

Observational Equivalence in Taylor Rule Estimation

What Is Identified Depends on the Output Gap and the Sample Period

Frank Schmid

frank.schmid@frankschmid.com

This version: 26 July 2026

First version: 3 March 2026

Abstract. Incompatible accounts of monetary policy can explain the same policy rate. This paper asks which of them the data separate, on a real-time dataset rebuilt from primary sources, and shows that the answer turns on which output gap the econometrician is given and over what span. Partial adjustment and a serially correlated policy shock nest in a common autoregressive distributed lag model. On thirty-seven years of data with the Federal Reserve staff's judgmental gap, both restrictions are rejected by nearly identical margins; on sixty years with a mechanically constructed gap, partial adjustment survives and the common factor is rejected. Because the tests involve a near-unit lagged dependent variable and a variance at its boundary, inference is by bootstrap throughout. The drift in the policy intercept cannot be separated from a constant under the staff's gap but can under the mechanical one over the same quarters, so that result is a property of the regressor rather than of the sample. No pairwise difference in fit among seven calibrated specifications is statistically significant. A normative counterpart finds the transmission from policy rate to inflation at approximately zero, on a simulated interval that excludes values of economic consequence, and this survives instrumentation by high-frequency policy surprises. Only the first link of that chain is instrumented, so the result bears on what can be measured rather than on what exists.

Contents

Abbreviations and Acronyms	ii
1 Introduction	1
1.1 Plan of the Paper	2
1.2 Related Literature	2
2 Data	3
2.1 Provenance	3
2.2 Summary	5
3 Fixed-Coefficient Specifications	5
3.1 Fit by Episode	6
3.2 Whose Forecast?	6
4 Time-Varying Parameter Models	7
4.1 Estimation	7
5 What the Data Distinguish	7
5.1 Inertia and Omitted Shocks	8
5.2 The Intercept	9
5.3 The Inflation Coefficient	12
5.4 Backward and Forward Looking	13
6 The Normative Counterpart	14
6.1 A Three-Equation Model	14
6.2 Instrumenting the Policy Gap	15
6.3 Optimized Rules	16
6.4 What the Chain Does Not Establish	17
7 Robustness	18
8 Limitations	19
9 Conclusion	20
A Verification	21
Bibliography	22

Abbreviations and Acronyms

ADL	autoregressive distributed lag
AIC	Akaike information criterion
ALFRED	Archival Federal Reserve Economic Data
BEA	Bureau of Economic Analysis
BIC	Bayesian information criterion
CBO	Congressional Budget Office
COVID	coronavirus disease 2019
CPI	Consumer Price Index
FOMC	Federal Open Market Committee
FRED	Federal Reserve Economic Data
GDP	gross domestic product
HAC	heteroskedasticity- and autocorrelation-consistent
HLW	Holston–Laubach–Williams
HP	Hodrick–Prescott
IS	Investment–Savings
MAD	mean absolute deviation
NAIRU	non-accelerating inflation rate of unemployment
OLS	ordinary least squares
PCE	Personal Consumption Expenditures
RMSD	root mean squared deviation
RTDSM	Real-Time Data Set for Macroeconomists
SPF	Survey of Professional Forecasters
USMPD	United States Monetary Policy Event-Study Database
2SLS	two-stage least squares

Greek Letters

β	beta (BAY-tah)	Okun’s law conversion coefficient (§3)
γ	gamma (GAM-ah)	Coefficient vector in the ADL model (§5)
ε	epsilon (EP-sih-lon)	Observation noise; regression disturbance
η	eta (AY-tah)	State innovation in the Kalman filter (§4)
π	pi (PIE)	Inflation rate; π^* is the inflation target
ρ	rho (ROH)	Interest rate smoothing parameter
σ	sigma (SIG-mah)	Standard deviation of an innovation or disturbance

1 Introduction

A Taylor Rule describes the policy rate as a response to inflation and to slack.¹ Estimating one is harder than the arithmetic suggests, because several incompatible descriptions of Federal Reserve behavior generate nearly the same path for the federal funds rate. A rule with interest rate smoothing and a rule with serially correlated omitted shocks both produce a sluggish policy rate.² A rule responding to forecasts and one responding to realized data agree whenever forecasts track outcomes. A drifting equilibrium real rate and a drifting policy intercept are one object seen from two directions.

The customary response is to report which specification fits best. This paper argues that such rankings are uninformative, and that the productive question is which restrictions the data reject. Three findings follow.

First, the leading accounts of policy sluggishness are nested restrictions on a common autoregressive distributed lag model, so they can be tested rather than compared. Over 1987–2024, with the Federal Reserve staff’s judgmental output gap, both restrictions are rejected at Wald statistics of 10.509 and 10.384, so neither account is preferred to the other. Over 1965–2024, with an output gap constructed mechanically from real-time output, partial adjustment survives at 3.228 while the common-factor restriction is rejected at 20.015. The equivalence is thus not an intrinsic feature of the estimation problem but a property of the span available.

Second, whether the policy intercept can be shown to drift depends on the output gap rather than on the sample. Under the staff’s gap the likelihood ratio against a constant intercept is 1.29, with a bootstrap p -value of 0.073, and the restriction is not rejected; under the mechanical gap over the same 149 quarters it is 3.42, at $p < 0.001$, and the restriction is rejected. The judgmental gap absorbs low-frequency variation that a mechanical trend leaves in the intercept.

Third, no pairwise difference in fit among seven calibrated specifications is statistically significant, including a gap of 0.606 percentage points in root mean squared deviation, which carries a p -value of 0.277. Rankings of the kind routinely reported do not survive a test of equal accuracy.

A second half asks the normative question and finds a sharper answer. The transmission from the policy rate to inflation is estimated at -0.000025 per quarter, with a simulated ninety-five percent interval of $[-0.0076, +0.0075]$, and the near-zero first link survives instrumentation by high-frequency policy surprises at a first-stage F of 18.18. The channel is measured rather than merely unmeasured.

The dataset is itself a result. Mapping a vintage archive to one observation per quarter requires an extraction rule, and the rule is not innocuous: four natural rules applied to the same output-gap release differ by 0.245 percentage points on average and by as much as 1.360 in a single quarter.

¹ John B. Taylor, “Discretion versus Policy Rules in Practice,” *Carnegie-Rochester Conference Series on Public Policy* 39 (1993): 195–214.

² Glenn D. Rudebusch, “Term Structure Evidence on Interest Rate Smoothing and Monetary Policy Inertia,” *Journal of Monetary Economics* 49, no. 6 (2002): 1161–1187.

Every series here is built from a primary release under a rule stated in advance and tested against the alternatives it excludes. §2 documents that work, because the conclusions turn on it.

1.1 Plan of the Paper

§2 documents the dataset and the two output gaps on which everything turns. §3 estimates seven calibrated specifications and tests whether their differences in fit are distinguishable. §4 allows the coefficients to drift. §5 is the paper's principal section: it states the equivalence between partial adjustment and a serially correlated shock as a proposition, tests both restrictions, and finds the inflation coefficient's drift resolvable while the intercept's is resolvable under only one of the two gaps. §6 turns to the normative question and estimates the transmission from the policy rate to inflation at approximately zero, a result that survives instrumentation. §7 reports the robustness program, §8 the limitations, and §9 concludes.

1.2 Related Literature

Clarida, Galí, and Gertler read a pre-Volcker coefficient on inflation below unity as accommodation and a post-Volcker coefficient above unity as the source of stabilization.³ Orphanides showed that the result largely dissolves on real-time data, because the output gaps then believed were far larger than those later measured.⁴ The estimates below reproduce that reversal: on real-time data the inflation coefficient sits *above* unity through the 1970s.

Time-varying parameter methods have been applied to policy by Boivin and by Cogley and Sargent.⁵ They carry a statistical difficulty: the variance governing drift sits at a boundary under the null of no drift, so the likelihood ratio lacks its textbook distribution and estimates pile up at zero.⁶ That difficulty is treated in §5 rather than assumed away.

Orphanides and van Norden establish that statistical detrending in real time does not reproduce judgmental output gaps.⁷ §7 confirms this—the two gaps used here correlate 0.659—and turns it to account, since the divergence is what allows the intercept result to be attributed to the regressor rather than to the sample.

³Richard Clarida, Jordi Galí, and Mark Gertler, "Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory," *Quarterly Journal of Economics* 115, no. 1 (2000): 147–180.

⁴Athanasios Orphanides, "Monetary Policy Rules Based on Real-Time Data," *American Economic Review* 91, no. 4 (2001): 964–985.

