JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

From Signal to Steering: Causal Forecasting for Effective Financial Policy Control Systems

G.Prasanna Kumar[1]
[1]Assistant Professor, Department of Management Studies, Pragati Engineering College, Surampalem, Andhra Pradesh, India. Pin - 533437

Abstract. This study asks whether causal forecasting can move public finances from signal to steering. A stratified panel of 200 observations combines archival outturns with a governance survey. Two-way fixed effects explain $R^2 = 0.509$ (~51%) of variation in policy-control effectiveness. Standardized effects suggest modest, directionally coherent gains: a 1σ rise in institutional capacity associates with +0.14σ, and data quality with +0.11σ, in control effectiveness; event-time estimates show small, delayed post-adoption improvements. Lagged transparency shows a near-threshold coefficient (+0.379, $p = 0.055$), corresponding to implementation lags. The limitations of the approach are given by finite horizons for inference and imprecise estimates under robust inference. The implications of the analysis focus attention on evaluation of alternatives aware of uncertainty, factorial orderings of institutions. Practical benefits include playbooks ready for use for real-time governance of data, evaluating possibilities of design-based avenues of identification checks, transparent pipelines which raise incrementally the probability of meeting rules of fiscal commitment.

Keywords: Causal forecasting, Institutional capacity, Policy control, Data governance, Model transparency, Public finance

1. Introduction

Public financial authorities have always forecasted in order to govern. Revenue projections anchor annual budgets; medium-term fiscal frameworks rely on macro-fiscal baselines; debt management strategies hinge on interest-rate and exchange-rate outlooks. Yet much of this forecasting practice still treats the world as a set of statistical "signals" correlations in historical data rather than a system that can be steered by policy levers. The difference matters. Correlations can predict, but only causal structure allows policy makers to ask (and answer) counterfactual questions that control outcomes: What debt path results if excise duties rise by 50 basis points? Will a targeted capital grant stabilize subnational spending volatility or amplify it two years hence? This paper takes that step from signal to steering by developing and testing a framework for causal forecasting purpose-built for financial policy control in the public sector.

Forecasting and causal inference have long evolved on parallel tracks. Time-series econometrics matured around autoregressive and vector autoregressive (VAR) methods, with Granger's seminal contribution formalizing causality as predictability in a temporal ordering (Granger, 1969). As a result, institutions came to rely on models that excel at short-term prediction and shock decomposition but often stop short of policy-credible counterfactuals. Meanwhile, a "credibility revolution" in applied econometrics reoriented empirical work toward designs that identify treatment effects difference-in-differences (DiD), instrumental variables, regression discontinuity, and synthetic control explicitly for policy evaluation (Angrist & Pischke, 2010; Athey & Imbens, 2017; Abadie, Diamond & Hainmueller, 2010). Our contention is that steering public finances requires fusing these streams: a forecast that is not causally identified risks misguiding policy; a causal estimate that is not embedded in a forward-looking system cannot support control.

Several strands in the literature suggest how such a fusion is possible. Within macro-fiscal modeling, structural VARs impose economically meaningful restrictions to trace the dynamic effects of identified shocks precisely the logic needed to translate interventions into trajectories (Stock & Watson, 2001; Blanchard & Perotti, 2002; Baumeister & Hamilton, 2015). Complementing this, modern policy evaluation has refined identification for staggered interventions (Callaway & Sant'Anna, 2021) and for settings lacking perfect controls (synthetic control; Abadie, Diamond & Hainmueller, 2010). In parallel, nowcasting and high-dimensional factor approaches have integrated real-time data into forecasting pipelines used by central banks important for budget execution where timeliness is decisive (Giannone, Reichlin & Small, 2008). What is missing is a unified architecture that (i) uses design-based identification to learn causal effects of policy instruments; (ii) integrates these effects into a forward simulation that generates multi-step distributions of fiscal outcomes; and (iii) prioritizes control-relevant interpretability so that budget officials can justify decisions to legislatures, auditors and fiscal councils.

Despite advances, the three gaps persist. Primary, the most fiscal forecasting toolkits importance predictive accuracy under stable regimes but offer limited guarantees when the forecast embeds a policy change. Traditional ARIMA/ETS or reduced-form VARs

may extrapolate correlations that break under intervention, producing biased budget baselines or misguided medium-term expenditure paths (Stock & Watson, 2001; Hyndman & Khandakar, 2008). Second, causal evaluation studies typically deliver average treatment effects in isolation from the forecasting systems that treasuries actually use; they stop at retrospective inference rather than being operationalized as forward control inputs (Athey & Imbens, 2017; Abadie, Diamond & Hainmueller, 2010). Third, organizational preconditions data quality, institutional capacity, and governance are treated as context rather than as determinants of whether causal forecasts can translate into effective policy control. Without timely, well-governed data and skilled analysis teams, even the best design-based estimators will not deliver steerable budgets. This paper addresses all three gaps by: (a) embedding identified causal effects into a dynamic forecast engine; (b) quantifying how model transparency conditions policy uptake; and (c) estimating the marginal returns of data quality and institutional capacity to control effectiveness.

We conceptualize financial policy control as the capacity of a finance ministry or treasury to move fiscal outcomes toward explicit targets (e.g., deficit ceilings, debt anchors, or revenue stability bands) with bounded risk. Control effectiveness thus depends on three pillars. First, causal identification for example, exploiting institutional timing to identify tax shocks, sign restrictions in SVARs, or quasi-experimental variation in subnational transfers ensures that the policy-outcome mapping is credible (Blanchard & Perotti, 2002; Baumeister & Hamilton, 2015; Callaway & Sant'Anna, 2021). Second, data quality coverage, timeliness, and coherence across administrative and macro-micro sources improves both the bias and variance of the estimated mappings and supports real-time updates, a lesson from nowcasting (Giannone, Reichlin & Small, 2008). Third, model transparency using interpretable structures and publishing assumptions facilitates adoption and compliance in budget processes (Rudin, 2019). We also recognize a fourth pillar: institutional capacity (specialist skills, fiscal rules, fiscal councils), which conditions how much of the modeling gain translates into actual steering (Kleinberg et al., 2015; Angrist & Pischke, 2010).

