experiment on that domain.

Frontiers move across empirical experiments as response restrictions and structural fitting change. Each result is reported with its sampling scheme and structural domain. Structural sign separation concerns the policy image of a declared response-compatible set. Sampling sign exclusion concerns a confidence set that propagates the stated estimation uncertainty.

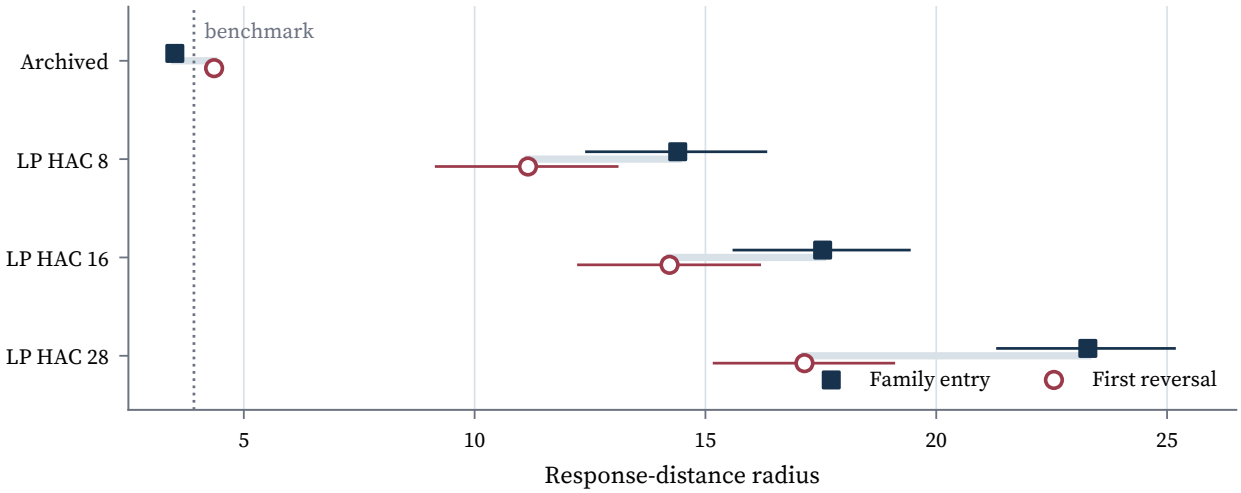
7.1. The experiment-indexed result

The baseline response experiment yields family-entry radius 3.5058 and first evaluated reversal 4.3567, so its evaluated-support gap is 0.8509. Refitting parameters and shock paths to the fifty-coordinate LP experiment yields negative gaps under all three HAC bandwidths: −3.2435, −3.3159, and −6.1391. Figure 1 displays the change in ordering. The four-experiment collection has $\inf_{\xi \in \Xi} \Delta(\xi) < 0$ on the declared evaluated supports.

On a common set of 102 parameter points, the original shock-fitting protocol gives a HAC-28 gap of 6.2580 and LP shock profiling gives −6.1391. Holding the LP-profiled fitted responses fixed, the baseline active covariance gives gap −1.0918 and the LP HAC-28 covariance gives −6.1391. The first LP crossings are positive by at most 0.000284 for output and 0.000814 for inflation. Tables 6 and 10 report the decomposition and target-size frontiers.

Sampling statements depend on how the response experiment is resampled. Fixed-support Gaussian directional ranges lie below zero for each LP bandwidth. Calendar block-multiplier ranges include zero after preprocessing, LP estimation, impact identification, and normalization are rerun. Tables 7 and 8 report these two calculations. Each result retains its stated structural support, policy targets, covariance treatment, and profiling regime.

FIGURE 1. Family entry and first sign reversal across empirical experiments



Squares mark family entry and circles the first evaluated nonnegative target. A circle to the right of its square gives a positive gap. The row labeled "Archived" is the baseline protocol. Horizontal intervals on the LP entries report the displayed directional sampling range. The dotted line is the benchmark radius 3.9199. Each row retains its stated support and covariance construction.

7.2. Testing-risk coverage and radius path

The benchmark total testing risk is 0.05, its Hansen–Sargent detection-error probability is 0.025, and its radius is 3.9199. Table 3 reports the retained baseline evaluations and the best recorded nonlinear coverage witness from the coordinate and damped Gauss–Newton searches.

