Generated Instruments with Validation Data

Xinran Liu
University of California, Riverside

Dingli Wang
University of California, Riverside

Abstract

Which conclusions about the yield response to Federal Open Market Committee (FOMC) news survive when the policy instrument is imperfect? We study nominal and real yields and inflation compensation around 92 FOMC meetings using an instrument constructed from changes in money-market rates around FOMC announcements. Because the instrument may also reflect nonpolicy news, we use historical benchmark data to quantify the resulting instrument error. Treating the instrument as valid, both nominal and real yields rise, with real yields rising more. Allowing the instrument to be imperfect, the evidence that both yields rise remains more robust than their ordering. This validation evidence therefore distinguishes which parts of the yield response continue to support a policy interpretation.

Keywords: instrumental variables; validation data; monetary policy; real yields; inflation compensation.
JEL codes: C26, C52, E43, E52.

1 Introduction

Does tighter monetary-policy news raise nominal yields because real yields rise, because inflation compensation changes, or both? Since nominal yields combine real yields and compensation for inflation, these components give different economic interpretations of the same movement in nominal interest rates. A rise in real yields implies tighter real financial conditions. If real yields rise by more than nominal yields, inflation compensation falls. Interpreting these responses as effects of monetary policy, however, requires the instrument for policy news to isolate policy variation from other information arriving around FOMC announcements. We ask which conclusions remain supported when that requirement is relaxed.

We measure policy news with a rate factor constructed from money-market surprises around FOMC announcements. The factor captures changes in expected policy, but it may also reflect news about economic conditions or risk premia. Because such news can also move bond yields, the factor need not isolate policy variation. A strong first stage or close agreement with a policy benchmark therefore does not by itself establish exclusion.

Existing evidence documents how monetary-policy surprises move nominal yields, real yields, and inflation compensation, while imperfect-IV methods show how conclusions change when exclusion is relaxed (Conley et al., 2012; Cinelli and Hazlett, 2025). What remains unresolved in our setting is how

external benchmark information should constrain that sensitivity analysis. Historical observations of the instrument and a policy benchmark can reveal the scale of instrument error, but they do not reveal how strongly that error is related to other forces moving yields. We therefore separate these two objects: validation data determine the error scale, while a sensitivity parameter bounds the error–yield correlation. This allows us to ask which signs and relative sizes of the yield responses remain supported as instrument validity is relaxed.

For a yield change Y and a one-year overnight index swap (OIS) rate change D, write $Y = \beta D + \varepsilon$. Let Z denote the rate factor, S the policy benchmark, and u the part of Z not explained by S. For simplicity, we omit intercepts and impose

$$Z = \lambda S + u, \qquad \operatorname{Cov}(S, \varepsilon) = 0, \qquad |\operatorname{Corr}(u, \varepsilon)| \leq \bar{\rho}.$$

The benchmark satisfies exclusion, while the rate factor need not. Historical observations of Z and S estimate the scale of u. We then carry this scale to the target sample under a stated stability assumption, while $\bar{\rho}$ bounds how strongly the instrument error can be related to the yield disturbance.

Treating the factor as a valid instrument, all three communication windows show the same basic pattern: nominal and real yields rise, with the real-yield response larger than the nominal response. On a common ten-basis-point scale for the one-year OIS response, five-year nominal yields rise by 10–12 basis points, real yields by 14–16, and inflation compensation falls by about four. Equity futures also decline. The fall in inflation compensation is concentrated in meetings in which rates rise while equities fall.

Allowing the instrument to be imperfect affects these conclusions differently. At $\bar{\rho} = 0.50$, the 95% confidence sets for both nominal and real yields remain positive in every communication window under either assumption about how the validation-error scale carries to the target sample. When the instrument error is assumed to remain a constant share of the factor's variation, joint inference also supports a larger real-yield response in the statement and full communication windows. For press conferences, both yield responses remain positive, but the data no longer determine which response is larger. Thus the conclusion that both yields rise survives a broader relaxation of instrument validity than the conclusion that real yields rise more.

The closest empirical study is Acosta et al. (2026). Using the same data, they show that tightening surprises are associated with lower inflation compensation. Related work shows that monetary-policy surprises can move real yields substantially and can change the decomposition of nominal yields into real rates and inflation compensation (Beechey and Wright, 2009; Hanson and Stein, 2015; Nakamura and Steinsson, 2018). We ask a different question: which of these yield conclusions remain supported when the policy instrument is allowed to violate exclusion? We evaluate nominal yields, real yields, and inflation compensation using the same meetings, instrument, and validation evidence, so differences across the three conclusions are directly comparable. For inflation compensation, we also compare results with and without validation to show how much the benchmark information reduces uncertainty about the response.

Our empirical design builds on futures-based measures of policy surprises (Kuttner, 2001) and the distinction between current policy actions and news about the expected policy path (Gürkaynak et al., 2005). Different forms of Fed communication can reveal different information sets (Swanson, 2023; Acosta et al., 2026). More broadly, high-frequency policy surprises may also contain central-bank information (Jarociński and Karadi, 2020; Miranda-Agrippino and Ricco, 2021), risk-premium news (Cieslak and Schrimpf, 2019), or responses to public information that are stronger than markets expected (Bauer and Swanson, 2023a,b). These possibilities motivate allowing the rate factor to violate exclusion rather than treating it as a pure policy shock. Evidence that instrument reliability may vary over time (Amir-Ahmadi et al., 2026) further motivates examining how the validation-error scale is carried from the historical sample to later meetings.

Our econometric analysis combines sensitivity analysis for imperfect instruments (Ashley, 2009; Ashley and Parmeter, 2015; Kang et al., 2025) with historical validation data on the instrument and a policy benchmark. Existing approaches use additional outcomes or other restrictions to calibrate departures from exclusion (van Kippersluis and Rietveld, 2018). We instead use observations of the instrument and a policy benchmark to estimate the scale of instrument error, and then bound how strongly that error may be related to the outcome disturbance. This separates what is learned from validation data from what is imposed through sensitivity analysis. Proposition 1 characterizes the resulting coefficient set, and Theorem 1 provides two-sample inference that accounts for estimation of the validation-error scale. Related Anderson–Rubin methods motivate inference that does not rely on a strong first stage (Wang et al., 2018).

The paper first describes the empirical setting and documents the yield responses under exact exclusion in Sections 2 and 3. It then develops the validation-based framework in Section 4 and uses it in Section 5 to determine which conclusions remain supported when the policy instrument is allowed to be imperfect.

2 Empirical setting and validation design

2.1 Meetings and communication windows

We use the U.S. Monetary Policy Event-Study Database of Acosta et al. (2026), which reports intraday changes in money-market rates, nominal Treasury yields, TIPS yields, and equity prices around FOMC statements and press conferences. We use the archived September 17, 2026 release, with the sample ending on June 17, 2026.

We analyze the same 92 meetings in three communication windows. The statement is released first, followed by the press conference. The statement window runs from ten minutes before to twenty minutes after the statement release. The press-conference window runs from ten minutes before to sixty minutes after the press conference begins. The full communication window runs from ten minutes before the statement release to sixty minutes after the press conference begins, covering both events and the intervening period. It is measured directly in the database rather than constructed by adding the two shorter-window changes.

We require the entire press-conference window to fall within the weekday trading session from 9:30 a.m. to 4:00 p.m. Eastern time. This restriction is based on market hours rather than whether a meeting was scheduled: it excludes the Sunday March 15, 2020 conference but retains the March 3, 2020 conference, which occurred during trading hours. We use the same 92 meeting dates in all three windows, so differences across windows do not reflect changes in sample composition. Appendix A gives the full timing and matching rules.

2.2 Instrument construction and external validation

Our instrument Z is a rate factor that summarizes unexpected changes in the current policy rate and in expected policy rates over the following several quarters. It combines five money-market surprises: $MP1$ and $MP2$ measure surprises for the current and next FOMC meetings from federal funds futures, while $ED2$–$ED4$ measure changes in quarterly Eurodollar/SOFR futures rates.[1]

$$Z_{tw} = \widehat{w}_1 MP1_{tw} + \widehat{w}_2 MP2_{tw} + \widehat{w}_3 ED2_{tw} + \widehat{w}_4 ED3_{tw} + \widehat{w}_5 ED4_{tw}, \tag{1}$$

where t indexes meetings and $w \in \{\text{statement}, \text{press conference}, \text{full}\}$ indexes the three communication windows defined above. We estimate the weights from the first principal component of 80 scheduled, rate-complete statement events in 1994–2003 and then hold them fixed across later samples and communication windows. Combining contracts allows the factor to capture news about both the current policy rate and the expected policy path (Gürkaynak et al., 2005). Appendix A reports the estimated weights.

The validation sample contains 56 scheduled statement events in 2004–2010 for which both the fixed rate factor Z and the published Jarociński–Karadi (JK) policy component S are available.[2] We project Z on S and denote the residual by u. We summarize the scale of this disagreement by

$$\widehat{\tau}_V = \widehat{\mathrm{Var}}_V(\widehat{u}_V),$$

where V denotes the validation sample. This is the historical instrument-error scale carried into the later sensitivity analysis. We account for uncertainty in estimating this scale while treating the published JK series as fixed. Appendix A describes the benchmark construction and this treatment of uncertainty.

The three samples therefore have distinct roles:

$$\underbrace{\text{1994–2003: 80 events}}_{\text{Estimate factor weights}} \longrightarrow \underbrace{\text{2004–2010: 56 events}}_{\text{Estimate instrument-error scale}} \longrightarrow \underbrace{\text{2011–2026: 92 meetings}}_{\text{Estimate yield responses}}.$$

The factor weights are estimated before validation and then held fixed, while the validation

[1] The underlying intraday market data in the USMPD are primarily from LSEG Tick History. The construction of $MP1$, $MP2$, and the Eurodollar/SOFR futures series is described in Appendix A of Acosta et al. (2026).

[2] Jarociński and Karadi (2020) use the joint response of interest rates and stock prices around FOMC announcements to separate monetary-policy news from information about the economic outlook. We use their policy component as the external benchmark and maintain its exclusion restriction in the sensitivity analysis.

sample provides the historical error scale used in the sensitivity analysis. Applying this scale to later statement windows requires stability across time. Applying it to press-conference and full communication windows additionally requires an assumption about how the error scale changes across communication windows.

2.3 Outcomes and normalization

Let D_{tw} denote the one-year OIS rate change, and let Y_{tw}^N and Y_{tw}^R denote the five-year nominal Treasury and real TIPS yield changes, all measured in basis points. Inflation compensation is

$$Y_{tw}^{BE} = Y_{tw}^N - Y_{tw}^R. \tag{2}$$

This spread reflects expected inflation, inflation risk compensation, and relative liquidity (D'Amico et al., 2018; Kim et al., 2019).

For each communication window, we estimate

$$Y^k = \beta_k D + \varepsilon_k, \qquad k \in \{N, R, BE\},$$

where N, R, and BE denote nominal yields, real yields, and inflation compensation, respectively. Under exact exclusion,

$$\mathrm{Cov}(Z, \varepsilon_k) = 0,$$

the conventional IV estimator is

$$\widehat{\beta}_k = \frac{\widehat{\mathrm{Cov}}(Z, Y^k)}{\widehat{\mathrm{Cov}}(Z, D)}.$$

Thus β_k measures the response of outcome k relative to the OIS response associated with the same policy shock (Stock and Watson, 2018). Tables and figures report $10\widehat{\beta}_k$, so all responses are expressed per ten-basis-point OIS response. This is a normalization, not the effect of an independently imposed ten-basis-point policy-rate change.

Because the three outcomes use the same meetings and instrument,

$$\widehat{\beta}_{BE} = \widehat{\beta}_N - \widehat{\beta}_R.$$

We estimate the three responses jointly so that inference on the difference between nominal and real yields retains their covariance. Unless otherwise stated, standard errors use a two-lag Bartlett HAC estimator, allowing for heteroskedasticity and dependence over two event lags (Newey and West, 1987).

Section 3 reports these conventional estimates under exact exclusion. Sections 4 and 5 then assess how the conclusions change when exclusion is relaxed using the validation evidence.

3 Nominal yields, real yields and equity prices

3.1 The conventional yield responses

We first report the conventional IV estimates under exact exclusion. Across all three communication windows, nominal and real yields rise, with the real-yield response larger, so inflation compensation falls. Equity futures also decline. Table 1 reports the magnitudes, normalized to a ten-basis-point OIS response.