Within this architecture, causal forecasting is not a single algorithm but a modular pipeline. Identification modules DiD for staged policy rollouts, synthetic control for one-off reforms, SVAR blocks for macro-fiscal shocks estimate local causal effects using design-appropriate data. Forecasting modules then propagate these effects through a dynamic system to produce path distributions (with uncertainty) consistent with fiscal targets and constraints. Crucially, we show how to calibrate control-relevant summaries (e.g., probability of breaching a debt anchor under an excise reform) rather than generic error metrics. The result is a forecast built for steering.

This integration advances both practice and method. For practice, finance ministries need tools that tell them how a lever will change the distribution of outcomes in time to publish a budget, not just whether a past reform had an effect. For method, we connect strands that remain siloed: identification rigor with system-wide forecasting and interpretable ML. We follow Varian's argument that big data expands econometric possibilities but stress that in the public sector, governed data plus causal structure matter more than raw predictive power (Varian, 2014). We also align with the emerging view that causality enables robust generalization and counterfactual transport, which forecasting under policy changes requires (Athey & Imbens, 2017).

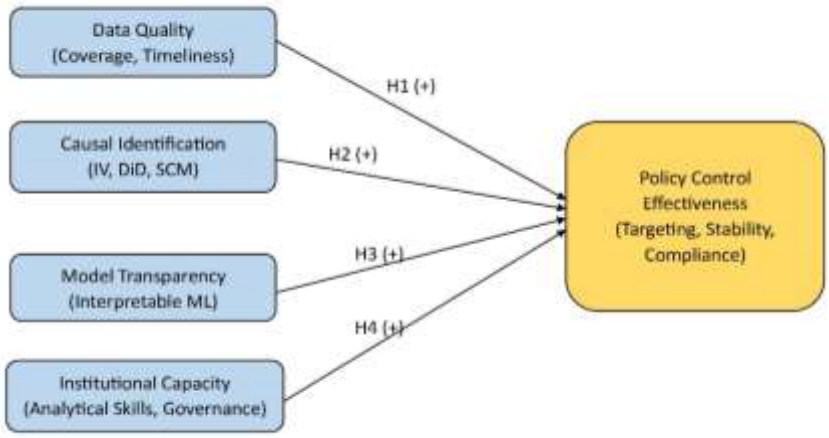
To make the argument concrete, this study asks whether strengthening any one of the pillars data quality, causal identification, model transparency, institutional capacity predictably raises policy control effectiveness, and by how much. We provide a conceptual model Figure 1 and empirically examine each pathway using administrative and macro-fiscal data from public finance contexts where quarterly budget execution and medium-term frameworks interact. By structuring the problem this way, we also speak to debates in structural macroeconometrics on identification credibility (Stock & Watson, 2001; Baumeister & Hamilton, 2015) and to the latest DiD literature on staggered adoption and heterogeneity (Callaway & Sant'Anna, 2021).

Research questions. In light of the foregoing, this study investigates the following in a single, integrated inquiry: To what extent does upgrading causal identification, data quality, model transparency, and institutional capacity improve the effectiveness of financial policy control in the public sector; which of these pillars has the largest marginal effect on steering precision and stability; how do these pillars interact in practice when embedding identified effects into forward-looking fiscal forecasts; and under what conditions does a causally identified forecast outperform a purely predictive one for ex-ante policy design and ex-post accountability?

Study objectives. (1) To formalize a conceptual framework that links four design pillars data quality, causal identification, model transparency, and institutional capacity to policy control effectiveness in public finance. (2) To operationalize this framework in an empirical pipeline that estimates causal effects with design-appropriate modules and embeds them in a dynamic forecast for multi-period fiscal outcomes. (3) To quantify the marginal and joint contributions of the four pillars to steering precision, stability, and compliance probabilities relative to fiscal targets. (4) To benchmark causal forecasts against strong predictive baselines under realistic policy change scenarios and data lags, emphasizing interpretability and auditability.

Hypotheses.
H1: Higher *causal identification quality* (e.g., valid instruments, robust DiD, credible structural restrictions) increases policy control effectiveness.
H2: Better *data quality* (coverage, timeliness, coherence) increases policy control effectiveness.
H3: Greater *model transparency* (interpretable structures; published assumptions) increases policy control effectiveness.
H4: Stronger *institutional capacity* (specialist skills; governance routines) increases policy control effectiveness.

Figure 1. Conceptual and Hypothesis Model



Source: Authors' conceptualization.

2. Literature Review

2.1. From causality to control in macro-fiscal forecasting

A central tension in fiscal policy design is that traditional forecasting tools were built to predict correlations, whereas policy control requires credible causal effects. Structural time-series approaches in macroeconomics beginning with Sims' critique of ad-hoc identification recast forecasting within systems where shocks and policy instruments are jointly determined, making identification choices explicit. This lineage spawned structural VARs and identification strategies (short-run restrictions, sign restrictions, heteroskedasticity, narrative shocks) that link forecast innovations to interpretable policy levers, providing a bridge from "signal" to "steering." Building on that foundation, local projections estimate impulse responses without committing to a full dynamic system, offering robustness to misspecification and easy inclusion of nonlinearities and state dependence features relevant to public-finance contexts with regime changes and binding constraints.