TABLE 3. Response coverage at total testing risk 0.05

Family	Retained minimum	Retained count	Nonlinear witness	Witness max target
RANK–RE	3.4880	613	3.4280	−0.005974
HANK–RE	3.5058	561	3.4245	−0.007058
RANK–CD	3.4838	664	3.3857	−0.007743
HANK–CD	3.4623	619	3.4089	−0.009090

The radius is 3.9199. Each family supplies 1,000 retained baseline evaluations. The nonlinear-witness column reports the best distance recorded by coordinate search and damped Gauss–Newton refinement; its projected-gradient norms are 4.299, 6.555, 1.675, and 4.700 in row order. Each witness is a feasible coverage point and its 34 delayed-policy targets lie below zero.

Each family has a nonlinear coverage witness with distance from 3.3857 to 3.4280. The projected-gradient norms range from 1.675 to 6.555, so these distances serve as feasible witness values. The 4,000 retained points are 1,000 retained parameter draws per family. Their counts summarize the evaluated design; sampler density governs these counts. At the benchmark, 613 RANK–RE, 561 HANK–RE, 664 RANK–CD, and 619 HANK–CD points enter the response ball.

Table 4 traces the same baseline design over eight testing-risk levels. The radius associated with total risk 0.10 admits zero retained points. All four families have retained coverage by total risk 0.07. The largest policy upper endpoint remains below zero through total risk 0.03 and becomes positive by total risk 0.02.

TABLE 4. Baseline-design policy image over the testing-risk path

Total risk	p_{HS}	Radius	Family counts	Total	Largest policy upper
0.10	0.050	3.2897	0/0/0/0	0	Coverage empty
0.08	0.040	3.5014	1/0/3/8	12	−0.005861
0.07	0.035	3.6238	55/59/64/95	273	−0.004444
0.06	0.030	3.7616	279/289/305/331	1,204	−0.002351
0.05	0.025	3.9199	613/561/664/619	2,457	−0.002193
0.04	0.020	4.1075	872/806/901/820	3,399	−0.000963
0.03	0.015	4.3402	958/947/987/948	3,840	−0.000011
0.02	0.010	4.6527	998/995/1,000/995	3,988	0.002816

Family counts are RANK–RE/HANK–RE/RANK–CD/HANK–CD. The last column is the largest upper endpoint across 34 delayed and eight persistent target cells. Counts depend on the retained-chain design.

The first evaluated sign reversal enters at distance 4.3567, equivalent to total testing risk 0.02938 and Hansen–Sargent error 0.01469. It is HANK–RE draw 443: inflation at delay three equals 0.000524, and the $\rho = 0.90$ inflation response equals 0.002816. The benchmark therefore has a baseline-design sign-radius slack of 0.4367.

7.3. Structural refitting to the LP experiment

Let $G_f(\theta)q$ collect the 150 model responses, where q stacks twenty-five coefficients for each shock. For response coordinates A, the refitting criterion is

$$\widetilde{d}_{f,A,b}(\theta) = \min_{q \in \mathcal{Q}_f(\theta)} \left\| \widehat{\Sigma}_{A,b}^{-1/2} \left([G_f(\theta)q]_A - \hat{s}_A^{LP} \right) \right\|_2 . \tag{55}$$

Both shock-path profiling and parameter search minimize this same criterion. The set $\mathcal{Q}_f(\theta)$ fixes the AD rate response at quarter two and the RR rate response at quarter one to 0.0625 quarterly percentage points, fixes RR output and inflation impact responses to zero, and bounds the absolute rate response at every quarter zero–twenty-four by 0.0625. These restrictions preserve the declared recursive zeros and peak-normalization cell. The earlier mixed-objective search profiled inputs on 146 coordinates and searched parameters using a 50-coordinate subdistance, with two peak equalities. Equation (55) records the common criterion and the additional normalization restrictions used here.

The parameter search uses the retained-evaluation coordinate hulls and fixes inactive parameters to the first retained evaluation. These are computational domains defined by the retained draws. The economic-domain and household-grid declarations retain their separate certification requirements. The principal LP experiment uses the fifty stochastic output and inflation coordinates through quarter twelve. The 146-coordinate experiment also fits the rate path and responses through quarter twenty-four, omitting four deterministic coordinates. Each experiment profiles its own shock paths. Policy targets retain the original announced 25-basis-point wedge definition.