⁵Jean Boivin, "Has U.S. Monetary Policy Changed? Evidence from Drifting Coefficients and Real-Time Data," *Journal of Money, Credit and Banking* 38, no. 5 (2006): 1149–1173; Timothy Cogley and Thomas J. Sargent, "Drifts and Volatilities: Monetary Policies and Outcomes in the Post WWII U.S.," *Review of Economic Dynamics* 8, no. 2 (2005): 262–302.

⁶James H. Stock and Mark W. Watson, "Median Unbiased Estimation of Coefficient Variance in a Time-Varying Parameter Model," *Journal of the American Statistical Association* 93, no. 441 (1998): 349–358.

⁷Athanasios Orphanides and Simon van Norden, "The Unreliability of Output-Gap Estimates in Real Time," *Review of Economics and Statistics* 84, no. 4 (2002): 569–583.

2 Data

Every series carries the value a policymaker could have observed at the date in question: the first published vintage, not the current one.⁸ For variables that are never revised this is the same thing; for output, prices, and unemployment it is not.

2.1 Provenance

Each series is built from its authoritative source under an extraction rule fixed in advance and tested against the alternatives it excludes; Table 1 records the sources and rules, and Table A.1 the alternatives.

Table 1: Data sources and construction. Each series maps a source release to one observation per quarter under the rule stated; Table A.1 reports the alternatives considered and the dispersion among them. Coverage is the span of usable quarterly observations.

Series	Source and construction	Coverage
Federal funds rate	FRED, quarterly average of monthly averages	1954–2026
Unemployment rate	RTDSM RUC, first vintage	1965–2026
CPI inflation	RTDSM PCPI, first vintage	1990–2026
GDP deflator inflation	RTDSM P, first vintage, four-quarter change	1965–2026
Output gap (staff)	Greenbook/Tealbook to 2020:Q4; CBO and BEA thereafter	1987:Q3–2024:Q4
Output gap (mechanical)	RTDSM real output, trend fitted within each vintage	1965:Q4–2024:Q4
Expected inflation	SPF median PGDP, one year ahead	1968:Q4–2026
Baa–Aaa spread	FRED, quarterly average of daily	1986–2026
Core PCE inflation	FRED, quarterly average, four-quarter change	1960–2026
r^*	HLW one-sided, 2026:Q1 vintage	1961–2026
π^*	Assumption; 2% from 1994	1965–2024
Policy surprises	Nakamura–Steinsson; USMPD	1994–2026

The two output gaps. This is the paper’s most consequential regressor: it defines the sample and enters six of seven rules and two of three state-space models. The Federal Reserve staff’s real-time estimates are released with a five-year lag.⁹ For quarter t this paper takes the staff estimate of t from the last Greenbook published *during* t . Two releases cover the period and agree to a mean absolute difference of 0.031 over their seventy-two overlapping quarters, which validates the extraction. The earliest vintage in either is August 1987, so this series cannot begin before 1987:Q3; neither release passes 2020:Q4, after which the gap is built from Congressional Budget Office potential against Bureau of Economic Analysis output.

Figure 1 shows the two side by side.

⁸Orphanides, “Monetary Policy Rules Based on Real-Time Data,” 964–985.

⁹Federal Reserve Bank of Philadelphia, *Output Gap and Financial Assumptions from the Board of Governors Tealbook/Greenbook* (data set, Federal Reserve Bank of Philadelphia, 2026).

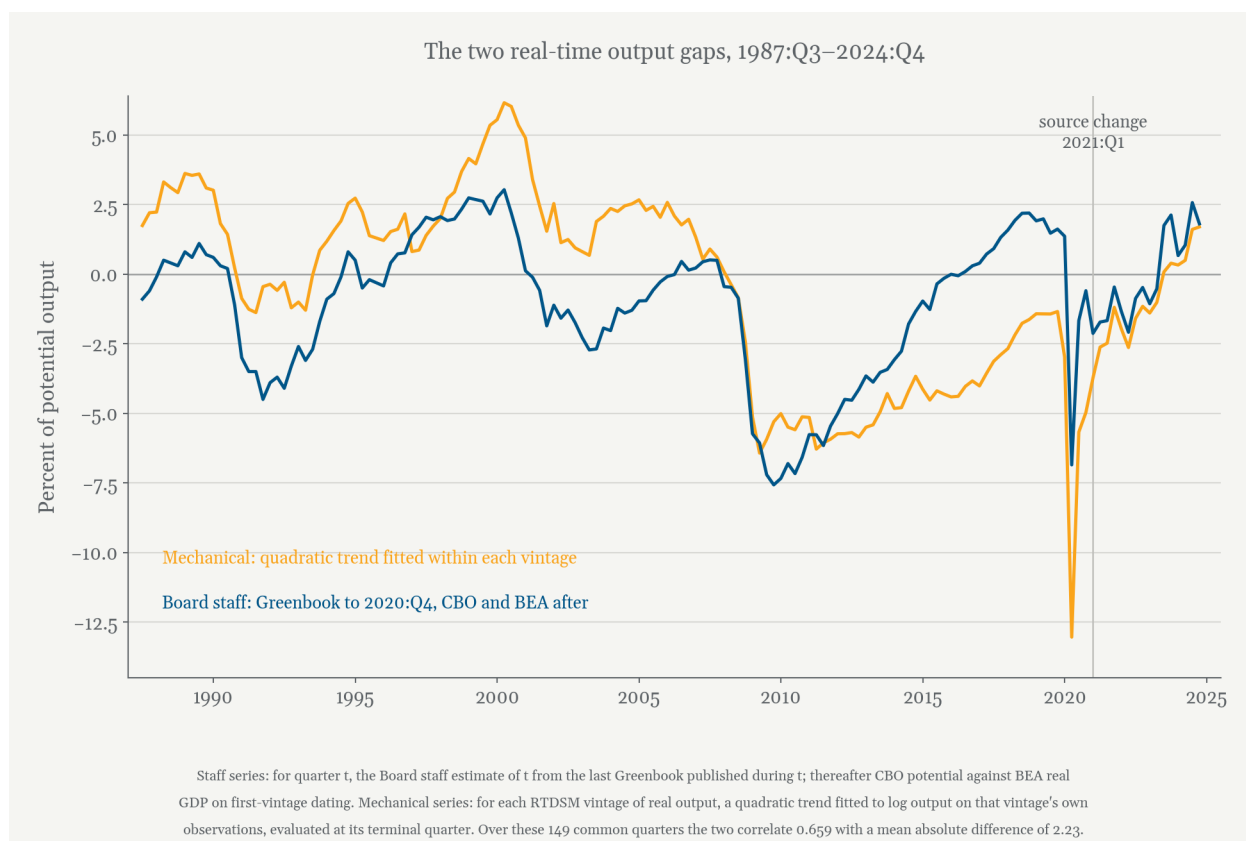


Figure 1: The two real-time output gaps over their common quarters, 1987:Q3–2024:Q4. The horizontal axis is calendar time at quarterly frequency; the vertical axis is percent of potential output. The Board staff's series is a solid deep-blue line and the mechanical series a solid amber line, each labelled directly, so the figure carries no legend. A light vertical rule marks 2021:Q1, where the staff series changes source from Greenbook estimates to Congressional Budget Office potential against Bureau of Economic Analysis output; a horizontal rule marks zero. Neither series is smoothed. The mechanical series is the more volatile of the two, with a standard deviation of 3.37 against 2.51, and the pair correlate 0.659. Construction is given in the bottom band.

The second gap is mechanical and reaches 1965. For each vintage of RTDSM real output, a quadratic trend is fitted to log output using only the observations that vintage contains, and the gap at its terminal quarter is the deviation from that trend. The construction is one-sided at every date by design. The two gaps correlate 0.659 with a mean absolute difference of 2.23 percentage points over their 149 common quarters—they are not substitutes, which is the reason both are carried.

The limits of the audit. A series that reproduces from its source is verified in construction but not in judgment; the extraction rule may still be inappropriate to the question. The Federal Reserve Bank of Philadelphia states of the staff gap that it was never published in the Tealbook, that some observations were reconstructed from records that were at times incomplete, and that errors may remain. That qualification attaches to the most important regressor in this paper and is not removed by rebuilding the series.

2.2 Summary

Table 2 gives the moments of the series over the two estimation windows.

Table 2: Summary statistics over the two estimation windows. All figures are in percent except the output gaps, which are in percent of potential output.