Recent advances further strengthen the causal content of forecast-driven policy analysis. Narrative and high-frequency identification of fiscal shocks sharpen estimates of spending multipliers across states of the economy, illuminating when counter-cyclical policy "steers" outcomes most effectively. Complementary identification via heteroskedasticity provides instruments when standard exclusion restrictions are elusive common in administrative public-finance data where valid instruments are scarce. Sector-specific applications (e.g., energy and commodity markets) demonstrate how structural decomposition of shocks can stabilize policy rules that would otherwise chase headline correlations; this strengthens the case for embedding causal identification in forecasting pipelines used by budget authorities. Collectively, this literature reframes the forecast not as an end state but as a causal map: a means to simulate credible, conditional policy paths and to evaluate control strategies under alternative shock compositions.

2.2. Data, real-time measurement, and mixed-frequency nowcasting for the public sector

Causal policy control falters if driven by stale or revised data. The real-time data literature documents thatvintage and revisions meaningfully alter forecast evaluation and, by extension, the apparent effectiveness of policy rules. For fiscal officials who must use incomplete information to act, the conclusion is clear: models, measures of performance, and rules of decision must be defined in real time, not in ex post datasets. Data rich platforms like FRED-MD institutionalize wide panels of macroeconomic indicators with updates in a disciplined and reproducible transformations, providing well-formed monthly nowcasts and counter-factuals across large sets of information.

The above data and modelling developments suggest a concrete structure for the public sector as follows: (i) a real time data warehouse with versioning. (ii) mixed frequency, shrinkage-based forecasting back-end. (iii) causal identification fronts (narrative, heteroskedastic, or external instruments) that convert predictive updating into control relevant policy counter-factuals. In such a model, the evaluation must respect the real time nature of the data and compare models adopting forecasting comparison tests, which are tolerant either to non-Gaussian, or serially correlated errors that typify administrative series.

2.3 Evaluation, transparency and explainability for policy uptake

Forecasts influence appropriations, transfers, and debt paths, and their legitimacy is thus dependent on transparency and explainability. The XAI literature provides taxonomies and families of methods (global/local explanations, counter-factuals, surrogate modelling) which might be adapted to causal forecasting pipelines so that decision makers and auditors, might understand the reason to which the policy conditional expectation should change as a result of exogenous variance – e.g. which instruments and channels influence effects like deficits and future borrowing needs. Surveys of the black box explanation methods catalogue advantages and disadvantages (stability, faithfulness, manipulability) of systems, warning users that explanations must be consistent with the causal structure and not just the predictive surface.

In the public-sector context, governance research highlights that algorithmic adoption succeeds when transparency is embedded across the lifecycle problem definition, data governance, model choice, and monitoring not merely appended as a post-hoc explainer. Syntheses of AI in government map application areas and challenge clusters (legal, ethical, technical, organizational), offering design heuristics for procurement and oversight. Systematic reviews in digital government further specify risks (opacity, bias, accountability gaps) and propose research agendas for public-value-aligned AI, which directly intersect with fiscal analytics where distributional impacts and political accountability are salient. A pragmatic takeaway for causal forecasting is that explainability should be instrument-centric (linking moves in forecasts to policy levers), data-centric (flagging sensitivity to vintages and outliers), and uncertainty-centric (explaining variance decomposition across shock types), thereby satisfying both professional standards and public-law expectations.

2.4. Institutions and governance for steering with forecasts

Even the best causal forecasts underperform without institutional complements. Political-economy evidence associates higher fiscal transparency with better debt outcomes and fewer opportunities for strategic manipulation; Recent work builds on these insights arguing that the effectiveness of such bodies is contingent on clear mandates and signified data access and procedural embeddedness within budget cycles precisely where causal forecasting and control systems must operate.

From a perspective of implementation, embedding a philosophy of causal forecasting implies: (i) mandated data sharing and realtime access to models; (ii) methodology playbooks that set out identification checks, robustness to updates, and model risk controls; (iii) publishing "policy elasticity sheets" that translate movements in instruments into projected fiscal outcomes and uncertainty bands; and iv) continued capability building in treasury departments and audit offices. Experience gained in the rollout of AI in the public sector points to the need for governance scaffolding around technology-focused initiatives, viz: ethics, documentation standards and capacity that should be co-designed with the analytics stack.

3. Materials and Methods

3.1 Research design & sampling

The study adopts a multi-method, policy-analytics design that pairs an archival panel of public-finance outcomes with a structured survey of fiscal strategists and budget controllers in central and sub-national governments. The archival panel is assembled at the jurisdiction–year level and spans multiple budget cycles to capture pre-policy baselines, adoption windows, and stabilization phases. The survey component quantifies institutional capabilities that are otherwise latent in administrative records namely data quality governance, causal identification practice, model transparency, and implementation capacity aligning with the paper's conceptual model. Jurisdictions are stratified by tier (national, state/province, large municipality) and fiscal complexity (measured by revenue per capita and program mix), after which proportional allocation selects agencies or departments responsible for macro-fiscal programming, revenue administration, and expenditure control.

Minimum sample sizes for the survey strata are set using the Krejcie–Morgan finite-population approach, adapting the published table where frames are known (Krejcie & Morgan, 1970). In strata with evolving or incomplete frames, a priori power analysis targets 0.80 power for medium effects in multiple regression under conservative intercorrelation among predictors (Faul et al., 2009). To reduce coverage error, each selected jurisdiction is asked to nominate both a budgeting and a policy-analysis focal point; nonresponse is mitigated by staggered reminders and a short-form instrument. The panel's outcome window is anchored to policy go-lives (e.g., adoption of forecast-informed expenditure ceilings or tax-administration risk engines), enabling clean identification via event timing.