Table 1: Yield and equity responses to FOMC news

Window	Nominal yield	Real yield	Inflation compensation	Equity futures
Statement	11.39	15.54	-4.15	-0.85
	[7.45, 15.32]	[10.91, 20.18]	[−5.59, −2.72]	[−1.18, −0.53]
Press conference	10.42	14.21	-3.79	-1.17
	[7.73, 13.10]	[11.63, 16.79]	[−5.05, −2.53]	[−1.55, −0.79]
Full communication	12.38	16.39	-4.01	-1.14
	[9.00, 15.76]	[12.11, 20.67]	[−5.39, −2.64]	[−1.57, −0.70]

All responses are normalized to a 10-basis-point increase in one-year OIS. Yield responses are in basis points; equity-futures responses are in percent. Brackets report conventional ratio-based 95% intervals with HAC(2), maintaining exact exclusion of the candidate instrument. The same 92 meetings enter each window. The full window is measured directly, rather than added from the two shorter windows.

Because nominal yields also rise, the fall in inflation compensation reflects a larger increase in real yields rather than a decline in nominal yields. Higher real yields imply tighter real financial conditions. The decline in equity futures is also consistent with tighter financial conditions: higher discount rates reduce the present value of future profits, although expected profits and risk premia can also change (Hanson and Stein, 2015; Cieslak and Schrimpf, 2019). The yield spread does not separately identify changes in expected inflation, inflation risk compensation and relative liquidity.

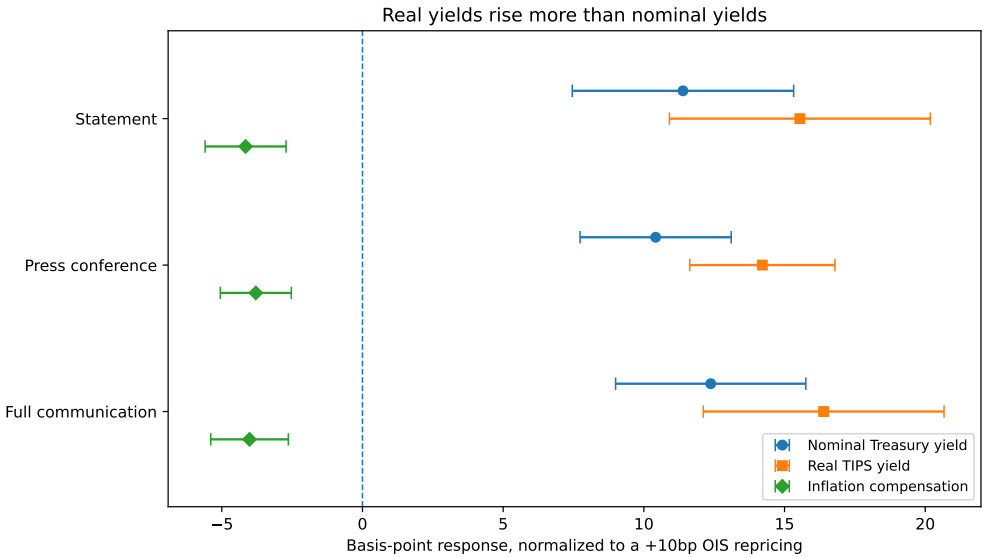


Figure 1: Yield responses across FOMC communication windows

Conventional IV responses under exact exclusion, normalized to a ten-basis-point one-year OIS response. Bars are marginal 95% ratio intervals using HAC(2). The three outcomes use the same observations; the covariance of the nominal and real responses enters the variance of their difference.

We next ask which rate–equity observations account for the decline in inflation compensation.

3.2 Where the negative covariance comes from

The joint movement of rates and equity prices provides a simple way to organize the observations. Jarociński and Karadi (2020) use opposite versus same-sign movements in rates and equities to distinguish monetary-policy news from information about the economic outlook. Other forces, including changes in risk premia, can also generate rate–equity comovement (Cieslak and Schrimpf, 2019).

To locate which observations contribute to the negative inflation-compensation covariance, we divide the instrument according to the signs of the rate factor and equity-futures change, Δp, before centering:

$$Z^- = Z\mathbf{1}\{Z\Delta p < 0\}, \qquad Z^+ = Z\mathbf{1}\{Z\Delta p > 0\}, \qquad Z^0 = Z - Z^- - Z^+.$$

Using the original first-stage covariance for each contribution gives

$$\widehat{\beta}_{BE} = \frac{\widehat{\mathrm{Cov}}(Z^-, Y^{BE})}{\widehat{\mathrm{Cov}}(Z, D)} + \frac{\widehat{\mathrm{Cov}}(Z^+, Y^{BE})}{\widehat{\mathrm{Cov}}(Z, D)} + \frac{\widehat{\mathrm{Cov}}(Z^0, Y^{BE})}{\widehat{\mathrm{Cov}}(Z, D)}. \tag{3}$$

Because the terms share the same denominator, they add exactly to the original inflation-compensation estimate. This partitions the observed covariance by sign group while retaining every meeting.

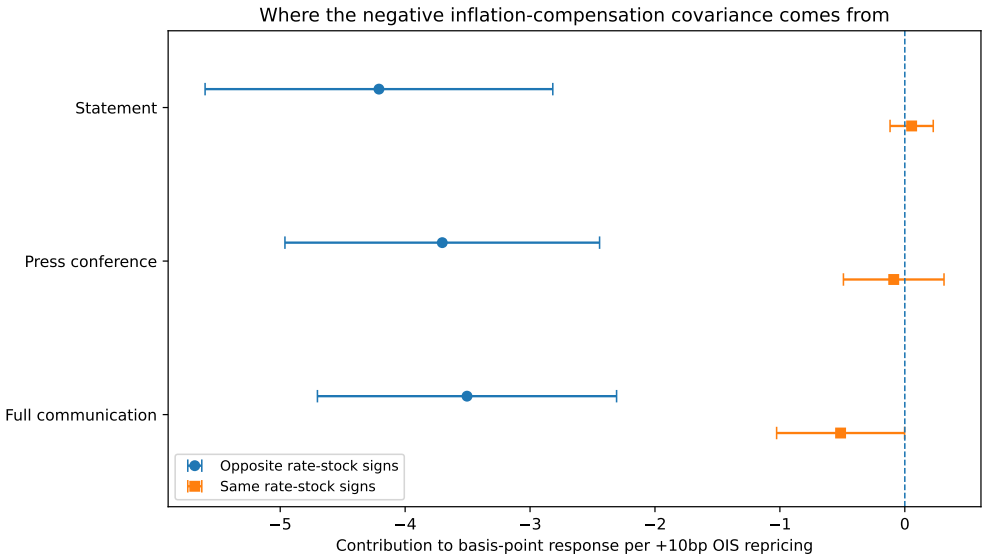


Figure 2: Contributions to the inflation-compensation response by rate–equity sign

The plotted contributions are $10\widehat{\mathrm{Cov}}(Z^{-}, Y^{BE})/\widehat{\mathrm{Cov}}(Z, D)$ and $10\widehat{\mathrm{Cov}}(Z^{+}, Y^{BE})/\widehat{\mathrm{Cov}}(Z, D)$. Bars are regular HAC(2) intervals that include estimation of the common denominator. Zero-sign contributions are retained in the decomposition and are negligible at this scale. The groups partition observed covariance rather than identify structural shocks.

Opposite-sign observations account for most of the negative inflation-compensation response in all three windows. The decline is therefore concentrated in meetings in which rates and equities move in opposite directions, rather than in observations in which they move in the same direction. In Jarociński and Karadi (2020), opposite-sign rate–equity movements are associated with monetary-policy news, while same-sign movements can reflect information about the economic outlook. The decomposition therefore shows that the fall in inflation compensation is not mainly driven by the same-sign movements associated with information about the economic outlook. The same pattern appears during press conferences, so the communication stage alone does not separate policy news from information news. The sign groups organize the observed price movements but do not identify structural shocks.

We next use external validation to ask whether nominal and real yields remain positive, and whether the real-yield response remains larger, when exact exclusion is relaxed.

4 Model, identification and inference

Section 3 shows the yield responses under exact exclusion. We now ask which signs and ordering remain supported when the rate factor may violate exclusion.The analysis separates two questions: how large is the instrument error, and how strongly can that error be related to the yield disturbance? The validation sample measures the historical error variance, and a cross-sample stability assumption maps that variance into the target sample. A sensitivity parameter then bounds the correlation between the transported error and each yield disturbance. Proposition 1 characterizes the resulting coefficient set; Theorem 1 accounts for sampling uncertainty in both the target and validation

samples.

4.1 Validation information and imperfect exclusion

For each yield outcome, retain the equation in Section 2: Y is the yield change, D is the change in the one-year OIS rate, and Z is the generated rate factor. The target sample observes (Y, D, Z); the historical validation sample observes (Z, S), where S is the JK policy benchmark. We suppress outcome and window indices because the same procedure applies throughout. All variables below are residuals from their population projections on the same list of controls, including an intercept. Let T and V denote the target and validation populations, respectively, and write $q_{AB,P} = \mathrm{Cov}_P(A, B)$ for the covariance between any variables A and B in population $P \in \{T, V\}$.

We maintain benchmark exclusion, $\mathrm{Cov}_T(S, \varepsilon) = 0$, and $q_{SS,T} > 0$. To separate the part of the rate factor explained by the benchmark from the remaining error, project the rate factor on the benchmark:

$$Z = \lambda_T S + u_T, \qquad \lambda_T = \frac{q_{ZS,T}}{q_{SS,T}}, \qquad \tau_T = \mathrm{Var}_T(u_T) \leq q_{ZZ,T}. \tag{4}$$

Projection orthogonality gives $\mathrm{Cov}_T(S, u_T) = 0$, while benchmark exclusion implies

$$\mathrm{Cov}_T(Z, \varepsilon) = \mathrm{Cov}_T(u_T, \varepsilon).$$

Thus the benchmark component contributes no covariance with the yield disturbance. Any exclusion violation by the rate factor operates through the benchmark-relative error u_T.

The analogous projection in the validation sample identifies the historical error variance τ_V and factor variance $q_{ZZ,V}$. To use that information for later meetings and different communication windows, we consider two assumptions:

$$\text{Constant absolute variance:} \quad \tau_T = \tau_V, \tag{5}$$

$$\text{Constant relative share:} \quad \tau_T = r_V q_{ZZ,T}, \qquad r_V = \tau_V / q_{ZZ,V}. \tag{6}$$

In sample, these restrictions map the estimated validation scale into the target scale as

$$\widehat{\tau}_T = \widehat{\tau}_V$$

under constant absolute variance, and

$$\widehat{\tau}_T = \frac{\widehat{\tau}_V}{\widehat{q}_{ZZ,V}} \widehat{q}_{ZZ,T}$$

under constant relative share. Validation therefore estimates the historical error scale; the transport assumption determines how that scale applies to the target sample.

The two assumptions differ in how the error scale changes across samples. Constant absolute variance keeps the projection-error variance fixed in squared basis points. Constant relative share instead keeps the unexplained share of factor variance fixed, allowing the error variance to rise or

fall with total factor variance. Both assumptions allow different projection loadings across samples.

This distinction matters across communication windows. The validation and target statement windows both cover thirty minutes, whereas full communication spans a longer period in which both policy and other news can move prices. Constant absolute variance keeps the error variance fixed across windows; constant relative share lets it change with factor variance. Section 5.4 compares the two assumptions with the available stability evidence.

Let τ denote the target error variance implied by the selected transport assumption. Transport determines the scale of the error, but not how strongly it is related to the yield disturbance. We therefore impose

$$|\operatorname{Corr}_T(u_T, \varepsilon)| \leq \bar{\rho}, \qquad 0 \leq \bar{\rho} \leq 1. \tag{7}$$

Validation determines the error scale, while $\bar{\rho}$ limits the error–disturbance correlation. At $\bar{\rho} = 0$, the rate factor satisfies exact exclusion. Positive values allow its projection error to be correlated with the yield disturbance. We apply the bound to each yield equation's own disturbance.

4.2 Identification of the yield responses

For a candidate response β, the imperfect-exclusion restriction requires

$$|\operatorname{Cov}_T(Z_T, Y_T - \beta D_T)| \leq \bar{\rho}\sqrt{\tau_T}\sqrt{\operatorname{Var}_T(Y_T - \beta D_T)}.$$

The left side is the covariance that the candidate β leaves between the rate factor and the disturbance. The right side is the largest covariance allowed by the transported error scale and the correlation bound. To write the restriction compactly, define

$$e(\beta) = Y - \beta D, \quad m(\beta) = a - b\beta, \quad s^2(\beta) = q_{YY,T} - 2\beta q_{YD,T} + \beta^2 q_{DD,T}, \tag{8}$$

where $a = q_{ZY,T}$ and $b = q_{ZD,T}$. At a candidate coefficient, $m(\beta) = \operatorname{Cov}_T(Z, e(\beta))$ and $s^2(\beta) = \operatorname{Var}_T(e(\beta))$.