Fifty-two parameter-search runs generate 102 distinct parameter evaluations across the four families. Reprofiling every point under both coordinate choices and all three bandwidths gives 612 constrained fits on a common parameter support. Every fit satisfies the numerical inner-solver and normalization checks. The largest relative convex-profile gap is 3.74×10^{-10}, the largest normalization equality error is 3.56×10^{-15}, and the largest rate-bound violation is 2.64×10^{-15}. The parameter searches supply evaluated candidates; their global optimality remains a separate question.

TABLE 5. LP-refitted distances on a common evaluated parameter support

Family	Entry, 50	Reversal, 50	Entry, 146	Reversal, 146	Bounds
RANK–RE	23.2840	32.2795	274.28	352.71	30
HANK–RE	18.2001	25.8588	323.16	372.30	32
RANK–CD	22.3531	27.1542	240.11	290.05	39
HANK–CD	17.0400	17.1449	316.26	559.40	29
Family union	23.2840	17.1449	323.16	290.05	–

The covariance uses Bartlett bandwidth 28. Every row profiles both 25-quarter shock paths with the same objective used for parameter search. The 102 distinct parameter points contain 25 RANK–RE, 28 HANK–RE, 25 RANK–CD and 24 HANK–CD points. Each point is reprofiled under both experiments; the 50-coordinate experiment uses output and inflation at horizons zero–twelve, and the 146-coordinate experiment adds the other stochastic primitive responses. Rate peaks and recursive zeros are imposed as in (55). Bounds counts active normalized-rate bounds among 50 ordinates at the 50-coordinate entry point, including the two peak anchors. Union entry is the largest family minimum; union reversal is the smallest nonnegative-target distance. These are evaluated-support quantities on the declared parameter support.

For the fifty-coordinate HAC-28 experiment, family entry occurs at 23.2840 and the first evaluated reversal at 17.1449. Thus reversal precedes all-family entry, and the common-negative-sign interval is empty on this support. At radius 3.9199, the evaluated compatible set is empty. HAC bandwidths 8 and 16 give entry–reversal pairs (14.3999, 11.1563) and (17.5390, 14.2231). The same ordering appears in each row. These are refitted finite-support findings on the declared parameter support.

The first HAC-28 LP reversal is HANK–CD output at delay four. The stored target is 0.00001217, with reported consumption and investment components −0.00709636 and 0.00710854. These independently rounded components sum to 0.00001218. At a 250-quarter solve it remains positive at 0.00001102. Its baseline-metric distance with the original shock-fitting protocol is 4.9899. This comparison attaches the same structural counterfactual to two specified response experiments.

Table 6 records an ordered accounting path for the baseline–LP change on declared supports. Panel A holds the 102 LP-profiled fitted response vectors and policy targets fixed. Replacing baseline response dynamics moves the gap from −0.0683 to −0.9548; applying the LP peak normalization moves it to −1.0918; applying the LP HAC-28 covariance moves it to −6.1391. Panel B separates parameter support from shock fitting under the LP metric. The 4,000 baseline points and fitted inputs give gap 3.3811. Evaluating the original shock-fitting protocol on the 102 LP-search parameter points gives 6.2580. Applying the LP shock-profile criterion on those same points gives −6.1391. The fitted shock path and covariance geometry account for the sign-ordering change along this path. The intermediate contributions depend on the order of replacement and on the support held fixed.

TABLE 6. Decomposition of the baseline–LP frontier change

Panel A. Center, normalization, and covariance on fixed LP-profiled responses

Response experiment	Entry	Reversal	Gap
Baseline center and covariance	5.5149	5.4466	-0.0683
LP dynamics, baseline normalization	11.2252	10.2704	-0.9548
LP center, baseline covariance	11.5117	10.4199	-1.0918
LP center and HAC 28 covariance	23.2840	17.1449	-6.1391

Panel B. Parameter support and shock fitting under the LP HAC-28 experiment

Structural calculation	Points	Entry	Reversal	Gap
Baseline 4,000 points and fitted inputs	4,000	46.4344	49.8155	3.3811
LP-search points, original shock fitting	102	41.0373	47.2953	6.2580
LP-search points, LP shock profiling	102	23.2840	17.1449	-6.1391

Panel A uses the same 102 LP-profiled fitted response vectors and policy targets. The first row uses the baseline center and active 50×50 covariance. The second row replaces the response dynamics while retaining the baseline peak-normalization multipliers; the third applies the LP peak normalization; the fourth applies the LP HAC-28 covariance. Panel B starts from the 4,000 baseline evaluations and fitted inputs, then evaluates the original full-response shock-fitting protocol on the 102 LP-search parameter points, and finally applies the LP shock-profile criterion on those points. Each row is an evaluated-support calculation.