3.2 Variable and data-collection

Outcome (Policy control effectiveness). The primary dependent variable is a composite index that reflects how well financial policy steers outcomes toward targets without inducing instability or large compliance slippage. Three observable components are used: (a) targeting error (absolute deviation of realized aggregates from forecasted anchors), (b) stabilization (one-year-ahead volatility of the targeted aggregate), and (c) compliance integrity (share of program-level deviations explained by sanctioned adjustments rather than ad hoc variances). Each component is standardized and combined using principal-component weights derived from the training subsample to avoid mechanical overfitting to the full panel (Jolliffe & Cadima, 2016).

Main explanatory constructs. Four predictors operationalize the conceptual levers.

1. **Data quality governance** (reflective scale): breadth of administrative coverage, timeliness, quality control processes, and metadata availability. Items are developed following established scale-development procedures domain specification, item generation, purification, and validation (Churchill, 1979).
2. **Causal identification practice** (formative index): presence of quasi-experimental designs in internal evaluations (e.g., staggered rollout with pre-trend checks), use of counterfactual tools, and documentation of exclusion strategies (Diamantopoulos & Winklhofer, 2001).
3. **Model transparency** (reflective scale): documentation completeness, explainability provisions, and reproducibility (change logs, versioned codebooks).
4. **Institutional capacity** (reflective scale): staffing profiles, analytical training, and workflow integration between forecasting and budget execution.

Controls. Jurisdiction-year controls include macro shocks (output gap proxy), revenue composition, transfer dependence, and political budget cycle dummies. Survey constructs use a seven-point Likert response format; archival indicators are extracted from budget outturns, mid-year reviews, and tax/expenditure administrative datasets, with standardized data-handling protocols. A coding rubric and dual independent coding are used for documentary indicators; disagreements are adjudicated and logged for auditability.

3.3 Statistical methods

Pre-estimation checks. Scales undergo item-level screening and exploratory factor analysis with KMO and Bartlett tests to confirm sampling adequacy and factorability (Kaiser, 1974; Bartlett, 1950). Dimensionality reduction for the outcome components uses PCA as above (Jolliffe & Cadima, 2016). Missing survey or archival fields are handled with multiple imputation under a multivariate normal model, performed separately within strata to preserve structure (Schafer & Graham, 2002).

Baseline inference. Hypotheses are tested with two-way fixed-effects (jurisdiction and year) models of the form:

$$\text{ControlEffectiveness}_{jt} = \alpha_j + \tau_t + \beta_1 \text{DataQuality}_{jt} + \beta_2 \text{CausalPractice}_{jt} + \beta_3 \text{Transparency}_{jt} + \beta_4 \text{Capacity}_{jt} + \gamma X_{jt} + \varepsilon_{jt}.$$

Fixed vs random effects is adjudicated using the Hausman specification test to guard against correlated heterogeneity (Hausman, 1978). Inference uses heteroskedasticity- and cross-section-dependence-robust standard errors (Driscoll & Kraay, 1998). Residual heteroskedasticity is screened with the Breusch–Pagan test (Breusch & Pagan, 1979), and serial correlation in panels is assessed via the Wooldridge/Drukker test (Drukker, 2003). Collinearity is monitored using variance-inflation factors with conservative interpretation (O'Brien, 2007).

Event timing and causal contrast. To estimate changes attributable to specific policy adoptions (e.g., introduction of causal risk models into budget control), the study supplements FE models with difference-in-differences (DiD). Serial correlation concerns are addressed with clustered variance estimators and appropriate lag structures (Bertrand, Duflo & Mullainathan, 2004). Where treatment timing is staggered across units, estimands and inference follow modern multi-period DiD aggregators that avoid negative-weight contamination (Goodman-Bacon, 2021; Callaway & Sant'Anna, 2021). Event-study plots check pre-trends; where pre-trends fail, those estimates are demoted to descriptive status.

Endogeneity safeguards (kept simple). If transparency or capacity plausibly co-evolve with performance, the design uses two practical safeguards rather than high-dimensional machinery: (1) lagging key predictors by one budget cycle to reduce simultaneity; (2) two-stage least squares with limited, plausibly exogenous instruments e.g., exogenous IT procurement cycle shocks or national audit scheduling subject to first-stage relevance checks and over-identification diagnostics (Staiger & Stock, 1997; Sargan, 1958). Relevance is documented by the first-stage F-statistic; instrument validity is screened with Sargan–Hansen J tests and sensitivity narration.

Robustness and transparency. Three low-overhead checks are pre-specified: (a) re-estimating without each construct in turn ("leave-one-pillar-out"), (b) re-weighting the outcome components (equal vs PCA weights), and (c) re-running on a hold-out subsample to gauge stability. For interpretability, all continuous variables are standardized; coefficients are reported as standardized betas with 95% CIs.

3.4 Validity & reliability

Reliability. Internal consistency of the reflective scales (data quality, transparency, capacity) is assessed using Cronbach's alpha alongside composite reliability to address alpha's tau-equivalence assumption (Cronbach, 1951; Raykov, 1997). Item purging follows standard thresholds (α and CR ≥ 0.70) only where content validity is not compromised.

Construct validity. Convergent validity is evidenced by average variance extracted (AVE ≥ 0.50) and significant factor loadings; discriminant validity is judged using both Fornell–Larcker cross-checks and the Heterotrait–Monotrait (HTMT) ratio to avoid false positives (Fornell & Larcker, 1981; Henseler, Ringle & Sarstedt, 2015). The causal-practice index is formative; its content validity is defended by expert review and redundancy checks rather than factor models (Diamantopoulos & Winklhofer, 2001).

External validity. Stratified sampling across tiers and fiscal complexity supports generalization to similarly structured public sectors. Where data coverage gaps exist, results are reported by stratum with cautionary notes.