Proposition 1 (Identification with validation data)**.** *Suppose the target covariance matrix is positive definite and $0 < \tau \leq q_{ZZ,T}$. Given the target second moments, benchmark exclusion, the transported projection-error variance and* (7)*, the sharp coefficient set is*

$$\mathcal{B}(\bar{\rho}; \tau) = \left\{\beta \in \mathbb{R} : |a - b\beta| \leq \bar{\rho}\sqrt{\tau}\, s(\beta)\right\}. \tag{9}$$

Every coefficient in this set admits a joint second-moment completion satisfying the stated benchmark and projection restrictions. No restriction on $\operatorname{Cov}_T(u_T, D)$ or positive lower bound on $|b|$ is required.

At $\bar{\rho} = 0$, (9) reduces to $m(\beta) = 0$, the exact-exclusion condition behind the conventional IV benchmark. As $\bar{\rho}$ increases, the set allows more covariance between the instrument error and the yield disturbance. For nominal and real yields, the relevant question is whether the set excludes all nonpositive responses. For inflation compensation, it is whether the set excludes all nonnegative

responses. These are the three sign restrictions assessed in Section 5. Appendix C proves sharpness and gives the complete set geometry.

Dividing by $\sqrt{q_{ZZ,T}}s(\beta)$ gives the direct instrument–disturbance correlation form,

$$|\operatorname{Corr}_T(Z, e(\beta))| \leq \delta, \qquad \delta = \bar{\rho}\sqrt{\tau/q_{ZZ,T}}. \tag{10}$$

This expression makes the role of benchmark fit transparent: the residual-correlation bound is scaled by the share of factor variance left unexplained by the benchmark.

To assess sensitivity to a larger error scale across windows, we also allow $\tau_T \leq \kappa\tau_V$. The feasible union is

$$\mathcal{B}^U(\bar{\rho}, \kappa) = \left\{\beta : |m(\beta)| \leq \bar{\rho}s(\beta)\sqrt{\min\{\kappa\tau_V, q_{ZZ,T}\}}\right\}. \tag{11}$$

The projection bound prevents the error variance from exceeding the factor variance. For a fixed scale, the smallest error–disturbance correlation bound admitting a candidate is

$$\rho^*(\beta_c) = \frac{|m(\beta_c)|}{\sqrt{\tau}\, s(\beta_c)}. \tag{12}$$

At $\beta_c = 0$, this is the correlation needed to admit a zero yield response; values above one cannot be admitted by any $\bar{\rho} \in [0, 1]$. Establishing a sign requires excluding the entire opposite half-line, not only zero. Appendix C gives the corresponding sign and boundedness conditions, including disconnected sets.

4.3 Two-sample inference

The set in Proposition 1 is a population identified set. In sample, the target moments and the validation error scale are both estimated, so confidence sets must account for uncertainty from both samples. Write the two restrictions as

$$g_j(\beta) = jm(\beta) - \bar{\rho}\sqrt{\tau}\, s(\beta) \leq 0, \qquad j \in \{-1, 1\}. \tag{13}$$

For each candidate, use sample moments and a standard error $\widehat{\sigma}_j(\beta)$ for $\widehat{g}_j(\beta)$. The confidence set is

$$\mathcal{C}_{1-\alpha}(\bar{\rho}) = \left\{\beta \in \mathbb{R} : \frac{\widehat{g}_j(\beta)}{\widehat{\sigma}_j(\beta)} \leq z_{1-\alpha/2} \text{ for } j = -1, 1\right\}. \tag{14}$$

This Anderson–Rubin-type inversion tests a coefficient through its moment rather than dividing by first-stage covariance (Anderson and Rubin, 1949). At $\bar{\rho} = 0$, it reduces to the usual two-sided candidate-moment test.

For independent target and validation samples, let n_T and n_V denote the two sample sizes, and let $\psi_j^T(\beta)$ and $\psi_j^V(\beta)$ denote the corresponding first-order score contributions to $\widehat{g}_j(\beta)$. Then

$$\sigma_j^2(\beta) = \frac{\operatorname{Var}_T(\psi_j^T(\beta))}{n_T} + \frac{\operatorname{Var}_V(\psi_j^V(\beta))}{n_V}. \tag{15}$$

The first term reflects uncertainty in the target market moments; the second reflects uncertainty in the validation error scale. Under relative shares, both the estimated validation variance ratio and the target factor variance enter these scores. Appendix D gives the score formulas, their long-run covariance counterparts and the finite-sample conventions.

Theorem 1 (Two-sample inference with estimated validation information)**.** *Fix the construction and benchmark rules, a transport restriction and a sensitivity bound. Suppose the centered, studentized moment errors at the true coefficient converge uniformly to standard normal distributions over the asserted model class as both sample sizes diverge. Then*

$$\liminf_{n_T,n_V\to\infty} \inf_{P\in\mathcal{P}} P\{\beta_0(P) \in \mathcal{C}_{1-\alpha}(\bar{\rho})\} \geq 1-\alpha. \tag{16}$$

The class can include arbitrarily small or zero first-stage covariance and need not impose a fixed limiting sample-size ratio.

The theorem provides the basis for interpreting the confidence sets in terms of the yield responses. If the confidence set for a nominal or real yield lies entirely above zero, the data support a positive response after accounting for uncertainty in both the target moments and the validation-error scale. If the confidence set for inflation compensation lies entirely below zero, the data support a larger real-yield response than nominal-yield response. These conclusions remain valid under the stated sensitivity bound even when the first-stage covariance is small.

Appendix D gives the regularity conditions and proof. The application uses HAC(2) for both samples and treats their score covariance across periods as negligible. Appendix A states how uncertainty in the published benchmark is handled. We apply the procedure separately to N, R and BE, maintaining benchmark exclusion for each outcome. For joint statements within a window, we allocate the total $\alpha = 0.05$ across the three outcome sets rather than infer the ordering by comparing marginal intervals. A positive nominal-yield set together with a negative inflation-compensation set implies

$$\beta_R > \beta_N > 0, \qquad \text{since} \qquad \beta_{BE} = \beta_N - \beta_R.$$

The inflation-compensation equation therefore provides direct inference on whether real yields rise more than nominal yields while retaining their covariance. Section 5 uses these confidence sets to ask which parts of the conventional yield response remain supported once exact exclusion is relaxed.

5 Which parts of the yield response survive imperfect instruments?

5.1 Which yield conclusions survive?

Allowing the rate factor to violate exclusion affects the three conventional yield conclusions differently. At the main sensitivity checkpoint, $\bar{\rho} = 0.50$, the marginal 95% confidence sets for nominal and real yields lie entirely above zero in every communication window under both constant absolute variance and constant relative share. Thus the evidence that both yields rise survives the relaxation of exact exclusion. Table 2 reports the corresponding confidence sets.

Table 2: Yield responses allowing instrument imperfection

Window	Outcome	Estimate	Absolute variance	Relative share
Statement	Nominal yield	11.39	[5.07, 31.28]	[6.37, 23.31]
	Real yield	15.54	[7.71, 40.06]	[9.48, 30.06]
	Inflation compensation	-4.15	[−11.04, 0.005]	[−8.23, −1.37]
Press conference	Nominal yield	10.42	[5.71, 40.49]	[7.08, 21.21]
	Real yield	14.21	[6.60, 46.38]	[10.16, 25.61]
	Inflation compensation	-3.79	[−11.92, 5.99]	[−7.05, 0.40]
Full communication	Nominal yield	12.38	[7.29, 25.15]	[7.61, 23.01]
	Real yield	16.39	[9.68, 32.80]	[10.18, 30.05]
	Inflation compensation	-4.01	[−8.95, −0.96]	[−8.14, −1.36]

Responses are in basis points per 10-basis-point OIS repricing. Brackets are marginal 95% coefficient sets at $\bar{\rho} = 0.50$, using HAC(2) for the target and validation scores. Each outcome has its own structural residual. Within a window, all outcomes use the same candidate instrument, first stage and estimated validation scale. The statement inflation-compensation upper endpoint under absolute transport is positive (0.0049 basis points). Relative transport gives the same direct-correlation allowance, $\delta = 0.2304$, in all three windows. Reference construction is treated as fixed as described in Appendix A.

The response of inflation compensation is less robust. Under constant relative share, its confidence set lies below zero for statements and full communication but includes zero for press conferences. Under constant absolute variance, only the full-communication set lies entirely below zero. The distinction is economically important: higher real yields imply tighter real financial conditions, whereas a negative inflation-compensation response additionally requires real yields to rise by more than nominal yields.

Figure 3 displays the relative-share results across the three communication windows.

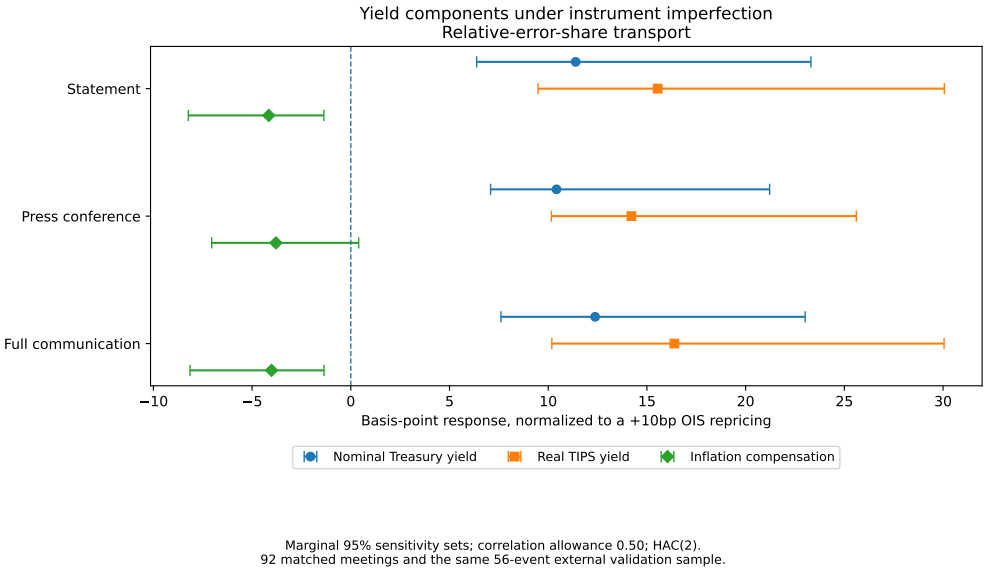


Figure 3: The yield components under imperfect exclusion

Marginal 95% confidence sets under constant relative share, $\bar{\rho} = 0.50$ and HAC(2), normalized to a ten-basis-point one-year OIS response. All panels use 92 meetings and 56 validation observations. Table 2 also reports constant absolute variance.

5.2 How the conclusions change as exclusion is relaxed

The ordering of the yield responses can be assessed directly from inflation compensation:

$$\beta_R > \beta_N > 0 \quad \Longleftrightarrow \quad \beta_N > 0 \text{ and } \beta_{BE} < 0, \qquad \beta_{BE} = \beta_N - \beta_R.$$

We form a 95% family for the three outcomes within each communication window, allocating $\alpha/3$ to each outcome. Table 3 shows how the resulting joint conclusions change as the allowed error–disturbance correlation increases from exact exclusion to $\bar{\rho} = 0.25$ and $\bar{\rho} = 0.50$.

Table 3: Joint yield conclusions across sensitivity allowances

	$\bar{\rho} = 0$	$\bar{\rho} = 0.25$	$\bar{\rho} = 0.50$
Panel A: Constant absolute variance			
Statement	$\beta_R > \beta_N > 0$	$\beta_R > \beta_N > 0$	$\beta_N > 0,\ \beta_R > 0$
Press conference	$\beta_N > 0,\ \beta_R > 0$	$\beta_N > 0,\ \beta_R > 0$	Signs unresolved
Full communication	$\beta_R > \beta_N > 0$	$\beta_R > \beta_N > 0$	$\beta_R > \beta_N > 0$
Panel B: Constant relative share			
Statement	$\beta_R > \beta_N > 0$	$\beta_R > \beta_N > 0$	$\beta_R > \beta_N > 0$
Press conference	$\beta_N > 0,\ \beta_R > 0$	$\beta_N > 0,\ \beta_R > 0$	$\beta_N > 0,\ \beta_R > 0$
Full communication	$\beta_R > \beta_N > 0$	$\beta_R > \beta_N > 0$	$\beta_R > \beta_N > 0$

Each entry reports the 95% joint conclusion for the three outcomes within a communication window, allocating $\alpha/3$ to each outcome. At $\bar{\rho} = 0$, the sensitivity restriction reduces to exact exclusion. The ordering $\beta_R > \beta_N > 0$ follows from a positive nominal-yield confidence set and a negative inflation-compensation confidence set, using $\beta_{BE} = \beta_N - \beta_R$. Positivity without an ordering means that both nominal and real yields rise but their difference is not signed. The family applies within a fixed communication window and transport model.