On 146 coordinates, the HAC-28 entry radius is 323.1614 and the reversal radius is 290.0486. Its covariance condition number is 3.24×10^9, compared with 1.62×10^5 for fifty coordinates. These large discrepancies accompany the additional rate and long-horizon coordinates and their estimated covariance geometry. Their interpretation is a profiled response-distance diagnostic; sampling rejection probabilities require the generated-response distribution and constrained inversion. Bandwidth paths, active rate bounds, and horizon checks accompany the calculation.

7.4. Sampling variation in the LP gap

Table 7 reports family entry, first reversal, and their gap for the baseline experiment and the three fifty-coordinate LP experiments. Proposition 6 maps these experiment-specific rows into the collection criterion.

TABLE 7. Experiment-indexed family-entry and sign-reversal gaps

Experiment	Entry	Reversal	Gap	95% gap range	Pr(gap > 0) (%)
Baseline covariance	3.5058	4.3567	0.8509	–	–
LP, HAC 8	14.3999	11.1563	-3.2435	[-5.1655, -1.2974]	0.033
LP, HAC 16	17.5390	14.2231	-3.3159	[-5.2075, -1.3484]	0.003
LP, HAC 28	23.2840	17.1449	-6.1391	[-7.8345, -4.2955]	0.000

The gap is $c_{\text{sign}} - c_{\text{entry}}$. The baseline row is a point comparison under its baseline covariance. LP rows use 100,000 Gaussian perturbations of the declared raw-data response estimator and report descriptive perturbation ranges. Each profiled distance is linearized by its envelope derivative, and the family minima, all-family maximum, and reversal minimum are recomputed in every draw across the 102-point support. The calculation permits binding evaluations to change. Its scope conditions on the evaluated structural support, policy targets, covariance estimator, and active profiling regime.

For the LP rows, the empirical center receives Gaussian perturbations from the raw-data covariance. The envelope derivative of each profiled distance supplies its local response to the perturbation. Family minima, the all-family maximum, and the reversal minimum are recomputed in every draw across all 102 evaluated points. Binding evaluations change across draws: under HAC 28, the two most frequent entry evaluations account for 71.55 percent of draws, and three HANK–CD reversal evaluations account for the reversal minimum. The directional calculation retains these switches. The 95-percent gap ranges lie below zero for all three bandwidths. These ranges condition on the evaluated structural support, fixed policy targets, covariance estimator, and active profiling regime; parameter-domain optimization and full generated-response inference remain separate objects.

The selected calendar bootstrap uses 28-quarter circular blocks and adaptive peak normalization. Each of its 1,999 draws reruns Hamilton detrending, lag-augmented LP estimation, impact identification, response normalization, and all 102 constrained shock-path profiles, then recomputes the family-entry and reversal binders. Table 8 reports the resulting distribution. The point gap is −6.1391, the bootstrap median is −4.9865, and the central 95-percent resampling range is $[-10.2187, 2.2214]$. The gap is positive in 6.353 percent of draws. The peak quarter changes in 1,421 draws. All draws produce a frontier, and the maximum width of the propagated numerical gap envelope is 3.01×10^{-5}. The HAC metric, 102-point structural support, and policy targets remain fixed across draws.

TABLE 8. Full-reprofile calendar bootstrap for the selected LP gap

Block	Point	2.5%	Median	97.5%	Pr($\Delta > 0$) (%)	Peak switches
28	−6.1391	−10.2187	−4.9865	2.2214	6.353	1,421

The table reports descriptive percentile quantiles from 1,999 circular calendar block-multiplier draws with adaptive peak normalization. Each draw reruns Hamilton detrending, lag-augmented local projections, Cholesky impact identification, and all constrained shock-path profiles, then recomputes the family-entry and reversal binders over the 102-point support. The HAC-28 metric, structural support, and policy targets remain fixed. The positive share is a resampling frequency; formal coverage of this range is a separate requirement. All draws produce a frontier. The maximum numerical gap-envelope width is 3.01×10^{-5}.