Series	n	Mean	Std. dev.	Minimum	Maximum
<i>Staff gap window, 1987:Q3–2024:Q4</i>					
Federal funds rate	150	3.19	2.65	0.06	9.73
GDP deflator inflation	150	2.32	1.24	0.43	7.50
Output gap (staff)	150	−0.96	2.51	−7.57	3.03
Unemployment rate	150	5.69	1.68	3.50	13.03
SPF expected inflation	150	2.36	0.73	1.25	4.58
<i>Mechanical gap window, 1965:Q4–2024:Q4</i>					
Output gap (mechanical)	236	−0.78	3.37	−13.04	6.15

3 Fixed-Coefficient Specifications

Seven specifications span the main alternatives. None is fitted; each prescribes a rate from contemporaneous data, so the comparison is of calibrated rules against realized policy. M1 is the original rule with r^* and π^* fixed at 2%. M2 replaces both with time-varying series. M3 adds smoothing, $i_t = \rho i_{t-1} + (1 - \rho)i_t^{\text{TV}}$ with $\rho = 0.85$. M4 uses an unemployment gap converted through Okun’s law at $\beta = -2$. M5 replaces realized inflation with the one-year-ahead forecast. M6 adds $-0.5(s_t - \bar{s})$ in the Baa–Aaa spread, and M7 combines M3 with M6. Table 3 reports fit together with tests of equal squared deviation.¹⁰

Table 3: Fit and tests of equal accuracy, 1987:Q3–2024:Q4 (150 quarters), staff output gap. Deviation is the realized rate less the prescribed rate. The Diebold–Mariano statistic tests equal squared deviation against M5, with the long-run variance estimated by a Bartlett kernel at four lags and the Harvey–Leybourne–Newbold small-sample correction applied against Student critical values; because the rules are calibrated rather than fitted, no parameter estimation error contaminates the comparison.

Specification	Mean	MAD	RMSD	DM vs. M5	p
M5 Forward-looking	−0.936	1.333	1.583	—	—
M3 Inertial	−0.884	1.278	1.718	−0.73	0.469
M7 Kitchen sink	−0.902	1.285	1.719	−0.74	0.460
M2 Time-varying	−0.880	1.390	2.189	−1.09	0.277
M6 Financial conditions	−0.903	1.426	2.221	—	—
M1 Original Taylor	−0.809	1.690	2.438	—	—
M4 Unemployment gap	−0.733	1.530	2.489	—	—

¹⁰Francis X. Diebold and Roberto S. Mariano, “Comparing Predictive Accuracy,” *Journal of Business and Economic Statistics* 13, no. 3 (1995): 253–263.

Three observations follow. First, every mean deviation is negative: on published natural-rate estimates the rules prescribe rates above those set, which is a level property of r^* rather than a statement about fit.

Second, the three best-fitting specifications rest on incompatible accounts. M5 represents the Committee as responding to forecasts, M3 as responding to realized data with delay, M7 as doing both while conditioning on credit spreads. M3 and M7 are separated by 0.001 percentage points.

Third, and most importantly, no difference is statistically significant. The widest gap tested—M5 over M2, 0.606 percentage points—carries a p -value of 0.277. The ordering ranks the specifications without distinguishing them.

3.1 Fit by Episode

A full-sample ordering conceals whether any specification describes policy consistently. Table 4 splits the window at the turn of the century, at the financial crisis, at the exit from the zero lower bound, and at the pandemic.

Table 4: Root mean squared deviation by episode, in percentage points. Boundaries are institutional, not selected to favor any specification. The best fit in each row is italicized.

Episode	M1	M2	M3	M4	M5	M6	M7	n
1987:Q3–1999:Q4	1.21	0.71	1.22	0.66	1.00	0.73	1.23	50
2000:Q1–2007:Q4	1.15	1.67	1.97	1.83	1.59	1.71	1.99	32
2008:Q1–2015:Q4	1.54	1.36	1.48	1.76	1.36	1.39	1.39	32
2016:Q1–2019:Q4	2.94	2.01	1.00	1.79	2.31	2.07	1.04	16
2020:Q1–2024:Q4	5.30	4.90	2.82	5.70	2.24	4.95	2.85	20

Four specifications lead across five episodes, and the reversals are not marginal. M3 and M7—second and third over the full sample—are the two *worst* in each of the first two episodes and the two best in the fourth. M1, last overall, leads over 2000–2007. The full-sample ordering averages over periods in which different specifications describe policy.

3.2 Whose Forecast?

M5 uses the Survey of Professional Forecasters, a private-sector expectation, as the Federal Reserve’s own. The staff forecast is the more natural object, and the Greenbook archive supplies it. Constructing the one-year-ahead deflator forecast by chaining the staff’s four quarterly projections gives a series correlating 0.948 with the survey at a mean absolute gap of 0.297.

Over the 134 quarters both cover, M5 on the staff forecast attains a root mean squared deviation of 1.319 against 1.482 on the survey. The improvement is not attributable to smoothness: the staff forecast has a standard deviation of 0.86 against the survey’s 0.76. A *more* variable regressor fits better, which weakens the standard objection that M5’s advantage is mechanical.

4 Time-Varying Parameter Models

If the coefficients drift, every fixed-coefficient comparison is misspecified. Three state-space models are estimated by maximum likelihood with all coefficients following random walks:

$$i_t = B_{0,t} + B_{1,t}\pi_t + B_{2,t}\hat{y}_t + \varepsilon_t, \quad B_{j,t} = B_{j,t-1} + \eta_{j,t}. \quad (1)$$

Model B adds a lagged funds rate, so the smoothing parameter itself drifts; Model C uses the unemployment gap.

4.1 Estimation

The observation variance R and the state innovation variances cannot both be estimated freely from 150 observations: the likelihood is nearly flat along the ridge trading one against the other, so any split of the total variation between drift and noise fits about as well as any other. This is the paper's own subject appearing inside its estimator. R is therefore fixed at 40% of the ordinary least squares residual variance and the innovation variances optimized, with initial states at the least squares estimates. §7 reports the range from 20% to 80%. Where the paper draws inferences about drift it draws them from likelihood ratios rather than point estimates, precisely because the point estimates inherit this choice.

Table 5: Kalman filter estimates, 1987:Q3–2024:Q4, staff output gap. The information criteria count $k = 2s + 1$ parameters for s states; the ranking does not turn on the convention. RMSE is filtered (one-sided) and smoothed (two-sided).

Model	R	logL	AIC	BIC	RMSE filt.	RMSE smooth.
A: No smoothing	1.734	−228.26	471	492	0.589	0.509
B: With smoothing	0.068	−49.57	117	144	0.135	0.150
C: Unemployment gap	1.907	−234.27	483	504	0.559	0.455

Model B dominates on every criterion by a log-likelihood margin near 180. No parameter-count convention and no boundary correction disturbs it: interest rate smoothing belongs in the specification. Its estimated innovation standard deviation on the smoothing parameter is 0.032—the parameter drifts, which §5 takes up.

5 What the Data Distinguish

The preceding sections establish that several specifications fit comparably and that the coefficients drift. This section asks which of the competing accounts the data separate, taking in turn the two explanations of policy sluggishness (§5.1), the policy intercept (§5.2), the coefficient on inflation (§5.3), and the distinction between backward- and forward-looking rules (§5.4).

5.1 Inertia and Omitted Shocks

Let x_t collect the arguments of the rule. Under partial adjustment,

$$i_t = \rho i_{t-1} + (1 - \rho) \gamma' x_t + \varepsilon_t; \quad (2)$$

under a serially correlated omitted influence,

$$i_t = \gamma' x_t + u_t, \quad u_t = \rho u_{t-1} + \varepsilon_t. \quad (3)$$

Proposition 1. *Equations (2) and (3) are nested restrictions on*

$$i_t = \rho i_{t-1} + \gamma'_0 x_t + \gamma'_1 x_{t-1} + \varepsilon_t, \quad (4)$$

with partial adjustment imposing $\gamma_1 = 0$ and the omitted-shock account imposing the common-factor restriction $\gamma_1 = -\rho \gamma_0$. The two are observationally equivalent as x_t approaches a random walk, since the restrictions then differ only in $x_t - x_{t-1}$.

Proof. Substituting $u_{t-1} = i_{t-1} - \gamma' x_{t-1}$ into (3) gives $i_t = \rho i_{t-1} + \gamma' x_t - \rho \gamma' x_{t-1} + \varepsilon_t$, which is (4) with $\gamma_0 = \gamma$ and $\gamma_1 = -\rho \gamma$. Equation (2) is (4) with $\gamma_0 = (1 - \rho) \gamma$ and $\gamma_1 = 0$. Under the two restrictions the fitted values differ by $\gamma'_0 (x_t - x_{t-1}) \cdot \rho$, which vanishes in the random-walk limit. $\square$

Over this sample the first-order autocorrelation is 0.938 for inflation and 0.914 for the staff output gap, so the limit is nearly attained. Table 6 reports the tests.