Two patterns are clear. First, the full-communication window supports

$$\beta_R > \beta_N > 0$$

at all three sensitivity values under both transport assumptions. Second, as the allowed correlation increases, the ordering is generally lost before the evidence that both yields rise. Under constant absolute variance, statements still support two positive responses at $\bar{\rho} = 0.50$ but no longer determine their ordering, while the press-conference joint region no longer determines even the signs. Under constant relative share, statements retain the ordering and press conferences retain two positive responses through $\bar{\rho} = 0.50$.

The main economic conclusion is therefore not simply that sensitivity analysis widens the confidence sets. The evidence that real yields rise, and hence that real financial conditions tighten, is more robust than the stronger conclusion that real yields rise more than nominal yields and inflation compensation falls.

5.3 What validation adds

The preceding results use the historical validation sample to restrict the scale of instrument error. Table 4 isolates what this information contributes for inflation compensation. Without external

information about the error scale, the projection-error variance can be as large as the target factor variance. The second column uses the validation-based scale under constant absolute variance while treating that scale as known. The final column also accounts for uncertainty from estimating it.

Table 4: What external validation adds for inflation compensation

Window	No external scale	Scale treated as known	Estimated scale
Statement	$[-12.89, 0.61]$	$[-10.61, -0.19]$	$[-11.04, 0.005]$
Press conference	$[-10.46, 3.15]$	$[-11.20, 5.24]$	$[-11.92, 5.99]$
Full communication	$[-12.53, 0.76]$	$[-8.80, -1.06]$	$[-8.95, -0.96]$

Basis-point responses per 10-basis-point OIS repricing, $\bar{\rho} = 0.50$ and HAC(2). The no-scale comparison permits the projection error variance to reach the target instrument variance. The next two columns use absolute-variance transport. Treating the estimated scale as known is a diagnostic calculation that omits its sampling uncertainty. The final column includes that uncertainty. The comparison holds the residual-correlation allowance fixed; the implied direct-correlation bounds differ.

The comparison separates the identifying information supplied by validation from the sampling uncertainty introduced by estimating that information. For full communication, the confidence set includes zero when no external scale information is used, but lies below zero once the validation scale is imposed, even after its estimation uncertainty is included. For statements, accounting for estimation of the scale is enough to move the upper endpoint back to essentially zero. Validation can therefore sharpen sign inference, but its sampling uncertainty matters when the conclusion is close to the boundary.

First-stage strength alone does not determine this sign evidence. At the zero-response candidate,

$$m(0) = \mathrm{Cov}(Z, Y),$$

so the first-stage covariance $\mathrm{Cov}(Z, D)$ does not enter the candidate moment directly. A stronger first stage can therefore coexist with weaker evidence about whether a yield response is positive or negative. The sign conclusions depend on the outcome covariance relative to the amount of covariance permitted by the error restriction.

5.4 Stability of the validation error

The remaining issue is whether the historical validation error scale is informative about the later target sample and its different communication windows. Constant absolute variance keeps the projection-error variance fixed in squared basis points. Constant relative share instead lets the error variance move with total factor variance. Statements are closest to the validation design because both use thirty-minute announcement windows; applying the validation scale to longer communication windows requires a stronger cross-window stability assumption.

The same-sign-rule diagnostics in Table 9 provide evidence on this assumption. For statements, the estimated target-to-validation error-variance ratios lie close to the relative-share prediction. For full communication, the point estimates indicate a larger error variance, but the confidence regions are wide. The diagnostics therefore do not sharply distinguish constant absolute variance from

constant relative share.

Figure 4 gives a complementary sensitivity calculation for full communication. Under constant absolute variance, the inflation-compensation confidence set first reaches zero when the allowed target error variance is about 1.86 times the validation variance. The two same-rule diagnostic point estimates fall on opposite sides of this threshold. Because this diagnostic benchmark differs from the published JK benchmark used in the main analysis, these comparisons inform the stability assumption rather than provide a direct recalibration of the main error scale.

The additional checks in Appendix B lead to the same broader conclusion. The negative full-communication inflation-compensation response survives either removing the influential 2022 observations or accounting jointly for the three communication windows, but not both under constant absolute variance. Across the reported specifications, the evidence of higher real yields is more stable than the evidence that inflation compensation falls.

6 Conclusion

Across the same 92 FOMC meetings, the conventional IV estimates show higher nominal and real yields, with the real-yield response larger, so inflation compensation falls. Equity futures also decline, and the negative inflation-compensation response is concentrated in meetings in which rates and equities move in opposite directions.

Allowing the rate factor to violate exclusion separates these conclusions. The evidence that nominal and real yields rise is more robust than the evidence that real yields rise more than nominal yields. At the main sensitivity allowance, both yield responses remain positive across all communication windows under both transport assumptions, while the ordering survives only in some windows and specifications.

Our contribution is to use external validation to discipline this sensitivity analysis. Validation estimates the scale of instrument error, a stability assumption carries that scale to the target sample, and a sensitivity parameter bounds its relation to the yield disturbance. The resulting two-sample inference shows which parts of the yield response remain supported when the policy instrument is imperfect. The main economic conclusion is that uncertainty about lower inflation compensation does not overturn the evidence of higher real yields and tighter real financial conditions.

References

Acosta, M., Ajello, A., Bauer, M. D., Loria, F., and Miranda-Agrippino, S. (2026). Financial market effects of FOMC communication: Evidence from a new event-study database. Working Paper 2025-30, Federal Reserve Bank of San Francisco. Revised August 27, 2026.

Amir-Ahmadi, P., Matthes, C., and Wang, M.-C. (2026). High-frequency instruments with time-varying reliability: Understanding identification in macroeconomics. *Journal of Applied Econometrics*, 41(5):539–548.

Anderson, T. W. and Rubin, H. (1949). Estimation of the parameters of a single equation in a complete system of stochastic equations. *The Annals of Mathematical Statistics*, 20(1):46–63.

Ashley, R. (2009). Assessing the credibility of instrumental variables inference with imperfect instruments via sensitivity analysis. *Journal of Applied Econometrics*, 24(2):325–337.

Ashley, R. A. and Parmeter, C. F. (2015). Sensitivity analysis for inference in 2SLS/GMM estimation with possibly flawed instruments. *Empirical Economics*, 49(4):1153–1171.

Bauer, M. D. and Swanson, E. T. (2023a). An alternative explanation for the "Fed information effect". *American Economic Review*, 113(3):664–700.

Bauer, M. D. and Swanson, E. T. (2023b). A reassessment of monetary policy surprises and high-frequency identification. *NBER Macroeconomics Annual*, 37:87–155.

Beechey, M. J. and Wright, J. H. (2009). The high-frequency impact of news on long-term yields and forward rates: Is it real? *Journal of Monetary Economics*, 56(4):535–544.

Cieslak, A. and Schrimpf, A. (2019). Non-monetary news in central bank communication. *Journal of International Economics*, 118:293–315.

Cinelli, C. and Hazlett, C. (2025). An omitted variable bias framework for sensitivity analysis of instrumental variables. *Biometrika*, 112(2):asaf004.

Conley, T. G., Hansen, C. B., and Rossi, P. E. (2012). Plausibly exogenous. *The Review of Economics and Statistics*, 94(1):260–272.

D'Amico, S., Kim, D. H., and Wei, M. (2018). Tips from TIPS: The informational content of Treasury Inflation-Protected Security prices. *Journal of Financial and Quantitative Analysis*, 53(1):395–436.

Gürkaynak, R. S., Sack, B., and Swanson, E. T. (2005). Do actions speak louder than words? The response of asset prices to monetary policy actions and statements. *International Journal of Central Banking*, 1(1):55–93.

Hanson, S. G. and Stein, J. C. (2015). Monetary policy and long-term real rates. *Journal of Financial Economics*, 115(3):429–448.

Jarociński, M. and Karadi, P. (2020). Deconstructing monetary policy surprises—The role of information shocks. *American Economic Journal: Macroeconomics*, 12(2):1–43.

Kang, H., Guo, Z., Liu, Z., and Small, D. (2025). Identification and inference with invalid instruments. *Annual Review of Statistics and Its Application*, 12:385–405.

Kim, D. H., Walsh, C., and Wei, M. (2019). Tips from TIPS: Update and discussions. FEDS Notes, Board of Governors of the Federal Reserve System, May 21.

Kuttner, K. N. (2001). Monetary policy surprises and interest rates: Evidence from the Fed funds futures market. *Journal of Monetary Economics*, 47(3):523–544.

Masten, M. A. and Poirier, A. (2020). Inference on breakdown frontiers. *Quantitative Economics*, 11(1):41–111.

Miranda-Agrippino, S. and Ricco, G. (2021). The transmission of monetary policy shocks. *American Economic Journal: Macroeconomics*, 13(3):74–107.

Nakamura, E. and Steinsson, J. (2018). High-frequency identification of monetary non-neutrality: The information effect. *The Quarterly Journal of Economics*, 133(3):1283–1330.

Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708.

Stock, J. H. and Watson, M. W. (2018). Identification and estimation of dynamic causal effects in macroeconomics using external instruments. *The Economic Journal*, 128(610):917–948.

Swanson, E. T. (2023). The importance of Fed chair speeches as a monetary policy tool. *AEA Papers and Proceedings*, 113:394–400.

van Kippersluis, H. and Rietveld, C. A. (2018). Beyond plausibly exogenous. *The Econometrics Journal*, 21(3):316–331.

Wang, X., Jiang, Y., Zhang, N. R., and Small, D. S. (2018). Sensitivity analysis and power for instrumental variable studies. *Biometrics*, 74(4):1150–1160.

A Data and implementation

A.1 Data sources and matched samples

All empirical results use the September 17, 2026 USMPD archive, with observations through June 17, 2026. The three samples have separate roles. The factor-construction sample contains 80 scheduled, rate-complete statement events in 1994–2003. The validation sample contains 56 scheduled statements in 2004–2010 for which the fixed rate factor and the published JK policy series are both available. The target sample contains the same 92 meetings in 2011–2026 for each communication window and outcome.

A.2 Factor construction and event matching

The rate factor is the first principal component of $MP1$, $MP2$, $ED2$, $ED3$, and $ED4$ in the 80-event construction sample. The estimated weights are

$$(\widehat{w}_1, \widehat{w}_2, \widehat{w}_3, \widehat{w}_4, \widehat{w}_5) = (0.142399,\ 0.130598,\ 0.222527,\ 0.247224,\ 0.257252).$$

These weights are estimated once and then held fixed in the validation and target samples.

The USMPD contains 93 press conferences through the target endpoint. We retain a conference only when the full window, from ten minutes before to sixty minutes after the conference begins, falls within weekday trading hours from 09:30 to 16:00 Eastern time. This rule excludes the Sunday March 15, 2020 conference and retains the March 3, 2020 conference. It leaves 92 meetings.

Each retained meeting is matched across the Statements, Press_Conferences, and Monetary_Events panels. The same factor weights are applied in all three panels. Nominal yields, real yields, one-year OIS rates, and equity futures are read directly from the corresponding panel. All required variables are observed for the 92 retained meetings. The full-communication return is the database's continuous full-window measure, not the sum of the statement and press-conference returns. The money-market futures series switches from Eurodollar to SOFR contracts in 2022 following the source database.

A.3 Benchmark and validation-scale estimation

The published JK policy series is treated as the external benchmark. The reported confidence sets condition on its realized values and account for uncertainty in the validation projection, but not for uncertainty from the upstream construction of the published JK series. The interpretation therefore maintains benchmark exclusion for the outcome under study.