The positive-gap share is the frequency in this resampling distribution. Both exercises condition on a selected support, and the calendar version also fixes the covariance metric and policy targets. Recomputing frontier binders allows their identities to change. A confidence interpretation additionally requires a valid approximation to the joint law of the nonsmooth frontier. As Fang and Santos (2019) show for directionally differentiable maps, validity of the usual bootstrap needs conditions beyond directional differentiability alone. Active constraints, tied binders, and estimated peak selection are therefore part of a separate coverage assessment. Coverage from the projection theorem pertains to its own moment statistic and compatible nuisance union.

7.5. Magnitude thresholds

For a cumulative-response threshold $\Delta \geq 0$, define $c_j^-(\Delta) = \inf\{d(M) : T_j(M) \geq -\Delta\}$. It measures breakdown of the conclusion $T_j < -\Delta$. The sign frontier uses $\Delta = 0$. Table 9 reports a sensitivity grid in the target's discounted units. At $\Delta = 0.001$, all seventeen inflation cells satisfy the threshold at the benchmark and the first evaluated crossing is 4.0319. At $\Delta = 0.005$, five inflation cells satisfy it and the crossing is 3.5213. Output satisfies all four reported thresholds in every delayed cell at the benchmark.

TABLE 9. Evaluated frontiers for sign and magnitude conclusions

Δ	Output radius	Inflation radius	Output cells	Inflation cells
0.000	–	4.3567	17	17
0.001	–	4.0319	17	17
0.005	–	3.5213	17	5
0.010	4.1257	3.4795	17	0

The conclusion is $T_j < -\Delta$; a reversing evaluation has $T_j \geq -\Delta$. Thresholds use the discounted cumulative units of the policy target and provide a sensitivity grid. Cell counts are out of seventeen per outcome at radius 3.9199. A dash means the evaluated support contains zero reversing evaluations.

For a positive reversal of size $\delta \geq 0$, define $c_j^+(\delta) = \inf\{d(M) : \max_h T_{j,h}(M) \geq \delta\}$. Table 10 applies this object to the common 102-point LP support under HAC 28. The output and inflation

sign frontiers are 17.1449 and 25.8588. At $\delta = 0.0001$, they move to 27.3523 and 36.8739. The largest positive targets on the support are 0.000284 and 0.000814, so the support contains zero qualifying evaluations for $\delta \in \{0.001, 0.005, 0.010\}$. These thresholds are absent from the evaluated support.

TABLE 10. Positive-reversal thresholds in the LP-refitted support

δ	Output radius	Output points	Inflation radius	Inflation points
0.0000	17.1449	5	25.8588	12
0.0001	27.3523	1	36.8739	2
0.0010	–	0	–	0
0.0050	–	0	–	0
0.0100	–	0	–	0

The frontier is $c_j^+(\delta) = \inf\{d(M) : \max_h T_{j,h}(M) \geq \delta\}$ on the common 102-point LP support under HAC bandwidth 28. The zero row is the sign frontier. Positive thresholds distinguish a numerical crossing from a reversal of stated size. A dash records zero qualifying evaluations on this support and carries support-level information only. Targets use the discounted cumulative units of the policy exercise.

Supporting diagnostics. Evidence-block ablations, the CMR calibration point, alternative intervention profiles, convex-completion checks, and weak-rank diagnostics accompany the main frontier. The native-domain certification gate remains open. Each calculation states its units, support, and computational bounds. The baseline and refitted experiments are carried through their sampling and magnitude checks.

8. Conclusion

The distance to a common policy conclusion depends on the response experiment used to compare models. Family entry asks when every declared family has a compatible representation. The reversal radius asks when a compatible representation first reaches a nonnegative target. Their interval gives the range of response tolerances over which family coverage and a common negative sign coexist. Additional evidence moves both thresholds, and its value for this comparison depends on which representations it challenges.

The monetary results illustrate that dependence on declared finite supports. The baseline entry and reversal radii are 3.5058 and 4.3567. Refitting to the local-projection experiment reverses their order under three covariance bandwidths. Common-support accounting traces changes in fitted shock paths and covariance geometry. The positive target crossings are small, which gives magnitude thresholds a separate role alongside the sign boundary.