Table 6: Wald tests of the restrictions on equation (4), with Newey–West standard errors at three lags. “PA” is partial adjustment, $\gamma_1 = 0$; “CF” the common factor, $\gamma_1 = -\rho \gamma_0$. Because the regression carries a lagged dependent variable at ρ near 0.94, the asymptotic χ^2_2 distribution is unreliable; p -values are from a parametric bootstrap under each restriction with 499 replications, and the bootstrap five percent critical values are reported for comparison with the asymptotic 5.99.

Sample and output gap	n	Partial adjustment			Common factor		
		Wald	boot. p	5% cv	Wald	boot. p	5% cv
1987:Q3–2024:Q4, staff	149	10.509	0.048	10.28	10.384	0.036	8.77
1965:Q4–2024:Q4, mechanical	235	3.228	0.281	9.07	20.015	0.000	7.84
1965:Q4–1987:Q2, mechanical	86	17.62	—	—	16.43	—	—

The choice of critical value is material. Asymptotic values would place both first-row statistics in the rejection region, at p near 0.005, whereas the bootstrap distribution sets the five percent critical values at 10.28 and 8.77. The asymptotic test therefore over-rejects, and partial adjustment is rejected only marginally, at $p = 0.048$. The first row nonetheless retains its interpretation: both restrictions are rejected, by statistics differing by 0.125, so neither account is preferred to the other.

The second row differs, and the bootstrap reinforces it. Over six decades partial adjustment survives at $p = 0.281$ while the common factor is rejected at $p < 0.001$. The restrictions are separated, in favor of partial adjustment, which supplies the interpretation that Model B's dominance in Table 5 leaves open.

The third row indicates that the separation is not a consequence of sample size alone: the pre-1987 subsample rejects both restrictions, as the short modern sample does, on asymptotic statistics of 17.62 and 16.43 that exceed every bootstrap critical value obtained above. The separation emerges from the full span rather than from either half.

The state-space estimates point the same way. Model B's innovation standard deviation on the smoothing parameter, 0.032, indicates a coefficient that drifts rather than one held fixed, which is what a persistent omitted influence would produce and partial adjustment alone would not. The restriction tests and the filter therefore disagree on this sample and agree on the longer one.

5.2 The Intercept

Under the Taylor coefficient structure the intercept corresponds to $r^* - 0.5\pi^*$, so a drifting equilibrium real rate should appear as a drifting intercept. Figure 2 sets Model A's intercept against the Holston–Laubach–Williams natural rate.

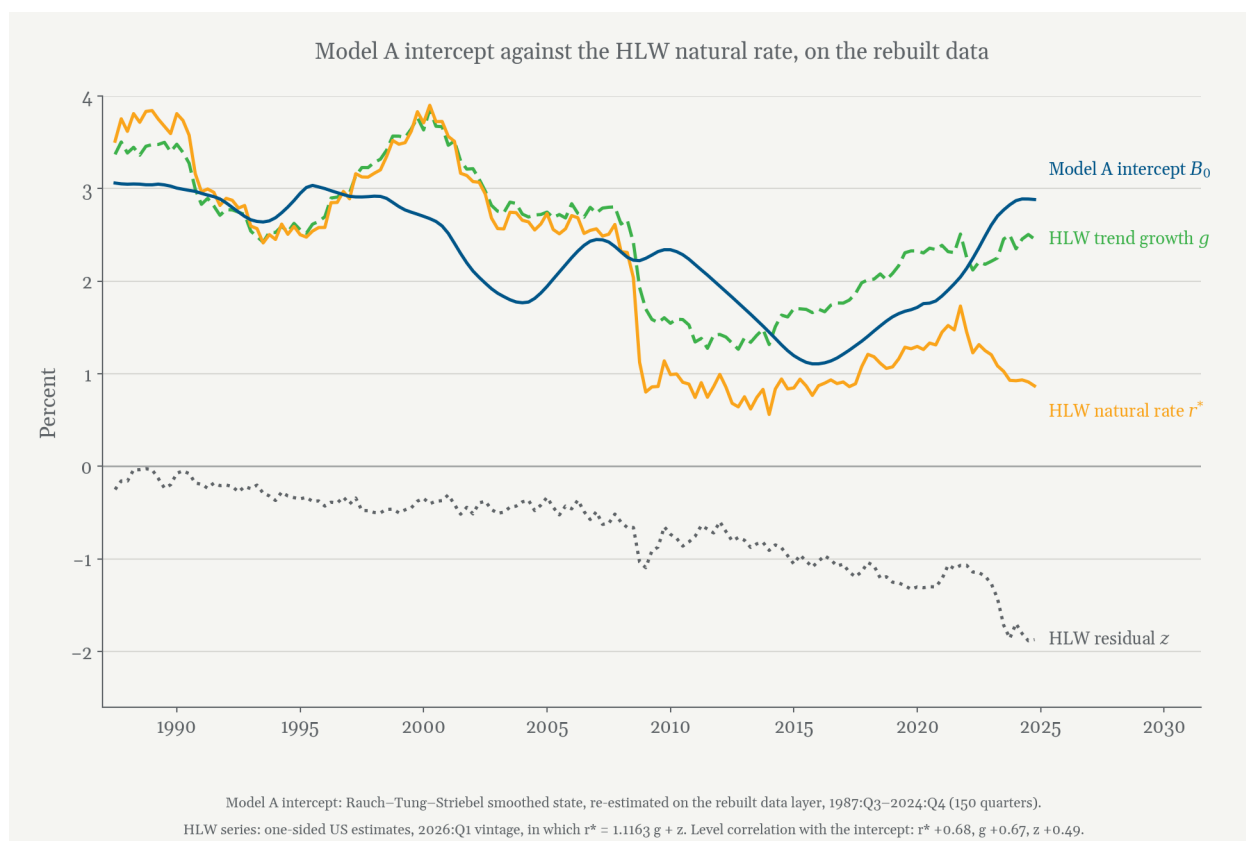


Figure 2: Model A's smoothed intercept against the Holston–Laubach–Williams natural rate and its decomposition, 1987:Q3–2024:Q4. The horizontal axis is calendar time at quarterly frequency, the vertical axis percent. The intercept B_0 is a solid deep-blue line, r^* a solid amber line, annualized trend growth g a dashed green line, and the residual z a dotted gray line, with $r^* = c g + z$ and $c = 1.1163$ in this vintage. Series are labeled directly. Sources: the re-estimated filter and the 2026:Q1 vintage of the Federal Reserve Bank of New York's real-time file.

The level correlation is $+0.68$ with the natural rate and $+0.67$ with trend growth; in four-quarter differences these fall to $+0.13$ and $+0.23$. The correspondence is one of trend, not of year-over-year movement. A qualification bounds it: the natural-rate model takes the real short rate as an input, so the funds rate enters both objects and part of any agreement is mechanical.

Whether the intercept moves at all is a separate question, and testing it is awkward because the null sets a variance to its boundary. Figure 3 profiles the likelihood.