For absolute transport, let SSE be the residual sum of squares from the validation projection with an intercept and the benchmark regressor. With $n_V = 56$,

$$\widehat{\tau}_V = \frac{SSE}{n_V - 2} = 7.444 \text{ bp}^2.$$

The moment mean uses

$$\widehat{\ell}_V = \sqrt{\widehat{\tau}_V} + \frac{\widehat{\mathrm{Var}}(\widehat{\tau}_V)}{8\widehat{\tau}_V^{3/2}},$$

while the variance calculation uses the first-order derivatives given in Appendix D.

For relative-share transport,

$$\widehat{r}_V = \frac{SSE/n_V}{\widehat{q}_{ZZ,V}} = 1 - \widehat{R}^2_V.$$

Both SSE/n_V and $\widehat{q}_{ZZ,V}$ are treated as estimated quantities when constructing the score variance.

A.4 Inference implementation and replication

Target and validation score variances use a two-lag Bartlett HAC estimator. The two period-specific score sums are treated as having negligible cross-sample covariance, so their estimated variances are added as in equation (15). When observations from 2022 are omitted, the original event indices are retained when forming HAC pairs.

Coefficient confidence sets are obtained by inverting both moment inequalities over the real line. All accepted connected components are retained, including unbounded branches. Familywise calculations allocate the stated error probability across the outcomes or communication windows covered by the corresponding table. A reported zero threshold is the first sensitivity value at which zero enters the confidence set; it is not an estimate of a change in the economic sign of the effect.

The replication archive contains the event-level series used in the paper, source hashes, fixed factor weights, processed inputs, table outputs, figures, and the programs used for estimation and numerical checks. All reported results use the same archived data vintage.

Disclosure. AI tools assisted computation and manuscript preparation. The replication material records the calculations, source files, and numerical checks. Raw proprietary tick histories are not included.

B Additional empirical results

B.1 Complete yield families and smaller sensitivity allowances

Table 5 reports the confidence sets underlying the joint conclusions in Table 3. The absolute-variance press-conference calculation shows why the full set must be retained: the nominal-yield set has two unbounded branches, while the real-yield and inflation-compensation sets are the whole real line. Table 6 reports the corresponding marginal results at the smaller allowance $\bar{\rho} = 0.25$.

Table 5: Intervals underlying the joint yield conclusions

Window	Outcome	Absolute variance	Relative share
Statement	Nominal yield	$[4.37, 44.37]$	$[5.89, 27.90]$
	Real yield	$[6.76, 55.86]$	$[8.89, 35.50]$
	Inflation compensation	$[-14.44, 1.05]$	$[-9.37, -0.89]$
Press conference	Nominal yield	$(-\infty, -18.11] \cup [3.87, \infty)$	$[6.70, 34.39]$
	Real yield	$(-\infty, \infty)$	$[9.56, 37.93]$
	Inflation compensation	$(-\infty, \infty)$	$[-8.47, 3.53]$
Full communication	Nominal yield	$[6.74, 30.37]$	$[7.16, 26.48]$
	Real yield	$[8.90, 39.40]$	$[9.58, 34.42]$
	Inflation compensation	$[-10.62, -0.41]$	$[-9.24, -0.98]$

Basis-point responses per 10-basis-point OIS repricing. Each outcome uses $\alpha = 0.05/3$, so the three sets form a conservative 95% family within a fixed window and transport model. Unbounded components are retained. All calculations use $\bar{\rho} = 0.50$ and HAC(2).

Table 6: Yield responses at a smaller correlation allowance

Window	Outcome	Absolute variance	Relative share
Statement	Nominal yield	$[6.96, 22.67]$	$[7.41, 20.36]$
	Real yield	$[10.15, 29.11]$	$[10.74, 26.23]$
	Inflation compensation	$[-7.81, -1.58]$	$[-6.99, -2.07]$
Press conference	Nominal yield	$[7.47, 23.05]$	$[7.81, 19.30]$
	Real yield	$[10.50, 27.15]$	$[11.24, 22.95]$
	Inflation compensation	$[-7.02, 0.85]$	$[-6.01, -0.60]$
Full communication	Nominal yield	$[8.63, 20.85]$	$[8.71, 20.10]$
	Real yield	$[11.51, 27.15]$	$[11.64, 26.20]$
	Inflation compensation	$[-7.17, -1.95]$	$[-6.89, -2.08]$

Marginal 95% sets at $\bar{\rho} = 0.25$, HAC(2), in basis points per 10-basis-point OIS repricing. All other inputs are identical to Table 2.

B.2 Window comparisons and candidate-zero scores

Table 7 compares the full-communication and press-conference conventional ratios while preserving same-meeting covariance. Both the HAC(2) comparison and the moving-block diagnostic include zero, so the data do not distinguish the two window-specific ratios.

Table 7: Paired coefficient-difference diagnostics

Contrast	Difference	HAC(2) 95% interval	p-value
Full minus PC	-0.022	$[-0.145, 0.101]$	0.723
Full minus Statement	0.014	$[-0.108, 0.136]$	0.821
PC minus Statement	0.036	$[-0.129, 0.201]$	0.666

Pairs share the same meeting dates and their covariance is retained. These are regular ratio diagnostics conditional on candidate construction, not weak-instrument-robust tests of a common structural coefficient. Failure to reject equality is not an equivalence test. The saved paired moving-block comparison is also reported in the text.

Table 8 reports the candidate-zero score components. The autocorrelation in that table refers to the influence score used by the test, not to the raw full-window return. The table also shows why similar conventional IV ratios need not generate similar evidence about the zero-response candidate.

Table 8: Components of the candidate-zero test

Window	Excess moment	Std. error	Statistic	Score ACF(1)	V share (%)
Statement	2.581	1.318	1.957	0.007	8.5
Press conference	2.652	2.246	1.181	0.443	3.8
Full communication	5.836	2.205	2.647	0.191	6.1

Absolute transport, $\bar{\rho} = 0.50$, all 92 meetings, HAC(2). Excess moment is $-\widehat{\mathrm{Cov}}(Z, Y) - \bar{\rho}\widehat{\ell s}_Y$; its uncertainty includes both samples. The one-sided critical value within the two-inequality procedure is 1.960. ACF is for the centered target influence score, not for raw returns. V share refers to this one test statistic, not to the length of its coefficient set.

B.3 Transfer of benchmark disagreement

To examine transport directly, define the same diagnostic benchmark in the validation and target samples,

$$S = Z\mathbf{1}\{Z\Delta p < 0\},$$

and estimate the projection separately in each sample. Table 9 reports the target-to-validation residual-variance ratios. Absolute transport predicts a ratio of one. Relative-share transport predicts the target-to-validation factor-variance ratio reported in the final column. The statement estimates are close to the relative-share prediction, whereas the full-communication point estimates imply a larger discrepancy variance. The confidence regions are wide under both comparisons.

Table 9: Same-rule evidence on transport of the discrepancy variance

Window	Cash ratio	95% region	Futures ratio	95% region	Share prediction
Statement	0.551	$[0.217, 2.033]$	0.524	$[0.211, 1.708]$	0.496
Press conference	1.057	$[0.377, 4.454]$	1.003	$[0.356, 3.649]$	0.512
Full communication	2.373	$[0.482, 10.778]$	1.611	$[0.516, 5.877]$	0.906

Ratio: target/source residual variance after separate projections using the same sign rule $S = Z\mathbf{1}\{Z\Delta p < 0\}$. Absolute transport predicts 1; relative-share transport predicts the final column. Regions use paired window-date blocks, independent source and target moving blocks of four events, and 3,999 draws. These sign benchmarks differ from the published JK median used in the main analysis, so their ratios cannot be inserted into that calibration without an additional assumption.

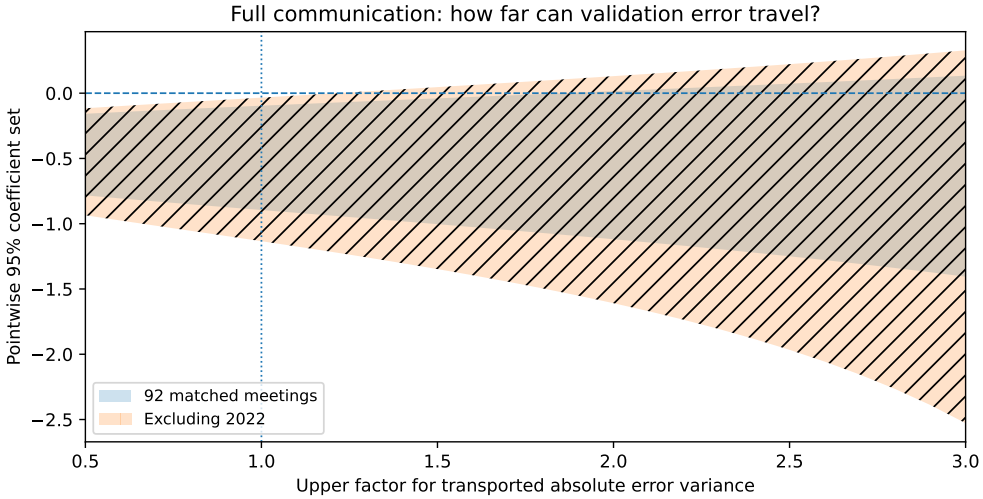


Figure 4: Full-window inflation-compensation sets as the error allowance increases

Pointwise 95% confidence sets under the smooth absolute-variance procedure, with $\bar{\rho} = 0.50$ and HAC(2). Coefficients are per basis point of one-year OIS response. The full-sample confidence set first includes zero near $\kappa = 1.86$; the no-2022 threshold is near 1.21. The same-sign-rule ratios in Table 9 use a different benchmark from the published JK series used in the main analysis.

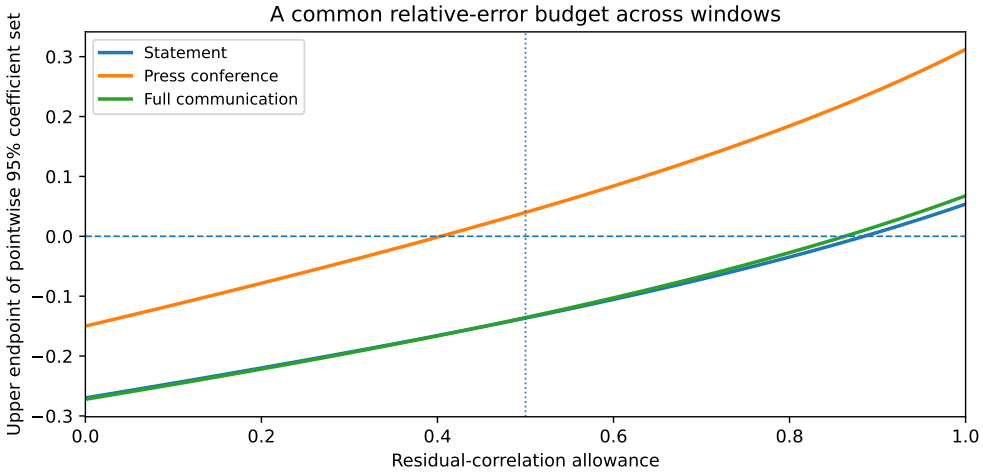


Figure 5: Sensitivity of the inflation-compensation sign across windows

Upper endpoints of pointwise 95% confidence sets under relative-share transport, HAC(2), and the same 56-event validation sample. Coefficients are per basis point of one-year OIS response. The curves vary the residual-correlation allowance and are not a simultaneous confidence band.

B.4 Influential meetings, dependence and window families

Leave-one-meeting-out estimates keep the press-conference conventional response negative. Removing the 2022 meetings makes the response less negative and sharply reduces the serial dependence in the candidate-zero score. Table 10 reports the corresponding results for all three communication windows.

Table 10: The common no-2022 sample across communication windows

Sample	Window	IV ratio	Absolute: 95% set	Relative: 95% set
All 92	Statement	-0.415	$[-1.104, 0.00049]$	$[-0.823, -0.137]$
All 92	Press conference	-0.379	$[-1.192, 0.599]$	$[-0.705, 0.040]$
All 92	Full communication	-0.401	$[-0.895, -0.096]$	$[-0.814, -0.136]$
Without 2022	Statement	-0.355	$[-1.672, 0.167]$	$[-0.912, -0.066]$
Without 2022	Press conference	-0.281	$[-2.083, 0.830]$	$[-0.769, 0.077]$
Without 2022	Full communication	-0.364	$[-1.134, -0.034]$	$[-0.933, -0.102]$

$\bar{\rho} = 0.50$, HAC(2). Removing 2022 removes the same eight meetings from every window and does not bridge the resulting gaps in event-time covariance estimation. The 92-meeting sample remains the main sample.