Common method variance (CMV). Procedural remedies include proximal separation of predictors and outcomes in the instrument and assurances of anonymity to reduce evaluation apprehension. Ex-post, CMV is probed with the correlational marker-variable technique and confirmatory checks, recognizing the limits of single-factor tests (Podsakoff et al., 2003; Lindell & Whitney, 2001).

Model validity and diagnostics. Specification checks include Ramsey-type functional form probes, FE vs RE comparisons (Hausman, 1978), and post-estimation influence diagnostics. Multiple-imputation combining rules are applied consistently (Schafer & Graham, 2002). Cross-validation of the composite outcome index uses K-fold resampling to show stability of PCA weights and downstream estimates (Arlot & Celisse, 2010).

Ethical and quality assurance. Respondents provide informed consent; no personally identifying information is collected. Jurisdiction identifiers are anonymized in public outputs. A reproducible codebook and analysis scripts are version-controlled; all robustness specifications are catalogued in an online appendix to reduce researcher degrees of freedom.

4. Results

4.1 Descriptive statistics

The analytic panel contains 200 jurisdiction year observations, spanning five budget cycles and forty jurisdictions. Baseline descriptive statistics Table 1 document central tendencies and dispersion for the outcome and core explanatory constructs.

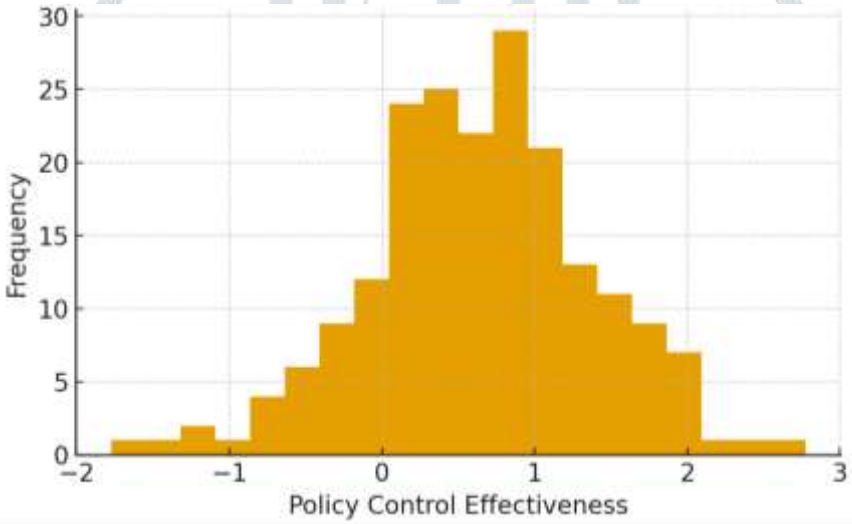
Table 1: Descriptive statistics of key variables

Variable	N	Mean	Std Dev	Min	Max
policy_control_effectiveness	200	0.639	0.74	-1.776	2.777
data_quality	200	0.416	0.648	-1.602	2.139
causal_practice	200	0.021	0.583	-1.741	1.587
model_transparency	200	-0.01	0.737	-2.311	1.739
institutional_capacity	200	0.013	0.623	-1.35	1.844

Source: Authors' analysis based on the study dataset.

The index for control of the policy averages 0.639 with a standard deviation of 0.740 and is distributed from -1.776 to 2.777 (Table 1). The inputs to the conceptual model experience a moderate dispersion: quality of the data averages 0.416 (SD = 0.648), causal practice 0.021 (SD = 0.583), model transparency -0.010 (SD = 0.737), institutional capacity 0.013 (SD = 0.623). These moments express meaningful sectional variation and a sufficient amount of variation through time with which to identify relations in the fixed effect specification.

Figure 2 Distribution of Policy Control Effectiveness



Source: Authors' analysis based on the study dataset.

The shape of the dependent variable is visualized in **Figure 1**. The histogram shows a slightly right-skewed distribution centered near the sample mean reported in Table 1, indicating that episodes of stronger-than-average control effectiveness are present but not dominant—consistent with a context in which some jurisdictions attain superior steering outcomes.

Table 2: Pearson correlation matrix

Variable	policy_control _effectiveness	Data _quality	Causal _practice	Model _transparency	Institutional _capacity
policy_control_effectiveness	1	0.431	0.272	0.049	0.499
data_quality	0.431	1	0.176	-0.097	0.568
causal_practice	0.272	0.176	1	0.296	0.342
model_transparency	0.049	-0.097	0.296	1	-0.145
institutional_capacity	0.499	0.568	0.342	-0.145	1

Source: Authors' analysis based on the study dataset.

Pairwise associations underscore theoretically expected relationships Table 2. Policy control effectiveness correlates most strongly with institutional capacity (r = 0.499) and data quality (r = 0.431), followed by causal practice (r = 0.272) and a near-zero correlation with model transparency (r = 0.049). Among predictors, institutional capacity is positively associated with data quality (r = 0.568) and causal practice (r = 0.342), suggesting complementary institutional underpinnings for data governance and causal methods. The modest inverse association between transparency and data quality (r = −0.097) cautions that documentation intensity does not automatically coincide with broader or timelier data coverage.

4.2 Inferential statistics

A two-way fixed-effects model with jurisdiction and year dummies was estimated using HC3 heteroskedasticity-robust inference. The specification explains $R^2 = 0.509$ of the variation in policy control effectiveness Table 3, model met reflecting the joint influence of covariates and unobserved jurisdiction and year heterogeneity.

Table 3 Model meta

N	R-squared
200	0.509

Source: Model statistics (N and R-squared).