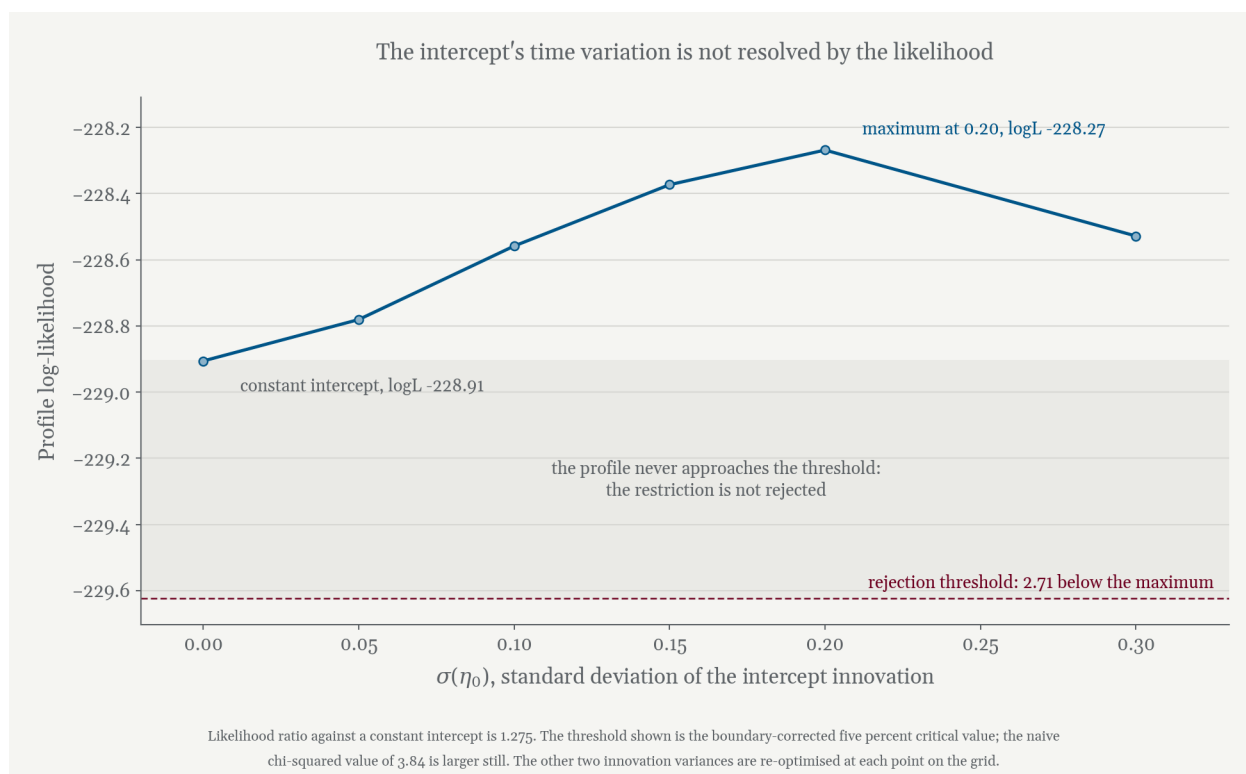


Figure 3: Profile log-likelihood for Model A over $\sigma(\eta_0)$ on the staff output gap. The horizontal axis is $\sigma(\eta_0)$ in percentage points, the vertical axis the profile log-likelihood, with the other innovation variances re-optimized at each grid point. The dashed maroon rule sits 2.71/2 log-likelihood units below the maximum, the boundary-corrected five percent threshold; the shaded band marks the region the profile never enters. The maximum lies at 0.20 and the restriction at zero. The grid is evaluated at intervals of 0.05, so its peak sits marginally below the optimized maximum; the likelihood ratio against the restriction is 1.29, with a bootstrap p -value of 0.073. The figure carries a single series and therefore no legend. Source: the verification harness, which re-estimates Model A from the data at each grid point.

The restriction is not rejected, at a bootstrap p -value of 0.073, and the ninety percent profile interval runs from zero to about 0.33. But Table 7 shows this is a property of the regressor rather than of the sample.

Table 7: Likelihood ratios against a constant intercept, by output gap and sample. Because the null sets a variance to its boundary, p -values are from a parametric bootstrap of the null with 178 replications under the seed recorded in the catalogue accompanying the estimation code; the bootstrap five percent critical values are reported beside them. The 2.71 that a 50:50 chi-squared mixture would imply is conservative here, and the two gaps generate different null distributions.

Sample and output gap	n	LR (B_0)	boot. p	5% cv	LR (B_1)
1987:Q3–2024:Q4, staff	150	1.29	0.073	2.04	12.08
1987:Q3–2024:Q4, mechanical	149	3.42	0.000	2.08	13.98
1965:Q4–2024:Q4, mechanical	236	3.62	—	—	35.64
1965:Q4–1987:Q2, mechanical	87	0.00	—	—	7.58

The first two rows differ only in the output gap. Over the same quarters, the intercept's drift is not resolvable under the staff's gap ($p = 0.073$) and is resolvable under the mechanical one ($p < 0.001$). The judgmental gap absorbs low-frequency variation that a mechanical trend leaves in the intercept. The implication is that no intercept path estimated in this way should be read as a measurement of the natural rate without reference to the gap it was estimated against.

Two properties of the bootstrap distributions bear on the conduct of such tests. Their ninetieth percentiles lie near unity and their ninety-ninth near four to five, with substantial mass at zero, as the boundary problem implies; the distribution is not the 50:50 chi-squared mixture often assumed. The two gaps also generate different null distributions, so a single tabulated critical value applied to both would misstate at least one.

5.3 The Inflation Coefficient

Figure 4 plots the path over six decades. The same test applied to B_1 rejects in every specification, at likelihood ratios from 7.58 to 35.64. The margin is decisive under the bootstrap as well: the null distribution for the staff gap has a ninety-ninth percentile of 7.00, which the smallest of these statistics exceeds.

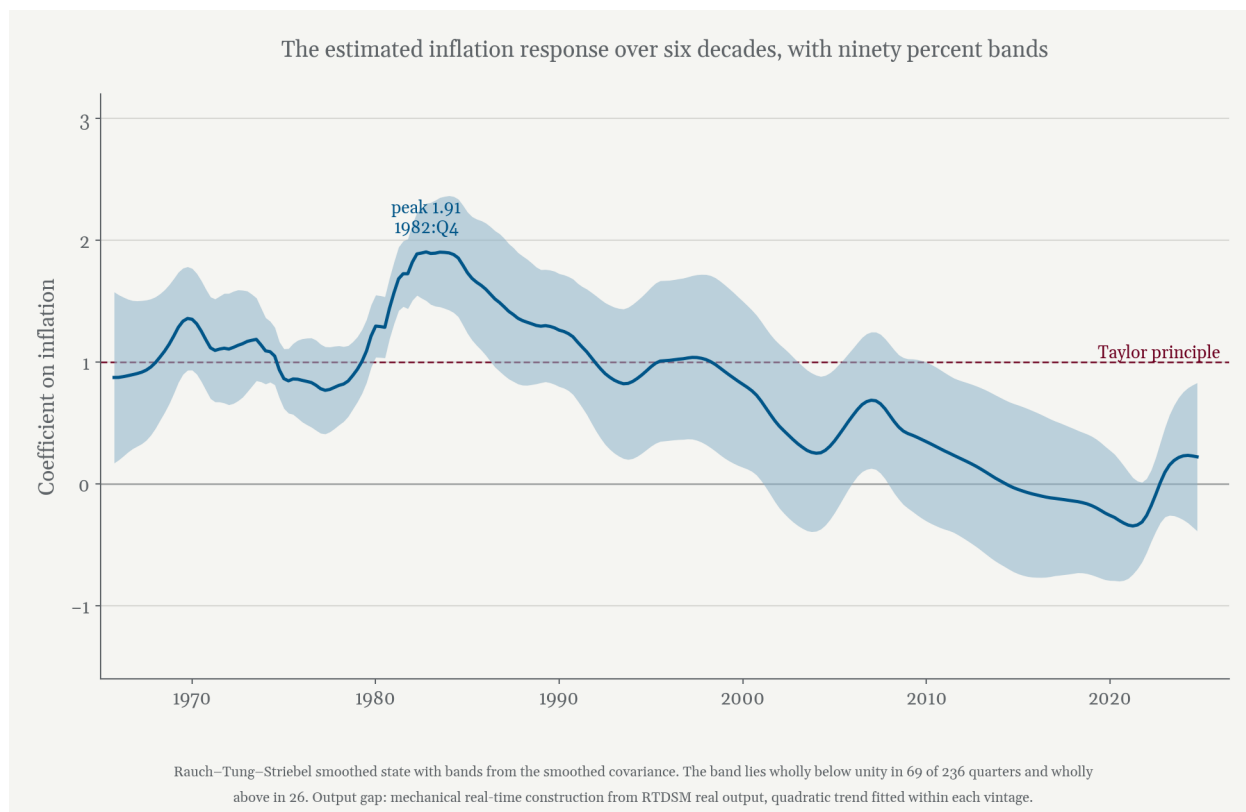


Figure 4: The smoothed coefficient on inflation over 1965:Q4–2024:Q4 with ninety percent bands from the smoothed covariance, on the mechanical output gap. The horizontal axis is calendar time at quarterly frequency; the vertical axis is the coefficient. The dashed maroon rule marks unity, the Taylor principle threshold. The band lies wholly below unity in 69 of 236 quarters and wholly above in 26. The figure carries one series and its band, and therefore no legend. The output gap is the mechanical construction of §2.1: for each RTDSM vintage of real output, a quadratic trend fitted to log output on that vintage’s own observations, evaluated at its terminal quarter. Source: the verification harness.

Over six decades the path traces a familiar sequence: 1.35 in 1970, a peak of 1.905 at 1982:Q4, 0.82 by 2000, and values below zero after 2015. Two features warrant attention. The disinflation-era peak is large. And the pre-Volcker coefficient lies above unity rather than below it, reproducing on an independently constructed real-time gap the reversal Orphanides identified.