Table 11 additionally allocates the error probability across the three communication windows. Under constant absolute variance, the full-communication inflation-compensation set remains negative when either the 2022 meetings are removed or the three-window family is imposed separately. When both changes are imposed together, the set includes zero.

Table 11: A three-window error-allocation check

Sample	Window	Absolute transport	Relative transport
All 92	Statement	$[-1.444, 0.105]$	$[-0.937, -0.089]$
All 92	Press conference	$\mathbb{R}$	$[-0.847, 0.353]$
All 92	Full communication	$[-1.062, -0.041]$	$[-0.924, -0.098]$
Without 2022	Statement	$[-5.068, 0.388]$	$[-1.165, -0.014]$
Without 2022	Press conference	$\mathbb{R}$	$[-0.917, 0.145]$
Without 2022	Full communication	$[-1.568, 0.026]$	$[-1.153, -0.071]$

$\bar{\rho} = 0.50$, HAC(2). Each window uses coefficient level $1 - 0.05/3$; the union bound gives nominal 95% family coverage for the three coefficients at a fixed transport model. This is not simultaneous coverage across sensitivity values, alternative transport models, or all preceding exploratory analyses. $\mathbb{R}$ denotes an uninformative coefficient set.

B.5 Additional benchmark diagnostics

When a benchmark is observed together with the target outcomes, benchmark exclusion and relevance identify its IV coefficient,

$$\beta_O = \frac{\mathbb{E}_T(SY)}{\mathbb{E}_T(SD)}.$$

Let

$$e_O = Y - \beta_O D, \qquad u = Z - \lambda S, \qquad \tau_T = \mathrm{Var}_T(u).$$

Then

$$\mu = \mathrm{Cov}_T(Z, e_O), \qquad \rho_u = \frac{\mu}{\sqrt{\tau_T \, \mathrm{Var}_T(e_O)}}. \tag{17}$$

For the press-conference sample, the cash- and futures-based sign benchmarks give similar point estimates of ρ_u, around 0.17, with projected confidence regions that include zero. These diagnostics therefore do not determine the sign of the discrepancy correlation. Appendix E gives the corresponding influence functions and projected-region construction.

Outcome-linked historical validation is also imprecise: the JK-relative correlation estimate is close to zero and its projected region is wide. Splitting the validation sample by time or by the zero lower bound does not provide a tighter correlation restriction.

Prior-information variables from Bauer and Swanson (2023b) explain part of the benchmark discrepancy in sample, but a frozen earlier-period regression does not improve centered out-of-period prediction. Lexical and semantic statement classifiers likewise produce small and uncertain incremental gains. These diagnostics therefore do not provide a stable additional restriction on the instrument error used in the main analysis.

C Identification, geometry and transport thresholds

C.1 Proof of Proposition 1

Fix a candidate β and define

$$e = Y - \beta D, \qquad q = \mathrm{Var}_T(Z), \qquad v = \mathrm{Var}_T(e), \qquad m = \mathrm{Cov}_T(Z, e).$$

Positive definiteness gives

$$q > 0, \qquad v > 0, \qquad qv - m^2 > 0.$$

For necessity, write the target projection as

$$Z = \lambda_T S + u_T, \qquad \mathrm{Cov}_T(S, u_T) = 0, \qquad \mathrm{Var}_T(u_T) = \tau.$$

Benchmark exclusion gives

$$\mathrm{Cov}_T(S, e) = 0.$$

Therefore

$$\begin{aligned} m &= \mathrm{Cov}_T(Z, e) \\ &= \mathrm{Cov}_T(\lambda_T S + u_T, e) \\ &= \lambda_T \,\mathrm{Cov}_T(S, e) + \mathrm{Cov}_T(u_T, e) \\ &= \mathrm{Cov}_T(u_T, e). \end{aligned}$$

The sensitivity restriction implies

$$\left| \frac{\mathrm{Cov}_T(u_T, e)}{\sqrt{\mathrm{Var}_T(u_T)\,\mathrm{Var}_T(e)}} \right| \leq \bar{\rho},$$

hence

$$m^2 \leq \bar{\rho}^2 \tau v.$$

Since $m = a - b\beta$ and $v = s^2(\beta)$,

$$|a - b\beta| \leq \bar{\rho}\sqrt{\tau}\, s(\beta).$$

For sufficiency, suppose

$$m^2 \leq \bar{\rho}^2 \tau v, \qquad 0 < \tau < q.$$

Because $0 \leq \bar{\rho} \leq 1$,

$$\tau v - m^2 \geq 0.$$

Let

$$\Sigma = \begin{pmatrix} q & m \\ m & v \end{pmatrix}, \qquad \det(\Sigma) = qv - m^2 > 0.$$

Define

$$u = c_Z Z + c_e e + \xi, \qquad \begin{pmatrix} c_Z \\ c_e \end{pmatrix} = \Sigma^{-1} \begin{pmatrix} \tau \\ m \end{pmatrix}, \qquad \mathrm{Var}(\xi) = \frac{(q-\tau)(\tau v - m^2)}{qv - m^2}, \tag{18}$$

where ξ is mean zero and independent of (Y, D, Z). Since

$$\Sigma^{-1} = \frac{1}{qv - m^2} \begin{pmatrix} v & -m \\ -m & q \end{pmatrix},$$

the coefficients are

$$c_Z = \frac{\tau v - m^2}{qv - m^2}, \qquad c_e = \frac{m(q-\tau)}{qv - m^2}.$$

The variance assigned to ξ is nonnegative because

$$q - \tau > 0, \qquad \tau v - m^2 \geq 0.$$

Now verify the required moments. First,

$$\begin{aligned} \mathrm{Cov}(Z, u) &= c_Z \,\mathrm{Var}(Z) + c_e \,\mathrm{Cov}(Z, e) \\ &= c_Z q + c_e m \\ &= \frac{q(\tau v - m^2) + m^2(q - \tau)}{qv - m^2} \\ &= \frac{\tau(qv - m^2)}{qv - m^2} \\ &= \tau. \end{aligned}$$

Second,

$$\begin{aligned} \mathrm{Cov}(e, u) &= c_Z \,\mathrm{Cov}(e, Z) + c_e \,\mathrm{Var}(e) \\ &= c_Z m + c_e v \\ &= \frac{m(\tau v - m^2) + mv(q - \tau)}{qv - m^2} \\ &= \frac{m(qv - m^2)}{qv - m^2} \\ &= m. \end{aligned}$$

Third,

$$\begin{aligned}\operatorname{Var}(u) &= c_Z \operatorname{Cov}(Z,u) + c_e \operatorname{Cov}(e,u) + \operatorname{Var}(\xi) \\ &= c_Z\tau + c_e m + \operatorname{Var}(\xi) \\ &= \frac{\tau(\tau v - m^2) + m^2(q-\tau)}{qv-m^2} + \frac{(q-\tau)(\tau v - m^2)}{qv-m^2} \\ &= \frac{\tau(qv-m^2)}{qv-m^2} \\ &= \tau.\end{aligned}$$

Set

$$S = Z - u.$$

Then

$$\begin{aligned}\operatorname{Cov}(S,u) &= \operatorname{Cov}(Z,u) - \operatorname{Var}(u) = \tau - \tau = 0, \\ \operatorname{Var}(S) &= \operatorname{Var}(Z) + \operatorname{Var}(u) - 2\operatorname{Cov}(Z,u) = q - \tau > 0, \\ \operatorname{Cov}(S,e) &= \operatorname{Cov}(Z,e) - \operatorname{Cov}(u,e) = m - m = 0.\end{aligned}$$

Thus

$$Z = S + u$$

is a valid projection with loading one, and the constructed benchmark satisfies exclusion for the candidate disturbance e. Finally,

$$|\operatorname{Corr}(u,e)| = \frac{|m|}{\sqrt{\tau v}} \le \bar{\rho}.$$

Hence every β satisfying (9) is feasible.

The boundary $\tau = q$ is obtained by taking

$$u = Z, \qquad \lambda_T = 0,$$

and choosing S independent of (Y, D, Z) with positive variance. Then

$$\operatorname{Var}(u) = q = \tau, \qquad \operatorname{Cov}(S,u) = 0, \qquad \operatorname{Cov}(S,e) = 0.$$

For the limiting case $\tau = 0$, the restriction becomes

$$m = 0.$$

Taking

$$u = 0, \qquad S = Z$$

satisfies the covariance restrictions. The correlation $\operatorname{Corr}(u,e)$ is undefined at $\tau = 0$, so this limit is expressed through the covariance restriction $m = 0$.

C.2 Coefficient geometry and sign statements

Starting from

$$|a - b\beta| \le \bar{\rho}\sqrt{\tau}\, s(\beta),$$

square both sides:

$$(a - b\beta)^2 \le \bar{\rho}^2\tau \left(q_{YY,T} - 2\beta q_{YD,T} + \beta^2 q_{DD,T}\right).$$

Collecting powers of β gives

$$A\beta^2 - 2C\beta + F \le 0,$$

where

$$A = b^2 - \bar{\rho}^2\tau q_{DD,T}, \qquad C = ab - \bar{\rho}^2\tau q_{YD,T}, \qquad F = a^2 - \bar{\rho}^2\tau q_{YY,T}.$$

Let

$$\Delta = C^2 - AF.$$

If $A > 0$, then

$$\Delta < 0 \quad \Longrightarrow \quad \mathcal{B}(\bar{\rho};\tau) = \varnothing,$$

while for $\Delta \ge 0$,

$$\mathcal{B}(\bar{\rho};\tau) = \left[\frac{C - \sqrt{\Delta}}{A}, \frac{C + \sqrt{\Delta}}{A}\right].$$

If $A < 0$ and $\Delta > 0$, then

$$\mathcal{B}(\bar{\rho};\tau) = \left(-\infty, \frac{C + \sqrt{\Delta}}{A}\right] \cup \left[\frac{C - \sqrt{\Delta}}{A}, \infty\right),$$

where the two roots are ordered after evaluating them. If $A < 0$ and $\Delta \le 0$,

$$\mathcal{B}(\bar{\rho};\tau) = \mathbb{R}.$$

If $A = 0$, the restriction is

$$-2C\beta + F \le 0.$$

Hence

$$\begin{cases} \beta \ge F/(2C), & C > 0, \\ \beta \le F/(2C), & C < 0, \\ \mathbb{R}, & C = 0,\ F \le 0, \\ \varnothing, & C = 0,\ F > 0. \end{cases}$$

These cases are the set components retained in numerical inversion.

For $b \neq 0$, define the conventional ratio

$$\beta_G = \frac{a}{b}$$

and

$$\rho_B = \frac{|b|}{\sqrt{\tau q_{DD,T}}}. \tag{19}$$

A nonempty coefficient set is bounded exactly when

$$A > 0.$$

Using the definition of A,

$$\begin{aligned} A > 0 &\iff b^2 > \bar{\rho}^2 \tau q_{DD,T} \\ &\iff \bar{\rho} < \frac{|b|}{\sqrt{\tau q_{DD,T}}} \\ &\iff \bar{\rho} < \rho_B. \end{aligned}$$

Next define

$$h(\beta) = \rho^*(\beta)^2 = \frac{(a - b\beta)^2}{\tau s^2(\beta)}.$$

Suppose $\beta_G < 0$. The opposite-sign half-line is $\beta \geq 0$, so the zero of the numerator, β_G, lies outside the domain. Differentiating,

$$h'(\beta) = \frac{2(a - b\beta)}{\tau s^4(\beta)} \left[a q_{YD,T} - b q_{YY,T} + \beta \left(b q_{YD,T} - a q_{DD,T} \right) \right].$$

Apart from $\beta = \beta_G$, there is at most one finite stationary point. If

$$B = b q_{YD,T} - a q_{DD,T} \neq 0,$$

that point is

$$\beta^* = \frac{b q_{YY,T} - a q_{YD,T}}{B}.$$

At β^*,

$$(a - b\beta^*)B = -\left(a^2 q_{DD,T} - 2ab q_{YD,T} + b^2 q_{YY,T} \right).$$

Positive definiteness of the covariance matrix of (Y, D) gives

$$a^2 q_{DD,T} - 2ab q_{YD,T} + b^2 q_{YY,T} > 0$$

whenever $(a, b) \neq (0, 0)$. Therefore

$$h''(\beta^*) < 0,$$

so any interior stationary point is a maximum. The infimum over $\beta \geq 0$ must therefore occur at the boundary:

$$h(0) = \rho^*(0)^2$$

or in the limit

$$\lim_{\beta \to \infty} h(\beta) = \frac{b^2}{\tau q_{DD,T}} = \rho_B^2.$$

Hence

$$\inf_{\beta \geq 0} \frac{|a - b\beta|}{\sqrt{\tau}\, s(\beta)} = \min\{\rho^*(0), \rho_B\}.$$

If $\beta_G > 0$, the same argument applies on $\beta \leq 0$. If this infimum exceeds one, no admissible $\bar{\rho} \in [0,1]$ can admit a coefficient with the opposite sign. This distinction between a candidate threshold and a sign threshold follows the sensitivity-frontier interpretation in Cinelli and Hazlett (2025) and Masten and Poirier (2020).