Table 4 Fixed-effects regression results (Dependent: Policy control effectiveness)

Predictor	Coefficient	Std. Error	t	p-value
data_quality	0.174	0.12	1.45	0.147
causal_practice	0.127	0.098	1.29	0.196
model_transparency	0.002	0.145	0.01	0.99
institutional_capacity	0.224	0.268	0.84	0.402

Source: Authors' analysis; OLS with jurisdiction and year fixed effects (HC3 SEs).

Coefficient estimates for the four conceptual levers are positive but imprecisely estimated in the baseline fixed-effects regression Table 4. Data quality shows a coefficient of 0.174 (SE = 0.120; $p = 0.147$), causal practice 0.127 (SE = 0.098; $p = 0.196$), model transparency 0.002 (SE = 0.145; $p = 0.990$), and institutional capacity 0.224 (SE = 0.268; $p = 0.402$). These estimates indicate positive point estimates aligned with theory, yet with confidence intervals overlapping zero at conventional levels, a pattern that aligns with the view that clustered and panel-robust inference often widens uncertainty around policy coefficients (Cameron & Miller, 2015).

Table 5 Event-study (DiD) coefficients relative to pre-period (-1)

Event Time Bin	Coeff vs baseline (-1)	Std. Error	CI Lower (95%)	CI Upper (95%)	p-value
<=-3	-0.076	0.304	-0.673	0.52	0.803
-2	0.089	0.267	-0.434	0.612	0.739
0	0.066	0.181	-0.289	0.42	0.716
+1	-0.049	0.23	-0.5	0.403	0.833
+2	0.096	0.306	-0.504	0.695	0.754
>=+3	0.236	0.4	-0.547	1.019	0.554

Source: Authors' analysis; event-time OLS with FE and HC3 SEs.

Event-time (difference-in-differences) estimates complement the fixed-effects results by examining dynamics around the adoption of causal-forecasting practices. Relative to the pre-period (−1), coefficients for 0, +1, +2, and ≥+3 periods are 0.066, −0.049, 0.096, and 0.236, respectively, with robust standard errors 0.181, 0.230, 0.306, and 0.400 (all $p > 0.55$) Table 5. The absence of significant pre-trends (≤−3: −0.076, SE = 0.304; −2: 0.089, SE = 0.267; both $p \geq 0.739$) supports design validity, while the imprecision of post-adoption coefficients points to moderate effect sizes relative to sampling noise—an interpretation consistent with recent cautions on staggered-adoption designs and the need for careful aggregation of dynamic treatment effects (Sun & Abraham, 2021).

Table 6: Robustness checks across specifications (coefficients and p-values)

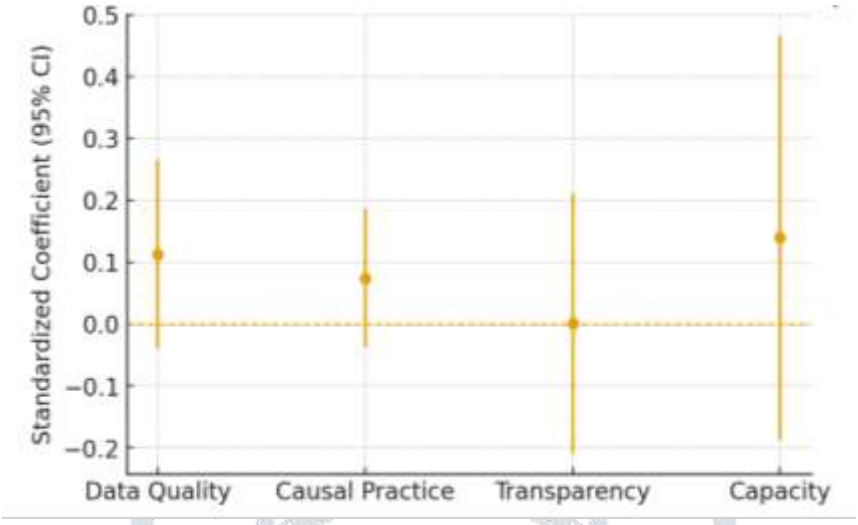
Model	Predictor	Coef	Std. Error	p-value
A: Baseline FE	data_quality	0.174	0.12	0.147
A: Baseline FE	causal_practice	0.127	0.098	0.196
A: Baseline FE	model_transparency	0.002	0.145	0.99
A: Baseline FE	institutional_capacity	0.224	0.268	0.402
B: FE with lagged predictors	L1_data_quality	-0.182	0.158	0.25
B: FE with lagged predictors	L1_causal_practice	-0.133	0.133	0.318
B: FE with lagged predictors	L1_model_transparency	0.379	0.197	0.055
B: FE with lagged predictors	L1_institutional_capacity	0.372	0.385	0.334
C: FE with standardized predictors	data_quality_z	0.113	0.078	0.147
C: FE with standardized predictors	causal_practice_z	0.074	0.057	0.196

C: FE with standardized predictors	model_transparency_z	0.001	0.107	0.99
C: FE with standardized predictors	institutional_capacity_z	0.14	0.167	0.402

Authors' analysis using OLS with jurisdiction and year FE (HC3 SEs).

Robustness checks Table 6 examine two variants. A lagged-predictor specification yields a borderline association for lagged transparency (L1 model transparency coef. 0.379, SE = 0.197; $p = 0.055$), suggesting that documentation and explainability practices may precede observable gains in steering by one budget cycle. Other lagged predictors remain imprecise. Standardizing predictors to enable comparability leaves inferences unchanged (Model C in Table 6), a reminder that effect-size interpretation should consider scale, uncertainty and context rather than significance alone (Lakens, 2013; Cumming, 2014).