The calendar re-estimation exercise has central 95-percent gap range $[-10.2187, 2.2214]$. It measures variation from the specified resampling calculation; the uniform projection theorem states conditions for coverage. Native-domain coverage beyond the evaluated support and its stated convex completion remains an open computational gate. Reporting the response protocol,

target, evaluated domain, and uncertainty calculation together specifies the scope of cross-model agreement.

A. Factorization proofs

These proofs record the classical quotient and constant-rank arguments underlying Section 3. All assumptions and result numbering refer to the main text.

A singleton in a Euclidean space is closed. This elementary fact, also recorded in Lemma A1, justifies taking the closure of a constant target image in the first proof.

PROOF OF THEOREM 1. Suppose first that T_p is constant on every fiber of S_A. For $s \in S_A(\mathfrak{E})$, choose any $M_s \in \mathfrak{E}$ satisfying $S_A(M_s) = s$ and define

$$\lambda_p(s) = T_p(M_s).$$

If M_s' is another preimage of s, fiber constancy gives $T_p(M_s') = T_p(M_s)$, so the definition is independent of the choice. For every $M \in \mathfrak{E}$, $\lambda_p\{S_A(M)\} = T_p(M)$, which proves factorization.

If $\widetilde{\lambda}_p$ is another map satisfying $T_p = \widetilde{\lambda}_p \circ S_A$, then for any $s \in S_A(\mathfrak{E})$ and any M_s in its preimage,

$$\widetilde{\lambda}_p(s) = \widetilde{\lambda}_p\{S_A(M_s)\} = T_p(M_s) = \lambda_p(s).$$

Thus the factor is unique on the empirical image.

For the reverse implication, if $T_p = \lambda_p \circ S_A$ and $S_A(M) = S_A(M')$, then $T_p(M) = \lambda_p\{S_A(M)\} = \lambda_p\{S_A(M')\} = T_p(M')$. For a nonempty fiber, its raw policy image is therefore a singleton, and Lemma A1 proves the identified-set conclusion. A response outside $S_A(\mathfrak{E})$ has an empty fiber and receives coverage-failure status. □

PROOF OF PROPOSITION 2. Assume $\ker D_f \subseteq \ker C_{f,p}$. For $y \in \operatorname{Im} D_f$, choose h such that $D_f h = y$ and define

$$\Lambda_{f,p,\theta_0}(y) = C_{f,p}h.$$

If $D_f\widetilde{h} = y$, then $h - \widetilde{h} \in \ker D_f$. Kernel inclusion implies $C_{f,p}(h - \widetilde{h}) = 0$, hence $C_{f,p}h = C_{f,p}\widetilde{h}$; the definition is well defined.

For $y_i = D_f h_i$ and scalars a_i,

$$\Lambda_{f,p,\theta_0}(a_1y_1 + a_2y_2) = \Lambda_{f,p,\theta_0}\{D_f(a_1h_1 + a_2h_2)\} = C_{f,p}(a_1h_1 + a_2h_2),$$

so Λ_{f,p,θ_0} is linear. By construction, $\Lambda_{f,p,\theta_0}D_f h = C_{f,p}h$ for every h. If $\widetilde{\Lambda}D_f = C_{f,p}$, then for every $y = D_f h$, $\widetilde{\Lambda}y = C_{f,p}h = \Lambda_{f,p,\theta_0}y$; uniqueness holds on $\operatorname{Im} D_f$.

For the reverse implication, suppose $C_{f,p} = \Lambda D_f$. If $h \in \ker D_f$, linearity gives $C_{f,p}h = \Lambda(0) = 0$, so $h \in \ker C_{f,p}$. □

PROOF OF THEOREM 2. The constant-rank theorem provides neighborhoods U_1 of θ_0 and V_1 of $S_{A,f}(\theta_0)$, and continuously differentiable coordinate changes under which

$$S_{A,f}(x, z) = (x, 0),$$

where $x \in \mathbb{R}^{r_f}$ and z indexes the remaining coordinates. Shrink U_1 to a product neighborhood $U_0 = X_0 \times Z_0$ whose vertical slices Z_0 are connected. The tangent vectors $(0, v)$ span $\ker DS_{A,f}(x, z)$. By the kernel inclusion,

$$D_z T_{p,f}(x, z)v = 0 \quad \text{for every } v.$$

Thus $D_z T_{p,f}(x, z) = 0$. For fixed x and any $z_0, z_1 \in Z_0$, integrate the derivative along a piecewise continuously differentiable path in the connected coordinate ball Z_0 to obtain $T_{p,f}(x, z_0) = T_{p,f}(x, z_1)$. Define $\lambda_{f,p}(x, 0) = T_{p,f}(x, z_0)$ for any $z_0 \in Z_0$. This is well defined and continuously differentiable because it can be represented using a fixed smooth local section, for example $z = 0$. Composition gives (10).