The shortfall below unity after 2008 is statistically distinguishable, but its span begins at the onset of the zero lower bound, where the funds rate could not respond to inflation whatever the Committee preferred. Part of the estimated shortfall is mechanical rather than behavioral.

5.4 Backward and Forward Looking

M5 leads Table 3 by 0.135 percentage points, a difference the Diebold–Mariano test cannot distinguish from zero. The two accounts are not nested, so no restriction test separates them as Proposition 1 separates inertia from omitted shocks. Identification would require variation in forecast

errors independent of the information the Committee acted on. This design supplies none, and the dimension is treated as unresolved.

6 The Normative Counterpart

The positive exercise asks which rule was followed. The normative asks what a different rule would have achieved, and the answer turns on the strength of the transmission from the policy rate to inflation.

6.1 A Three-Equation Model

A Phillips curve, an IS curve, and Okun's law are estimated over 2012:Q3–2024:Q4—fifty quarters beginning with the formal adoption of the two percent target. The unemployment gap is taken against a constant $u^* = 4.4\%$ and a COVID indicator spans 2020:Q2–2022:Q2. All three are estimated on persistent quarterly series, so the Newey–West column of Table 8 governs inference.

Table 8: Structural model, 2012:Q3–2024:Q4 (50 quarters). Standard errors in parentheses; SE (HAC) uses a Bartlett kernel at $\lfloor 4(T/100)^{2/9} \rfloor = 3$ lags. The policy gap is $i_t - \pi_t - r_t^*$.

Equation	Coefficient	Estimate	SE (OLS)	SE (HAC)	t (HAC)
Phillips curve $R^2 = 0.9504$	Constant	0.0764	0.1092	0.1198	0.64
	π_{t-1}	1.2486	0.1558	0.1884	6.63
	π_{t-2}	−0.3048	0.1473	0.1509	−2.02
	$u_{t-1} - u^*$	0.0152	0.0299	0.0413	0.37
	COVID	0.3240	0.1552	0.3456	0.94
IS curve $R^2 = 0.6121$	Constant	0.1934	0.3038	0.2549	0.76
	y_{t-1}	0.4968	0.1428	0.1792	2.77
	y_{t-2}	0.2445	0.1398	0.1255	1.95
	Policy gap $_{t-1}$	0.0029	0.1495	0.1300	0.02
	COVID	−1.0787	0.6150	0.9849	−1.10
Okun's law $R^2 = 0.7445$	Constant	0.2128	0.1469	0.2406	0.88
	$(u - u^*)_{t-1}$	0.2854	0.0886	0.1875	1.52
	y_t	−0.5685	0.0781	0.2328	−2.44

Three coefficients are distinguishable from zero: inflation persistence at both lags and Okun's output term. The two carrying policy to inflation are not. Multiplying the three links gives -0.000025 per quarter. The delta method is unreliable for this quantity, since it linearises at a point where two of the three coefficients are near zero; simulating from the three sampling distributions gives a ninety-five percent interval of $[-0.0076, +0.0075]$, some 3.4 times wider than the delta-method interval of $[-0.0022, +0.0022]$. The simulated interval is reported throughout. It excludes transmission of economic consequence: the simulated probability that the chain exceeds 0.05 in absolute value—a one-point change in the policy rate moving inflation by a fifth of a point over a year—is below 0.0001.

Adding the ten-year Personal Consumption Expenditures inflation expectation from the Survey of Professional Forecasters does not rescue the channel. Its coefficient is 1.2768 with a standard error of 0.8044, of the right sign and magnitude for an accelerationist curve but not resolved, and its inclusion drives the slack coefficient to 0.0010. The series varies little—standard deviation 0.140 over 2012–2024—so anchored expectations supply almost no identifying variation, which is itself why the channel cannot be measured here.

6.2 Instrumenting the Policy Gap

The policy gap is set by a Committee responding to output, so ordinary least squares is biased toward zero exactly where the paper concludes there is nothing. High-frequency policy surprises, measured in narrow windows around Federal Open Market Committee announcements, are plausibly orthogonal to contemporaneous output and provide an instrument. Table 9 reports the estimates and their diagnostics.

Table 9: Instrumental-variables estimates of the IS curve’s policy gap coefficient. F is the first-stage statistic on the excluded instruments, with ten the conventional threshold. In the levels specification instruments enter contemporaneously and cumulated; the differenced specification uses contemporaneous instruments only, avoiding the nonstationary cumulated term. NS denotes the Nakamura–Steinsson policy news shock; USMPD the current-target surprise from the United States Monetary Policy Event-Study Database. Hansen J tests the over-identifying restrictions where the specification is over-identified.

Instrument, window, form	n	F	OLS	2SLS	Hansen J	p
NS, 1995–2014, levels	77	18.18	0.0640 (0.0626)	0.0502 (0.0666)	0.312	0.577
USMPD, 1995–2014, levels	77	6.25	0.0640 (0.0626)	0.3419 (0.1368)	—	—
Both, 1995–2014, levels	77	15.76	0.0640 (0.0626)	0.1221 (0.0734)	5.390	0.145
Both, 1995–2014, differences	76	12.69	−0.1243 (0.1494)	0.2194 (0.3230)	0.002	0.965
USMPD, 1994–2024, levels	122	3.00	−0.0113 (0.0689)	−0.0919 (0.1578)	—	—

The Nakamura–Steinsson shock is a strong instrument at $F = 18.18$, and under it the two-stage estimate is 0.0502 with a standard error of 0.0666—indistinguishable from zero and within one-fifth of a standard error of the ordinary least squares estimate.¹¹ The near-zero coefficient is therefore not a simultaneity artifact.

The second row is not contradictory. The current-target surprise alone is weak at $F = 6.25$, and the estimate of 0.3419 it produces has the properties associated with weak instruments; the first-stage statistic is the reason it is not reported as evidence.

¹¹Emi Nakamura and Jón Steinsson, “High-Frequency Identification of Monetary Non-Neutrality: The Information Effect,” *Quarterly Journal of Economics* 133, no. 3 (2018): 1283–1330.

Two diagnostics support the specification. Hansen's test does not reject the over-identifying restrictions in any specification, at $p = 0.577$ on the preferred row. Because the cumulated instrument is nonstationary by construction, the fourth row re-estimates in first differences with contemporaneous instruments only; the instrument remains strong at $F = 12.69$ and the coefficient is 0.2194 with a standard error of 0.3230, again indistinguishable from zero. The point estimates vary across specifications, from 0.050 in levels to 0.219 in differences, but each standard error encompasses the full range, and no specification rejects the hypothesis of no transmission.

6.3 Optimized Rules

The reaction function is optimized to minimize the root mean squared deviation of Core Personal Consumption Expenditures inflation from two percent, with the estimated structural residuals held fixed across counterfactuals and the zero lower bound imposed. Under actual policy the simulation reproduces observed inflation to machine precision, which verifies the recursion; the resulting statistic is 1.35375.