C.3 Feasible upper transport

Suppose

$$0 \leq \tau_T \leq \min\{\kappa\tau_V, q_{ZZ,T}\}.$$

For a fixed β, feasibility requires

$$|m(\beta)| \leq \bar{\rho} s(\beta)\sqrt{\tau_T}.$$

The right-hand side is increasing in τ_T. Therefore there exists a feasible τ_T in the allowed interval if and only if

$$|m(\beta)| \leq \bar{\rho} s(\beta)\sqrt{\min\{\kappa\tau_V, q_{ZZ,T}\}},$$

which is exactly (11).

Now consider the zero candidate. Since

$$m(0) = a, \qquad s^2(0) = q_{YY,T},$$

zero is feasible if and only if

$$a^2 \leq \bar{\rho}^2 q_{YY,T} \min\{\kappa\tau_V, q_{ZZ,T}\}.$$

Assume $a \neq 0$ and $\bar{\rho} > 0$. Define

$$c_0 = \frac{|a|}{\sqrt{q_{ZZ,T} q_{YY,T}}}, \qquad R_0 = \frac{|a|}{\sqrt{\tau_V q_{YY,T}}}.$$

A finite κ can admit zero only if the projection cap itself is large enough:

$$\frac{a^2}{\bar{\rho}^2 q_{YY,T}} \leq q_{ZZ,T}.$$

Equivalently,

$$\bar{\rho} \geq c_0.$$

When this condition holds, the smallest κ satisfies

$$\kappa\tau_V = \frac{a^2}{\bar{\rho}^2 q_{YY,T}},$$

so

$$\kappa_0^*(\bar{\rho}) = \begin{cases} (R_0/\bar{\rho})^2, & \bar{\rho} \geq c_0, \\ +\infty, & \bar{\rho} < c_0. \end{cases} \tag{20}$$

If $a = 0$, zero is already feasible. If $\bar{\rho} = 0$ and $a \neq 0$, zero cannot be admitted.

C.4 Point-estimate robustness on a common scale

For a fixed candidate β_c,

$$\rho^*(\beta_c) = \frac{|m(\beta_c)|}{\sqrt{\tau}\, s(\beta_c)}.$$

Therefore

$$\begin{aligned}\frac{\tau}{q_{ZZ,T}}\rho^*(\beta_c)^2 &= \frac{\tau}{q_{ZZ,T}}\frac{m(\beta_c)^2}{\tau s^2(\beta_c)} \\ &= \frac{m(\beta_c)^2}{q_{ZZ,T}s^2(\beta_c)} \\ &= \mathrm{Corr}_T(Z, Y-\beta_c D)^2\,.\end{aligned}$$

This identity maps the candidate frontier to the direct instrument–disturbance correlation scale used in related robustness calculations (Cinelli and Hazlett, 2025).

D Two-sample scores and confidence-set construction

D.1 Influence functions under the two transport restrictions

All variables below are centered, controls-adjusted population residuals. For a fixed candidate β, define

$$e = Y - \beta D, \qquad s^2 = \mathbb{E}_T(e^2), \qquad m = \mathbb{E}_T(Ze).$$

The target-sample moment influences are

$$\phi_{m,T} = Ze - m, \qquad \phi_{s^2,T} = e^2 - s^2, \qquad \phi_{s,T} = \frac{e^2 - s^2}{2s}.$$

For the validation sample, define

$$\phi_{\tau,V} = u_V^2 - \tau_V, \qquad \phi_{Z,V} = Z_V^2 - q_{ZZ,V},$$

and for the target factor variance define

$$\phi_{Z,T} = Z_T^2 - q_{ZZ,T}.$$

The validation projection error is

$$u_V(\lambda) = Z_V - \lambda S_V, \qquad \tau_V(\lambda) = \mathbb{E}_V[u_V(\lambda)^2].$$

At the projection coefficient,

$$\mathbb{E}_V(S_V u_V) = 0,$$

so

$$\frac{\partial \tau_V(\lambda)}{\partial \lambda} = -2\mathbb{E}_V[S_V u_V(\lambda)] = 0.$$

Thus estimating the projection loading has no first-order effect on $\widehat{\tau}_V$.

Under absolute transport,

$$\tau = \tau_V, \qquad g_j(\beta) = jm - \bar{\rho}\sqrt{\tau}\, s.$$

The differential is

$$dg_j = j\, dm - \bar{\rho}\left(\sqrt{\tau}\, ds + \frac{s}{2\sqrt{\tau}}\, d\tau\right).$$

Hence the target and validation score contributions are

$$\psi_j^T = j(Ze - m) - \frac{\bar{\rho}\sqrt{\tau}}{2s}(e^2 - s^2), \tag{21}$$

$$\psi_j^V = -\frac{\bar{\rho}s}{2\sqrt{\tau}}\phi_{\tau,V}. \tag{22}$$

If $\tau = \kappa\tau_V$ for fixed κ, then

$$d\tau = \kappa\, d\tau_V,$$

and the validation contribution is multiplied accordingly.

Under relative-share transport,

$$r_V = \frac{\tau_V}{q_{ZZ,V}}.$$

The quotient rule gives

$$\phi_{r,V} = \frac{\phi_{\tau,V}}{q_{ZZ,V}} - \frac{\tau_V}{q_{ZZ,V}^2}\phi_{Z,V}. \tag{23}$$

Since

$$\tau = r_V q_{ZZ,T}, \qquad \sqrt{\tau} = \sqrt{r_V}\sqrt{q_{ZZ,T}},$$

the differential of $\sqrt{\tau}$ is

$$d\sqrt{\tau} = \frac{\sqrt{r_V}}{2\sqrt{q_{ZZ,T}}}\, dq_{ZZ,T} + \frac{\sqrt{q_{ZZ,T}}}{2\sqrt{r_V}}\, dr_V.$$

Substituting this expression into

$$dg_j = j\, dm - \bar{\rho}\left(s\, d\sqrt{\tau} + \sqrt{\tau}\, ds\right)$$

gives

$$\psi_{j,R}^T = j(Ze - m) - \frac{\bar{\rho}\sqrt{\tau}}{2s}(e^2 - s^2) - \frac{\bar{\rho}\sqrt{r_V}s}{2\sqrt{q_{ZZ,T}}}\phi_{Z,T}, \tag{24}$$

$$\psi_{j,R}^V = -\frac{\bar{\rho}\sqrt{q_{ZZ,T}}s}{2\sqrt{r_V}}\phi_{r,V}. \tag{25}$$

D.2 Regularity and proof of Theorem 1

For each $j \in \{-1, 1\}$, suppose

$$\widehat{g}_j - g_j = \frac{1}{n_T}\sum_{i=1}^{n_T}\psi_{j,i}^T + \frac{1}{n_V}\sum_{i=1}^{n_V}\psi_{j,i}^V + R_{j,n},$$

with

$$R_{j,n} = o_P(\sigma_{j,n})$$

uniformly over $P \in \mathcal{P}$. Also suppose

$$\frac{\widehat{\sigma}_{j,n}}{\sigma_{j,n}} \to_P 1$$

uniformly and

$$\frac{\widehat{g}_j - g_j}{\widehat{\sigma}_{j,n}} \Rightarrow N(0,1)$$

uniformly over the model class. For independent target and validation samples,

$$\sigma_{j,n}^2 = \frac{\mathrm{Var}_T(\psi_j^T)}{n_T} + \frac{\mathrm{Var}_V(\psi_j^V)}{n_V}.$$

Bounded $4+\eta$ moments, nonsingular fixed-dimensional projection designs, strictly positive scale denominators, and nondegenerate score variances are sufficient regularity conditions for the displayed expansion in the regular independent-sample case.

At the true coefficient β_0,

$$g_j(\beta_0) \le 0 \qquad \text{for } j = -1, 1.$$

Let

$$z = z_{1-\alpha/2}.$$

For each j,

$$\left\{ \frac{\widehat{g}_j(\beta_0)}{\widehat{\sigma}_j(\beta_0)} > z \right\} \subseteq \left\{ \frac{\widehat{g}_j(\beta_0) - g_j(\beta_0)}{\widehat{\sigma}_j(\beta_0)} > z \right\},$$

because $g_j(\beta_0) \le 0$. Therefore

$$\sup_{P \in \mathcal{P}} P\left\{ \frac{\widehat{g}_j(\beta_0)}{\widehat{\sigma}_j(\beta_0)} > z \right\} \le \frac{\alpha}{2} + o(1).$$

The confidence set excludes β_0 only if at least one of the two inequalities rejects:

$$\{\beta_0 \notin \mathcal{C}_{1-\alpha}(\bar{\rho})\} \subseteq \bigcup_{j \in \{-1,1\}} \left\{ \frac{\widehat{g}_j(\beta_0)}{\widehat{\sigma}_j(\beta_0)} > z \right\}.$$

By the union bound,

$$\begin{aligned} \sup_{P \in \mathcal{P}} P\{\beta_0 \notin \mathcal{C}_{1-\alpha}(\bar{\rho})\} &\le \sum_{j \in \{-1,1\}} \sup_{P \in \mathcal{P}} P\left\{ \frac{\widehat{g}_j(\beta_0)}{\widehat{\sigma}_j(\beta_0)} > z \right\} \\ &\le \alpha + o(1). \end{aligned}$$

Hence

$$\inf_{P \in \mathcal{P}} P\{\beta_0 \in \mathcal{C}_{1-\alpha}(\bar{\rho})\} \ge 1 - \alpha - o(1),$$

and therefore

$$\liminf_{n_T, n_V \to \infty} \inf_{P \in \mathcal{P}} P\{\beta_0(P) \in \mathcal{C}_{1-\alpha}(\bar{\rho})\} \ge 1 - \alpha.$$

When $\bar{\rho} = 0$, the validation score is zero and only the target expansion is needed. With dependent observations, the same proof applies after replacing the independent-sample variance by a consistent joint long-run variance. If a construction or benchmark rule is estimated using the evaluation data, its first-order contribution must also be included in the score.

D.3 Threshold inversion and the variance cap

For a fixed candidate β_c, define the first accepted sensitivity value

$$L(\beta_c) = \inf \{\rho \in [0,1] : \beta_c \in \mathcal{C}_{1-\alpha}(\rho)\},$$

with the reporting convention $L(\beta_c) = 1$ if no crossing occurs in $[0,1]$.

The population candidate frontier is

$$\rho^*(\beta_c) = \frac{|m(\beta_c)|}{\sqrt{\tau}\, s(\beta_c)}.$$

If $\rho^*(\beta_c) \in [0,1]$, then

$$\{L(\beta_c) > \rho^*(\beta_c)\} \subseteq \{\beta_c \notin \mathcal{C}_{1-\alpha}(\rho^*(\beta_c))\}.$$

Thus

$$P\{L(\beta_c) > \rho^*(\beta_c)\} \leq \alpha + o(1),$$

so $L(\beta_c)$ is a conservative lower confidence bound for the population frontier.

For the smooth absolute upper-bound procedure,

$$\tau_T \leq \kappa \tau_V,$$

the moment restriction depends on

$$\bar{\rho}\sqrt{\kappa}.$$

Let the untruncated first crossing for the zero candidate be $L(0)$. The corresponding variance factor solves

$$\bar{\rho}\sqrt{\kappa} = L(0),$$

hence

$$\kappa = \left(\frac{L(0)}{\bar{\rho}}\right)^2.$$

The projection cap gives

$$|m(\beta)| \leq \bar{\rho} s(\beta) \sqrt{\min\{\kappa\tau_V, q_{ZZ,T}\}}.$$

This is equivalent to the pair

$$|m(\beta)| \leq \bar{\rho}\sqrt{\kappa\tau_V}\, s(\beta),$$

and

$$|m(\beta)| \leq \bar{\rho}\sqrt{q_{ZZ,T}}\, s(\beta).$$

Construct a $1 - \alpha/2$ coefficient confidence set for each restriction and intersect them. If the two

component sets are denoted $\mathcal{C}_1$ and $\mathcal{C}_2$,

$$\begin{aligned} P\{\beta_0 \notin \mathcal{C}_1 \cap \mathcal{C}_2\} &\leq P\{\beta_0 \notin \mathcal{C}_1\} + P\{\beta_0 \notin \mathcal{C}_2\} \\ &\leq \frac{\alpha}{2} + \frac{\alpha}{2} = \alpha. \end{aligned}$$

Each component therefore uses two one-sided critical values $z_{1-\alpha/4}$.