For uniqueness, any factor map on $S_{A,f}(U_0)$ must take the value $T_{p,f}(x, z)$ at $(x, 0)$, so two such maps coincide. □

References

Andrews, Donald W. K. and Patrik Guggenberger, "Identification- and Singularity-Robust Inference for Moment Condition Models," *Quantitative Economics*, 2019, *10* (4), 1703–1746.

___ **and Xu Cheng**, "Estimation and Inference with Weak, Semi-Strong, and Strong Identification," *Econometrica*, 2012, *80* (5), 2153–2211.

Andrews, Isaiah and Anna Mikusheva, "Conditional Inference with a Functional Nuisance Parameter," *Econometrica*, 2016, *84* (4), 1571–1612.

___ **, Matthew Gentzkow, and Jesse M. Shapiro**, "Measuring the Sensitivity of Parameter Estimates to Estimation Moments," *Quarterly Journal of Economics*, 2017, *132* (4), 1553–1592.

___ **,** ___ **, and** ___ , "On the Informativeness of Descriptive Statistics for Structural Estimates," *Econometrica*, 2020, *88* (6), 2231–2258.

Aruoba, S. Boragan and Thomas Drechsel, "Identifying Monetary Policy Shocks: A Natural Language Approach," 2024. NBER Working Paper 32417; forthcoming, American Economic Journal: Macroeconomics.

Auclert, Adrien, "Monetary Policy and the Redistribution Channel," *American Economic Review*, 2019, *109* (6), 2333–2367.

___ **, Bence Bardoczy, Matthew Rognlie, and Ludwig Straub**, "Using the Sequence-Space Jacobian to Solve and Estimate Heterogeneous-Agent Models," *Econometrica*, 2021, *89* (5), 2375–2408.

Bauer, Michael D. and Eric T. Swanson, "An Alternative Explanation for the "Fed Information Effect"," *American Economic Review*, 2023, *113* (3), 664–700.

Beraja, Martin, "A Semistructural Methodology for Policy Counterfactuals," *Journal of Political Economy*, 2023, *131* (1), 190–201.

Bonhomme, Stéphane and Martin Weidner, "Minimizing Sensitivity to Model Misspecification," *Quantitative Economics*, 2022, *13* (3), 907–954.

Caravello, Tomas E., Alisdair McKay, and Christian K. Wolf, "Evaluating Monetary Policy Counterfactuals: (When) Do We Need Structural Models?," 2026. Working paper, March 4, 2026; earlier version NBER Working Paper 32988.

Cerreia-Vioglio, Simone, Lars Peter Hansen, Fabio Maccheroni, and Massimo Marinacci, "Making Decisions Under Model Misspecification," *Review of Economic Studies*, 2026, *93* (2), 892–925.

Christiano, Lawrence J., Martin Eichenbaum, and Charles L. Evans, "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy," *Journal of Political Economy*, 2005, *113* (1), 1–45.

___ **, Roberto Motto, and Massimo Rostagno**, "Risk Shocks," *American Economic Review*, 2014, *104* (1), 27–65.

Cloyne, James, Clodomiro Ferreira, and Paolo Surico, "Monetary Policy When Households Have Debt: New Evidence on the Transmission Mechanism," *Review of Economic Studies*, 2020, *87* (1), 102–129.

Cocci, Matthew D. and Mikkel Plagborg-Møller, "Standard Errors for Calibrated Parameters," *Review of Economic Studies*, 2025, *92* (5), 2952–2978.

Dufour, Jean-Marie, "Some Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models," *Econometrica*, 1997, *65* (6), 1365–1387.

Fang, Zheng and Andres Santos, "Inference on Directionally Differentiable Functions," *Review of Economic Studies*, 2019, *86* (1), 377–412.