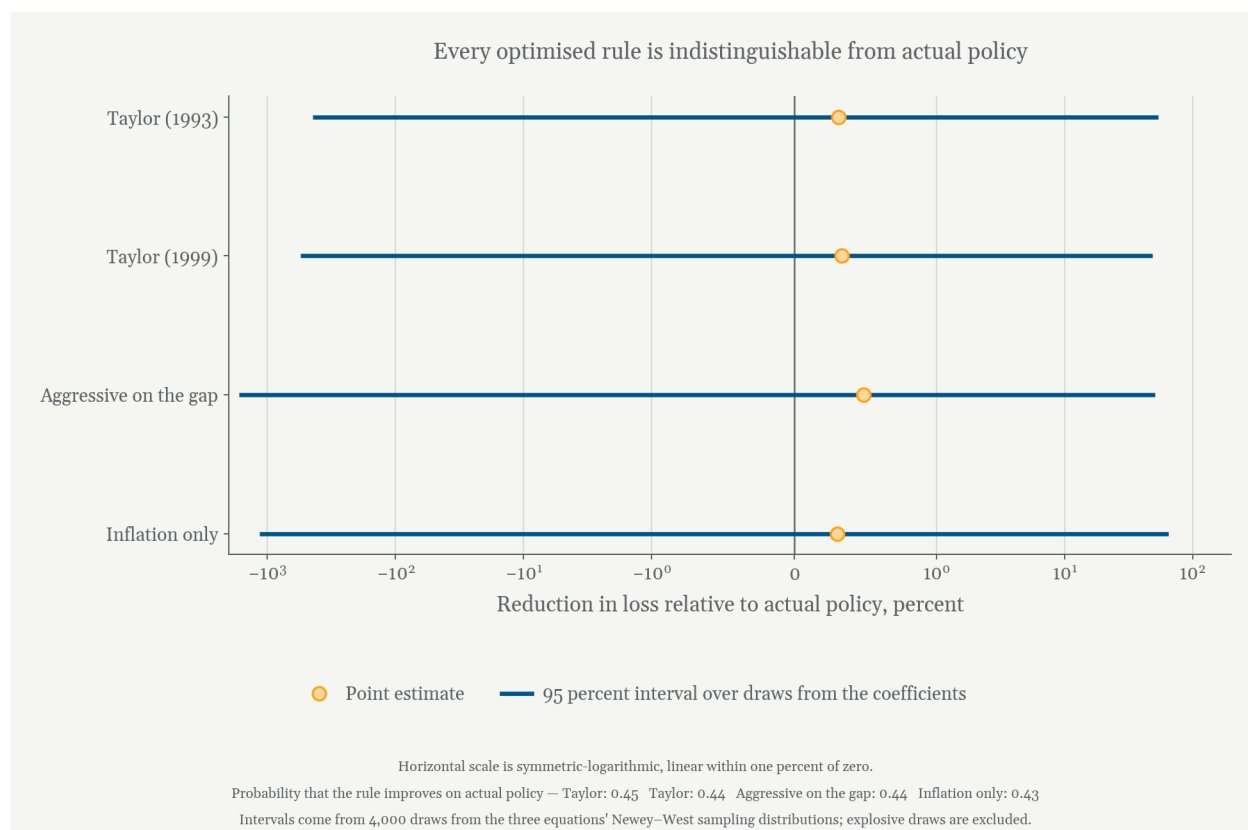


Figure 5: Reduction in loss relative to actual policy for four optimized rules, at the point estimates and over draws from the coefficients' sampling distributions. The vertical axis lists the rules; the horizontal axis is the percentage reduction in mean squared deviation of Core PCE inflation from two percent, on a symmetric-logarithmic scale linear within one percent of zero. The amber marker is the point estimate, the deep-blue bar the ninety-five percent interval over 4,000 draws from the Newey–West sampling distributions, with explosive draws excluded. Marker and bar are identified in a legend below the panel. Each draw re-simulates the three-equation system forward from 2012:Q3 with the estimated structural residuals held fixed and the zero lower bound imposed. Source: the verification harness and the seed recorded in the accompanying catalogue.

Figure 5 sets the point estimates against the intervals implied by coefficient uncertainty. At the point estimates every rule improves on actual policy, by between 0.30 and 0.49 percent of the loss. Carrying coefficient uncertainty through, the intervals span from roughly $-1,700$ to $+65$ percent and the probability that a given rule improves lies between 0.435 and 0.450. The exercise has no power to rank the rules, because the mechanism through which any would operate is measured at zero.

6.4 What the Chain Does Not Establish

The chain runs from the policy rate to the output gap, from the gap to unemployment, and from unemployment to inflation. Its first link is instrumented in §6.2; its last is not. The Phillips slope is estimated by ordinary least squares on aggregate data, and it carries the wrong sign for a Phillips

curve—slack raises inflation at 0.0152—with a standard error of 0.0413 that admits curves of economic consequence in the conventional direction. Cross-sectional designs identify the slope precisely because aggregate regressions conflate it with movements in expectations and with the policy response to both.¹²

The claim of this section is therefore about measurement, not about structure. The transmission from the policy rate to inflation cannot be detected on this sample; whether it is absent is a different question, and one this design does not answer. That distinction is the paper’s own argument applied to itself, and it sharpens the normative conclusion rather than softening it: the exercise cannot rank rules because the mechanism through which they would operate is unidentified.

One channel is absent by construction. The model carries no expectations block beyond the term discussed above, so an alternative rule changes the funds rate path without changing expectations. To the extent policy reaches inflation by moving expected inflation rather than slack, these counterfactuals understate what a different rule could achieve.

7 Robustness

Three of the paper’s inputs are choices rather than data: the detrending rule behind the mechanical gap, the split of variation between drift and noise in the state-space models, and the reference distributions used for inference. This section reports what each choice is worth.

The detrending method. The mechanical gap is a construction of this paper, so the restriction tests are re-run across the five detrending rules of Table 10.

Table 10: ADL restriction tests across detrending methods, 1965:Q4–2024:Q4. Each gap is built by fitting the named trend to log real output within every RTDSM vintage and taking the deviation at that vintage’s terminal quarter. Bootstrap five percent critical values are approximately 9.1 for partial adjustment and 7.8 for the common factor.

Trend	n	ρ	PA Wald	CF Wald
Linear	236	0.9120	5.517	3.509
Quadratic	236	0.8971	3.228	20.015
Cubic	236	0.9060	4.583	17.097
Hodrick–Prescott, $\lambda = 1600$	236	0.9420	1.912	28.261
Hodrick–Prescott, $\lambda = 6400$	236	0.9316	1.631	37.831

Partial adjustment survives under all five, so that conclusion is invariant to the detrending rule. The common-factor rejection holds under four but not under a linear trend, where the statistic of 3.509 carries a bootstrap p -value of 0.283. Under a linear trend, therefore, neither restriction is rejected and the data do not discriminate. A linear trend is a poor description of six decades of log output, and the four specifications that permit curvature agree; the exception is nonetheless recorded.

¹²Jonathon Hazell, Juan Herreño, Emi Nakamura, and Jón Steinsson, “The Slope of the Phillips Curve: Evidence from U.S. States,” *Quarterly Journal of Economics* 137, no. 3 (2022): 1299–1344.

The spliced quarters. The staff gap changes source at 2021:Q1. Estimating on Greenbook data alone, to 2020:Q4, leaves partial adjustment near its full-sample value at 9.793 but moves the common factor from 10.384 to 5.353, below even the asymptotic threshold. The short-sample finding that both restrictions are rejected therefore depends in part on the sixteen spliced quarters, and is reported as conditional on them.

The transmission chain's interval. The chain is a product of three estimated coefficients, two of which are near zero, so the delta method understates its dispersion: the simulated ninety-five percent interval is 3.4 times the delta-method interval. Simulation is used throughout, and the qualitative conclusion is unaffected, since the simulated probability that the chain exceeds 0.05 in absolute value remains below 0.0001.

The variance allocation. Fixing R at 40% of the least squares residual variance is a judgment. Across 20, 40, 60, and 80 percent, $\sigma(\eta_0)$ falls from 0.2555 to 0.1520 and the intercept path's amplitude compresses from 2.61 to 1.26 percentage points, while the inflation coefficient's minimum sits between -0.30 and -0.31 at every setting. The intercept results depend on the allocation; the inflation coefficient results do not.

Asymptotic critical values. Neither test in this paper can rely on its textbook distribution, and the two fail in opposite directions. For the ADL Walds, a lagged dependent variable at ρ near 0.94 makes the asymptotic test *liberal*: bootstrap five percent critical values run from 7.84 to 10.28 against χ^2_2 's 5.99. For the profile likelihood ratio, the boundary makes a 50:50 chi-squared mixture *conservative*: bootstrap critical values are 2.17 and 1.67 against the 2.71 that mixture implies. Both sets of tests are therefore reported on bootstrap p -values throughout.

The two output gaps. They correlate 0.659 with a mean absolute difference of 2.23 points over their 149 common quarters, confirming that statistical detrending in real time does not reproduce a judgmental gap. This paper treats the divergence as informative rather than as a nuisance: because the two gaps disagree while the sample is held fixed, Table 7 can attribute the intercept result to the regressor rather than to the span.

The forecast series. §3.2 substitutes the Federal Reserve staff's own projection for the private survey in M5 and finds the fit improves despite the staff series being more variable.

8 Limitations

The staff output gap changes source at 2021:Q1, joining two concepts of potential output with a level step of 1.53 percentage points; the final sixteen quarters rest on the spliced segment. The Philadelphia Fed states that the staff series was never published in the Tealbook and that errors may remain. The mechanical gap is a construction of this paper and reproduces no published series. The inflation target before 2012 has no observable counterpart and is an assumption. The natural rate is not independent of the funds rate, so §5.2 compares two objects sharing an input. Structural parameters are held fixed across counterfactuals, and the normative half rests on fifty quarters. The instrumental-variables result of §6.2 covers 1995–2014, the span of the Nakamura–Steinsson

series, and does not extend to the recent tightening; the Phillips slope is not instrumented at all, so the transmission chain rests on one identified link and one that is not. Two results are conditional in ways §7 documents: the common-factor rejection on the short sample fails under a linear trend and falls from 10.384 to 5.353 when the spliced quarters are dropped, and the mechanical gap is this paper's own construction rather than a published series.