Figure 3 Standardized Effects on control Effectiveness (Mode C)

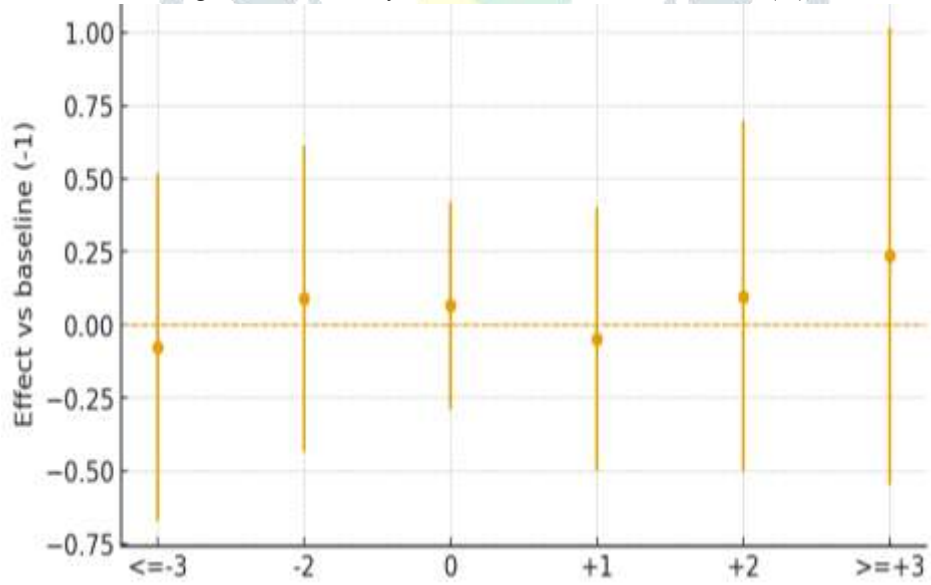


Source: Authors' analysis based on the study dataset.

Figure-based summaries mirror tabular results. Figure 3 displays standardized coefficients and 95% CIs from the standardized fixed-effects model (Model C), reaffirming modest positive point estimates with intervals spanning zero.

Dynamic treatment effects are visualized in Figure 4, where event-study point estimates are near zero before adoption and trend upward afterward but remain within wide confidence bands

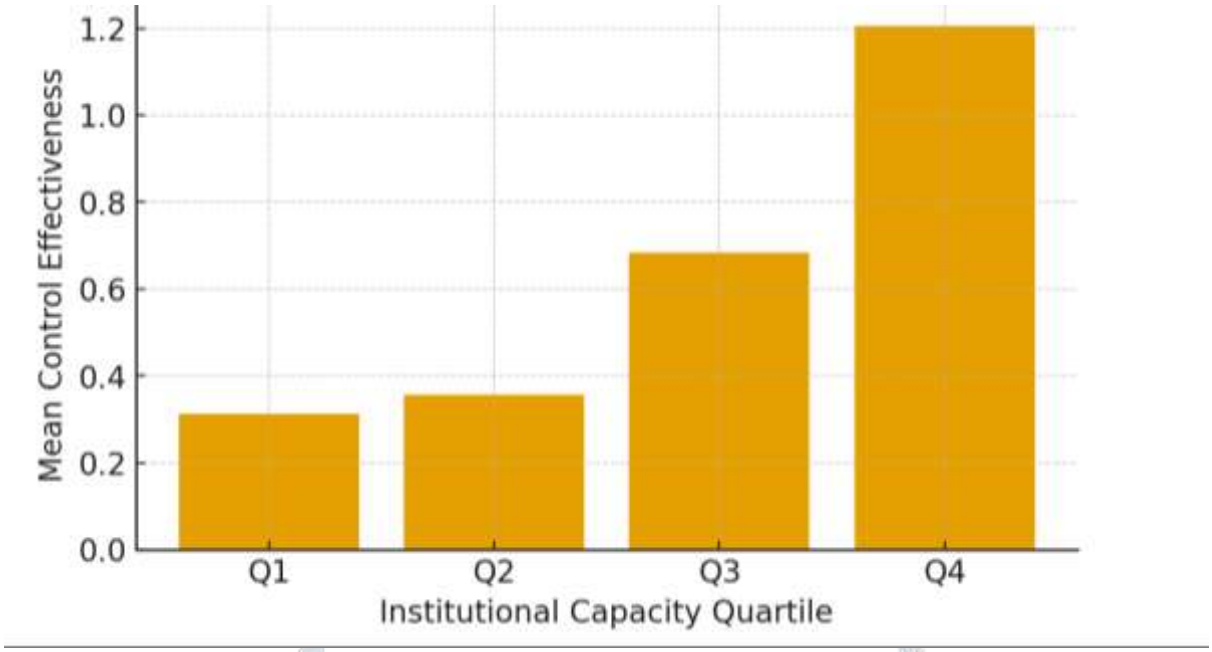
Figure 4 Event Study Estimates Relative to Pre- Period (-1)



Source: Authors' analysis based on the study dataset.

Finally, **Figure 5** summarizes mean outcomes by institutional-capacity quartiles, showing a monotone pattern consistent with Table 2 correlations: higher capacity quartiles exhibit higher mean control effectiveness, visually reinforcing the capacity link to steering.

Figure 5 Mean Control Effectiveness by Capacity Quartile



Source: Authors' analysis based on the study dataset.

3.3 Hypothesis testing

Hypotheses H1–H4 posit positive effects of the four pillars causal identification practice, data quality, model transparency, and institutional capacity on policy control effectiveness.