Freyberger, Joachim, "Normalizations and Misspecification in Skill Formation Models," *Review of Economic Studies*, 2026, *93* (4), 2574–2604.

Gabaix, Xavier, "A Behavioral New Keynesian Model," *American Economic Review*, 2020, *110* (8), 2271–2327.

Gertler, Mark and Peter Karadi, "Monetary Policy Surprises, Credit Costs, and Economic Activity," *American Economic Journal: Macroeconomics*, 2015, *7* (1), 44–76.

Giacomini, Raffaella, Toru Kitagawa, and Alessio Volpicella, "Uncertain Identification," *Quantitative Economics*, 2022, *13* (1), 95–123.

Gürkaynak, Refet S., Brian Sack, and Eric T. Swanson, "Do Actions Speak Louder Than Words? The Response of Asset Prices to Monetary Policy Actions and Statements," *International Journal of Central Banking*, 2005, *1* (1), 55–93.

Hamilton, James D., "Why You Should Never Use the Hodrick–Prescott Filter," *Review of Economics and Statistics*, 2018, *100* (5), 831–843.

Hansen, Lars Peter and Thomas J. Sargent, "Robust Control and Model Uncertainty," *American Economic Review*, 2001, *91* (2), 60–66.

___ **and** ___ , *Robustness*, Princeton, NJ: Princeton University Press, 2008.

___ **and** ___ , "Wanting Robustness in Macroeconomics," in Benjamin M. Friedman and Michael Woodford, eds., *Handbook of Monetary Economics*, Vol. 3, Elsevier, 2010, pp. 1097–1157.

Jarociński, Marek and Peter Karadi, "Deconstructing Monetary Policy Surprises—The Role of Information Shocks," *American Economic Journal: Macroeconomics*, 2020, *12* (2), 1–43.

Kalouptsidi, Myrto, Yuichi Kitamura, Lucas Lima, and Eduardo Souza-Rodrigues, "Counterfactual Analysis for Structural Dynamic Discrete Choice Models," *Review of Economic Studies*, 2026. Corrected proof, June 4, 2026; published May 11, 2026, rdag039.

Kang, Zi Yang and Shoshana Vasserman, "Robustness Measures for Welfare Analysis," *American Economic Review*, 2025, *115* (8), 2449–2487.

Kaplan, Greg, Benjamin Moll, and Giovanni L. Violante, "Monetary Policy According to HANK," *American Economic Review*, 2018, *108* (3), 697–743.

Kleibergen, Frank, "Pivotal Statistics for Testing Structural Parameters in Instrumental Variables Regression," *Econometrica*, 2002, *70* (5), 1781–1803.

Masten, Matthew A. and Alexandre Poirier, "Inference on Breakdown Frontiers," *Quantitative Economics*, 2020, *11* (1), 41–111.

McKay, Alisdair and Christian K. Wolf, "What Can Time-Series Regressions Tell Us about Policy Counterfactuals?," *Econometrica*, 2023, *91* (5), 1695–1725.

Negro, Marco Del, Marc P. Giannoni, and Christina Patterson, "The Forward Guidance Puzzle," *Journal of*

Political Economy Macroeconomics, 2023, *1* (1), 43–79.

Olea, José Luis Montiel and Mikkel Plagborg-Møller, "Local Projection Inference Is Simpler and More Robust Than You Think," *Econometrica*, 2021, *89* (4), 1789–1823.

Ottonello, Pablo and Thomas Winberry, "Financial Heterogeneity and the Investment Channel of Monetary Policy," *Econometrica*, 2020, *88* (6), 2473–2502.

Plagborg-Møller, Mikkel and Christian K. Wolf, "Local Projections and VARs Estimate the Same Impulse Responses," *Econometrica*, 2021, *89* (2), 955–980.

Romer, Christina D. and David H. Romer, "A New Measure of Monetary Shocks: Derivation and Implications," *American Economic Review*, 2004, *94* (4), 1055–1084.

Stock, James H. and Jonathan H. Wright, "GMM with Weak Identification," *Econometrica*, 2000, *68* (5), 1055–1096.

Swanson, Eric T., "Measuring the Effects of Federal Reserve Forward Guidance and Asset Purchases on Financial Markets," *Journal of Monetary Economics*, 2021, *118*, 32–53.

Tamer, Elie, "Partial Identification in Econometrics," *Annual Review of Economics*, 2010, *2*, 167–195.