9 Conclusion

The central finding of this paper is that what a Taylor Rule estimation identifies is a property of the output gap it is given and of the span over which it is run, rather than a fixed feature of the estimation problem. Partial adjustment and a serially correlated policy shock are nested restrictions on a single autoregressive distributed lag model. Over thirty-seven years with the Federal Reserve staff's judgmental gap, both are rejected, at bootstrap p -values of 0.048 and 0.036, and neither account is preferred. Over six decades with a mechanically constructed gap, partial adjustment survives at 0.281 while the common-factor restriction is rejected at below 0.001, a pattern that holds under four of the five detrending rules examined. The drift in the policy intercept is likewise not resolvable under the staff's gap, at $p = 0.073$, and is resolvable under the mechanical gap over the same quarters, at $p < 0.001$.

Fit rankings do not settle these questions. No pairwise difference among the seven calibrated specifications is statistically significant, including a gap of 0.606 percentage points in root mean squared deviation, and four different specifications lead across five episodes, with the full-sample runners-up ranking last over 1987–2007.

Two dimensions do resolve. Interest rate smoothing belongs in the specification, by a log-likelihood margin near 180 that no parameter-count convention disturbs. The estimated response to inflation drifts, at likelihood ratios from 7.58 to 35.64 against a bootstrap ninety-ninth percentile of 7.00, and over six decades traces a peak of 1.905 at 1982:Q4 with pre-Volcker values above unity.

The normative counterpart yields a sharper result than the positive one. The transmission from the policy rate to inflation is estimated at -0.000025 per quarter, with a simulated ninety-five percent interval of $[-0.0076, +0.0075]$, and the near-zero first link survives instrumentation by high-frequency policy surprises at a first-stage F of 18.18. The first link of the chain is therefore identified; the last is not, so the finding is that the transmission cannot be detected on this sample rather than that it is absent. Either way no optimized rule can be distinguished from actual policy once coefficient uncertainty is carried through.

Finally, the dataset is itself a result. Mapping a vintage archive to one observation per quarter requires a rule, and four natural rules applied to the same output-gap release differ by 0.245 percentage points on average and by 1.360 at the extreme—larger than every pairwise difference in fit the paper tests. A quantitative claim about monetary policy is therefore a claim about the extraction rule no less than about the economy, and the rule belongs in the record.

Appendix A Verification

An executable harness recomputes the paper's printed quantities from primary state and compares each against the value printed, at the precision at which it is printed; it exits nonzero on any discrepancy and is run as part of the build. It re-estimates the three state-space models from the data rather than reading stored output, replicates the seven rule formulas and their tests of equal accuracy, re-runs the restriction tests and the structural regressions with their Newey–West standard errors, re-estimates the instrumental-variables specifications, and re-simulates the transmission chain's interval. It verifies 165 quantities.

Seven groups of quantities the harness does not recompute are listed in its output with the record carrying each. They are the bootstrap p -values and critical values of Tables 6 and 7, which rest on stored draws under a fixed seed; the profile likelihood ratios of Table 7; three rows of Table 10; the mean absolute differences of Table 1; the extraction tests of Table A.1; and the counterfactual loss statistics of §6.3. External releases are read as primary state and are not re-derived.

Stochastic results are generated under the seeds recorded in the accompanying catalogue, so the bootstrap distributions regenerate rather than persisting. The profile likelihood bootstraps of Table 7 use 178 replications each.

Extraction rules were tested rather than assumed, with the outcome in Table A.1. Ten candidates were compared for the staff output gap and three for the deflator. The comparison establishes the magnitude of the choice rather than the correctness of any one rule: across the four output-gap rules shown, the mean pairwise absolute difference is 0.245 percentage points and the average spread 0.461, reaching 1.360 in a single quarter.

Where two independent releases covered the same quarters their agreement was measured: 0.031 mean absolute difference at a correlation of 0.99984, which validates the extraction independently. Vintage lags were verified rather than assumed—one quarter for 240 of 243 deflator vintages and 238 of 240 unemployment quarters.

Table A.1: Candidate extraction rules and the dispersion among them. Each maps the same source release to one observation per quarter. The final column reports the mean absolute difference between the rules of that family, computed over their common quarters; it measures how much the choice of rule moves the series, not the correctness of any one rule. The rule used in this paper is italicised.

Candidate rule	n	Dispersion
<i>Output gap: four rules over 72 common quarters</i>		
<i>Expanded release, last Greenbook published in q</i>	100	
Expanded release, last Greenbook published in q , read at $q - 1$	100	
Previous format, last Greenbook published in q	106	
Previous format, last Greenbook with $T = q + 1$, read at $T - 1$	106	
mean pairwise absolute difference		0.245
mean spread across the four		0.461
largest spread in a single quarter		1.360
<i>GDP deflator inflation: three constructions</i>		
<i>Within the first vintage, four-quarter change</i>	240	
Chained first-release quarterly growth over four quarters	240	
Vintage current during q , four-quarter change at $q - 1$	236	

Bibliography

- Boivin, Jean. “Has U.S. Monetary Policy Changed? Evidence from Drifting Coefficients and Real-Time Data.” *Journal of Money, Credit and Banking* 38, no. 5 (2006): 1149–1173.
- Clarida, Richard, Jordi Galí, and Mark Gertler. “Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory.” *Quarterly Journal of Economics* 115, no. 1 (2000): 147–180.
- Cogley, Timothy, and Thomas J. Sargent. “Drifts and Volatilities: Monetary Policies and Outcomes in the Post WWII U.S.” *Review of Economic Dynamics* 8, no. 2 (2005): 262–302.
- Diebold, Francis X., and Roberto S. Mariano. “Comparing Predictive Accuracy.” *Journal of Business and Economic Statistics* 13, no. 3 (1995): 253–263.
- Federal Reserve Bank of New York. *Holston–Laubach–Williams Real-Time Estimates*. Data file. New York: Federal Reserve Bank of New York, 2026.
- Federal Reserve Bank of Philadelphia. *Output Gap and Financial Assumptions from the Board of Governors Tealbook/Greenbook*. Data set. Philadelphia: Federal Reserve Bank of Philadelphia, 2026.
- Federal Reserve Bank of Philadelphia. *Real-Time Data Set for Macroeconomists*. Data set. Philadelphia: Federal Reserve Bank of Philadelphia, 2026.
- Federal Reserve Bank of Philadelphia. *Survey of Professional Forecasters*. Data set. Philadelphia: Federal Reserve Bank of Philadelphia, 2026.

- Federal Reserve Bank of San Francisco. *United States Monetary Policy Event-Study Database*. Data set. San Francisco: Center for Monetary Research, 2026.
- Hazell, Jonathon, Juan Herreño, Emi Nakamura, and Jón Steinsson. “The Slope of the Phillips Curve: Evidence from U.S. States.” *Quarterly Journal of Economics* 137, no. 3 (2022): 1299–1344.
- Holston, Kathryn, Thomas Laubach, and John C. Williams. *Measuring the Natural Rate of Interest after COVID-19*. Federal Reserve Bank of New York Staff Reports, no. 1063. New York: Federal Reserve Bank of New York, 2023.
- Holston, Kathryn, Thomas Laubach, and John C. Williams. “Measuring the Natural Rate of Interest: International Trends and Determinants.” *Journal of International Economics* 108 (2017): S59–S75.
- Laubach, Thomas, and John C. Williams. “Measuring the Natural Rate of Interest.” *Review of Economics and Statistics* 85, no. 4 (2003): 1063–1070.
- Nakamura, Emi, and Jón Steinsson. “High-Frequency Identification of Monetary Non-Neutrality: The Information Effect.” *Quarterly Journal of Economics* 133, no. 3 (2018): 1283–1330.
- Orphanides, Athanasios. “Monetary Policy Rules Based on Real-Time Data.” *American Economic Review* 91, no. 4 (2001): 964–985.
- Orphanides, Athanasios, and Simon van Norden. “The Unreliability of Output-Gap Estimates in Real Time.” *Review of Economics and Statistics* 84, no. 4 (2002): 569–583.
- Rudebusch, Glenn D. “Term Structure Evidence on Interest Rate Smoothing and Monetary Policy Inertia.” *Journal of Monetary Economics* 49, no. 6 (2002): 1161–1187.
- Stock, James H., and Mark W. Watson. “Median Unbiased Estimation of Coefficient Variance in a Time-Varying Parameter Model.” *Journal of the American Statistical Association* 93, no. 441 (1998): 349–358.
- Taylor, John B. “Discretion versus Policy Rules in Practice.” *Carnegie-Rochester Conference Series on Public Policy* 39 (1993): 195–214.