D.4 Finite-sample conventions

For absolute transport,

$$\widehat{\tau}_V = \frac{SSE}{n_V - 2}$$

with an intercept and one benchmark regressor. The moment mean uses

$$\widehat{\ell}_V = \sqrt{\widehat{\tau}_V} + \frac{\widehat{\mathrm{Var}}(\widehat{\tau}_V)}{8\widehat{\tau}_V^{3/2}},$$

while the first-order variance uses the scores in (22).

For relative-share transport,

$$\widehat{r}_V = \frac{SSE/n_V}{\widehat{q}_{ZZ,V}} = 1 - \widehat{R}^2_V,$$

and the variance uses (25). Treating the estimated validation scale as known sets the validation-score contribution to zero and is used only as the diagnostic comparison reported in the empirical results.

E Inference with a target-linked benchmark

Suppose S is observed together with the target outcomes. Define

$$q_S = \mathbb{E}_T(S^2), \qquad b_S = \mathbb{E}_T(SD), \qquad \lambda = \frac{\mathbb{E}_T(SZ)}{q_S}, \qquad u = Z - \lambda S.$$

Benchmark exclusion and relevance identify

$$\beta_O = \frac{\mathbb{E}_T(SY)}{b_S}.$$

Let

$$e = Y - \beta_O D, \qquad \tau = \mathbb{E}_T(u^2), \qquad q_e = \mathbb{E}_T(e^2),$$

and define

$$a = \mathbb{E}_T(ZY), \qquad b = \mathbb{E}_T(ZD), \qquad \beta_G = \frac{a}{b}, \qquad \mu = \mathbb{E}_T(Ze).$$

Because

$$\mathbb{E}_T(Se) = 0,$$

we also have

$$\mu = \mathbb{E}_T(ue).$$

The discrepancy correlation is

$$\rho_u = \frac{\mu}{\sqrt{\tau q_e}}.$$

For the benchmark IV ratio,

$$\beta_O = \frac{\mathbb{E}_T(SY)}{\mathbb{E}_T(SD)}.$$

The ratio influence function is

$$\phi_{\beta_O} = \frac{S(Y - \beta_O D)}{b_S} = \frac{Se}{b_S}.$$

Since

$$\mu = \mathbb{E}_T[Z(Y - \beta_O D)],$$

its influence is

$$\phi_\mu = Ze - \mu - b\phi_{\beta_O}.$$

Projection orthogonality gives

$$\phi_\tau = u^2 - \tau.$$

For

$$q_e = \mathbb{E}_T[(Y - \beta_O D)^2],$$

differentiation with respect to β_O gives

$$\phi_{q_e} = e^2 - q_e - 2\mathbb{E}_T(De)\phi_{\beta_O}.$$

Applying the delta method to

$$\rho_u = \mu(\tau q_e)^{-1/2}$$

gives

$$\phi_{\rho_u} = \frac{\phi_\mu}{\sqrt{\tau q_e}} - \frac{\rho_u}{2}\left(\frac{\phi_\tau}{\tau} + \frac{\phi_{q_e}}{q_e}\right).$$

The coefficient gap satisfies

$$\begin{aligned}\mu &= \mathbb{E}_T[Z(Y - \beta_O D)] \\ &= a - b\beta_O \\ &= b(\beta_G - \beta_O).\end{aligned}$$

For $b \neq 0$,

$$\phi_{\beta_G} = \frac{Z(Y - \beta_G D)}{b},$$

so

$$\phi_{\beta_G - \beta_O} = \frac{Z(Y - \beta_G D)}{b} - \phi_{\beta_O}.$$

For an Anderson–Rubin projected region, fix a candidate β and define

$$e_\beta = Y - \beta D, \qquad s^2(\beta) = \mathbb{E}_T(e_\beta^2),$$

$$n_\beta = \mathbb{E}_T(ue_\beta), \qquad c_\beta = \mathbb{E}_T(Se_\beta).$$

The loading influence is

$$\phi_\lambda = \frac{Su}{q_S}.$$

Since

$$u = Z - \lambda S,$$

the numerator influence is

$$\phi_{n_\beta} = ue_\beta - n_\beta - c_\beta \phi_\lambda.$$

Define

$$r(\beta) = \frac{n_\beta}{\sqrt{\tau s^2(\beta)}}.$$

Then

$$\phi_{r(\beta)} = \frac{\phi_{n_\beta}}{\sqrt{\tau s^2(\beta)}} - \frac{r(\beta)}{2}\left[\frac{u^2 - \tau}{\tau} + \frac{e_\beta^2 - s^2(\beta)}{s^2(\beta)}\right].$$

Let

$$\mathcal{A}_{1-\alpha/2}$$

be a $1-\alpha/2$ Anderson–Rubin confidence set for β_O, and let

$$\mathcal{I}_{1-\alpha/2}(\beta)$$

be the $1-\alpha/2$ interval for $r(\beta)$, intersected with $[-1, 1]$. Define

$$\mathcal{R}_{1-\alpha} = \bigcup_{\beta \in \mathcal{A}_{1-\alpha/2}} \mathcal{I}_{1-\alpha/2}(\beta).$$

If both

$$\beta_O \in \mathcal{A}_{1-\alpha/2}$$

and

$$\rho_u = r(\beta_O) \in \mathcal{I}_{1-\alpha/2}(\beta_O),$$

then

$$\rho_u \in \mathcal{R}_{1-\alpha}.$$

Therefore

$$\begin{aligned} P\{\rho_u \notin \mathcal{R}_{1-\alpha}\} &\le P\{\beta_O \notin \mathcal{A}_{1-\alpha/2}\} \\ &\quad + P\{\rho_u \notin \mathcal{I}_{1-\alpha/2}(\beta_O)\} \\ &\le \alpha + o(1). \end{aligned}$$

Taking the interval hull of $\mathcal{R}_{1-\alpha}$ is conservative.

F Restriction strength and sampling normalization

For $a \neq 0$, define the two zero-candidate frontiers

$$F_A = \frac{|a|}{\sqrt{\tau_V q_{YY,T}}}, \qquad F_R = \frac{|a|}{\sqrt{(\tau_V / q_{ZZ,V}) q_{ZZ,T} q_{YY,T}}}.$$

Their ratio is

$$\begin{aligned} \frac{F_R}{F_A} &= \frac{\sqrt{\tau_V q_{YY,T}}}{\sqrt{(\tau_V / q_{ZZ,V}) q_{ZZ,T} q_{YY,T}}} \\ &= \sqrt{\frac{q_{ZZ,V}}{q_{ZZ,T}}}. \end{aligned}$$

For the log frontiers, define

$$A_T = \frac{ZY - a}{a} - \frac{Y^2 - q_{YY,T}}{2q_{YY,T}}, \qquad H_T = \frac{Z^2 - q_{ZZ,T}}{2q_{ZZ,T}},$$
$$A_V = -\frac{u_V^2 - \tau_V}{2\tau_V}, \qquad K_V = \frac{Z_V^2 - q_{ZZ,V}}{2q_{ZZ,V}}.$$

Absolute transport has log-frontier influence pair

$$(A_T, A_V),$$

while relative-share transport has

$$(A_T - H_T,\ A_V + K_V).$$

For independent target and validation samples,

$$\operatorname{Var}(\log \widehat{F}_A) = \frac{\operatorname{Var}(A_T)}{n_T} + \frac{\operatorname{Var}(A_V)}{n_V},$$

and

$$\operatorname{Var}(\log \widehat{F}_R) = \frac{\operatorname{Var}(A_T - H_T)}{n_T} + \frac{\operatorname{Var}(A_V + K_V)}{n_V}.$$

Subtracting gives

$$\begin{aligned} \operatorname{Var}(\log \widehat{F}_R) - \operatorname{Var}(\log \widehat{F}_A) = &\frac{\operatorname{Var}(H_T) - 2\operatorname{Cov}(A_T, H_T)}{n_T} \\ &+ \frac{\operatorname{Var}(K_V) + 2\operatorname{Cov}(A_V, K_V)}{n_V}. \end{aligned}$$

The covariance terms have no fixed sign, so neither normalization is uniformly more precise. If $a = 0$, the log expansion is undefined and inference should use the candidate-moment inversion directly.

G Simulation: validation-scale uncertainty and transport misspecification

The simulation separates two issues: sampling uncertainty in the validation error scale and misspecification of the restriction that carries this scale to the target sample. Let

$$S, U, V, W \overset{\text{i.i.d.}}{\sim} N(0,1)$$

and define the target variables by

$$Z = 0.8S + \sqrt{\kappa_T \tau_V}\, U, \qquad D = \pi S + V,$$

$$Y = D + 0.3U + 0.5V + \sqrt{0.66}\, W, \qquad \tau_V = 0.25.$$

The structural coefficient is

$$\beta_0 = 1,$$

because

$$Y = \beta_0 D + \varepsilon, \qquad \varepsilon = 0.3U + 0.5V + \sqrt{0.66}\, W.$$

The disturbance variance is

$$\mathrm{Var}(\varepsilon) = 0.3^2 + 0.5^2 + 0.66 = 1.$$

The target projection error relative to S is

$$u_T = \sqrt{\kappa_T \tau_V}\, U,$$

so

$$\mathrm{Var}(u_T) = \kappa_T \tau_V.$$

Moreover,

$$\mathrm{Cov}(u_T, \varepsilon) = 0.3\sqrt{\kappa_T \tau_V},$$

and therefore

$$\mathrm{Corr}(u_T, \varepsilon) = \frac{0.3\sqrt{\kappa_T \tau_V}}{\sqrt{\kappa_T \tau_V}\sqrt{1}} = 0.30.$$

Thus the true discrepancy–disturbance correlation lies on the reported sensitivity boundary in this design.

The first-stage covariance is

$$\begin{aligned} \mathrm{Cov}(Z, D) &= \mathrm{Cov}(0.8S + \sqrt{\kappa_T \tau_V} U, \pi S + V) \\ &= 0.8\pi. \end{aligned}$$

Varying π therefore changes first-stage relevance without changing the structural coefficient or the discrepancy–disturbance correlation.

The validation sample is generated independently:

$$Z_V = 0.8S_V + 0.5U_V, \qquad S_V, U_V \overset{\text{i.i.d.}}{\sim} N(0,1).$$

Hence

$$u_V = 0.5U_V, \qquad \mathrm{Var}(u_V) = 0.25 = \tau_V.$$

Each design in Table 12 uses 5,000 Monte Carlo replications.

Table 12: Coverage under estimated scale and transport misspecification

n	n_v	True κ	Known scale	Full, correct	Ignore V	Impose $\kappa = 1$
92	56	1.00	0.977	0.980	0.978	0.980
92	56	1.44	0.978	0.977	0.975	0.954
500	100	1.00	0.972	0.972	0.968	0.972
500	100	1.44	0.974	0.974	0.969	0.896

5,000 replications per row, nominal 95% coverage, Gaussian boundary design at $\bar{\rho} = .30$. Simulation standard errors range from .002 to .004. Ignore V uses the correct transport factor but omits validation sampling uncertainty; impose $\kappa = 1$ retains that uncertainty but misspecifies transport in the 1.44 rows. At the truth the tested moment is independent of first-stage relevance under this DGP, so the same coverage check applies to strong, local and zero relevance without counting these as independent experiments.

The known-scale and full-uncertainty procedures use the correct target-to-validation variance factor. The known-scale calculation treats τ_V as fixed, while the full procedure includes the validation-score contribution. At sample sizes close to the application, accounting for estimation of the validation scale changes coverage only modestly in these designs.

The misspecified-transport calculation instead imposes

$$\kappa_T = 1$$

when the true target error variance is larger. In that case coverage can fall below its nominal level, and increasing the sample sizes does not repair the misspecified restriction. Larger samples reduce sampling uncertainty around the wrong transported scale.

At the true coefficient,

$$Y - \beta_0 D = \varepsilon,$$

which does not depend on π. This explains why the coverage calculation at β_0 is relatively insensitive to first-stage strength in this particular simulation. It does not imply that the width or geometry of the full confidence set is independent of first-stage relevance.