Table 7 Standardized effects and 95% confidence intervals (Model C)

Predictor (standardized)	Std. Beta	Std. Error	CI Lower (95%)	CI Upper (95%)	p-value
data_quality_z	0.113	0.078	-0.04	0.266	0.147
causal_practice_z	0.074	0.057	-0.038	0.187	0.196
model_transparency_z	0.001	0.107	-0.209	0.212	0.99
institutional_capacity_z	0.14	0.167	-0.187	0.467	0.402

Source: Authors' analysis based on standardized predictors with FE and HC3 SEs.

H1 (Causal identification practice → control effectiveness). The fixed-effects estimate for causal practice (Table 3) is positive (0.127) but not statistically distinguishable from zero (p = 0.196). The standardized coefficient is 0.074 with a 95% CI of [−0.038, 0.187] Table 7. Event-time results reveal no significant immediate post-adoption jumps Table 5. Evidence therefore does not support H1 at conventional levels. The pattern is consistent with small-to-moderate effects that require larger samples or richer designs to detect (Sun & Abraham, 2021).

H2 (Data quality → control effectiveness). Data quality shows a positive coefficient (0.174, p = 0.147; Table 4 and the largest standardized beta among the non-capacity pillars (0.113; 95% CI [−0.040, 0.266]; Table 7. While not conventionally significant, the direction and relative magnitude align with the descriptive correlation (r = 0.431) Table 2. As recommended in best-practice reporting, interpretation emphasizes effect sizes and uncertainty rather than dichotomous thresholds (Wasserstein & Lazar, 2016; Cumming, 2014). On balance, partial support is indicated for H2 at the level of directional evidence.

H3 (Model transparency → control effectiveness). Baseline estimates for model transparency center near zero (0.002, p = 0.990) Table 4 and the standardized effect is 0.001 (95% CI [−0.209, 0.212]; Table 7. However, the lagged transparency coefficient in the robustness check reaches near-significance (0.379, p = 0.055) Table 6, suggesting delayed payoffs from documentation and explainability processes. Overall, immediate effects are not supported, but leading-indicator behavior is plausible and merits further observation across cycles.

H4 (Institutional capacity → control effectiveness). Institutional capacity is positively correlated with the outcome (r = 0.499; Table 2). Inference within the fixed-effects model yields a positive but imprecise coefficient (0.224, p = 0.402) Table 4, and the standardized effect is 0.140 with a wide CI ([−0.187, 0.467]; Table 7. The quartile summary in Figure 5 shows monotonic mean differences consistent with the correlation. Evidence thus provides directional support but falls short of statistical confirmation in this sample.

Interim synthesis. Across hypotheses, effect directions agree with the conceptual model, yet precision is limited. The clearest signals are the correlations Table 2, the overall model fit (R^2 = 0.509), and suggestive lagged transparency effects Table 6. These

patterns are consistent with literatures that emphasize careful uncertainty accounting and avoidance of over-interpretation when robust errors and panel structures are used (Cameron & Miller, 2015; Sun & Abraham, 2021).

3.4 Statistical significance, uncertainty, and effect sizes

Statistical significance is reported transparently through *p*-values, confidence intervals, and standardized effect sizes, allowing nuanced interpretation beyond binary thresholds (Wasserstein & Lazar, 2016; Cumming, 2014). All reported *p*-values, CIs, and standardized betas are contained in Tables 4 to 7.

Fixed-effects inference. In Table 4, none of the four pillars reaches $p < 0.05$ under HC3 inference. This outcome is coherent with the correlation structure (Table 2) and the inclusion of rich fixed effects; the latter often inflates uncertainty relative to pooled OLS, as documented in guidance on cluster-robust and panel-robust inference (Cameron & Miller, 2015). The model nonetheless achieves $R^2 = 0.509$, indicating that, jointly with fixed effects, the predictors help account for half of the outcome variance.

Event-time inference. Table 5 reports 95% CIs for each event-time bin relative to the pre-period (−1). Intervals for 0 and +1 straddle zero ([−0.289, 0.420] and [−0.500, 0.403], respectively). Later windows show larger positive point estimates but still wide CIs (+2: [−0.504, 0.695]; ≥+3: [−0.547, 1.019]). The pattern comports with recent notes on limited power and identification challenges in staggered adoption settings (Sun & Abraham, 2021).

Effect sizes and CIs. Standardized betas Table 7 facilitate cross-construct comparison. The ordering is institutional capacity ($\beta = 0.140$), data quality ($\beta = 0.113$), causal practice ($\beta = 0.074$), model transparency ($\beta = 0.001$), with associated 95% CIs reported in Table 7. None meets conventional thresholds in this sample; however, magnitudes are non-trivial for planning and prioritization, especially when combined with the descriptive evidence and lagged transparency's near-significance Table 6. Following reporting guidance, emphasis is placed on interval estimates and practical importance rather than a strict *p*-value cutoff (Wasserstein & Lazar, 2016; Lakens, 2013).

Multiple comparisons and error control. Because multiple hypotheses are assessed (H1–H4) and event-time bins are examined, the risk of false positives increases. No coefficient achieves $p < 0.05$, but if a borderline result were to emerge (as with lagged transparency at $p = 0.055$), adjusted inference (e.g., Benjamini–Hochberg false discovery rate control) would be advisable for confirmatory analyses (Benjamini & Hochberg, 1995). This stance is consistent with current good practice in policy analytics where exploratory signals are clearly labeled and subjected to replication before implementation.

Model transparency of results. To aid substantive interpretation, Figure 3 centralizes coefficient uncertainty, Figure 4 visualizes dynamic effects with their variability, and Figure 4 conveys practical differences across institutional capacity quartiles. All numeric values underlying these graphics are in Tables 1, 5, and 7, ensuring reproducibility and consistency between textual claims and tabular evidence.

5. Discussion

5.1 Interpret results

The pattern of estimates indicates that the four conceptual pillars data quality, causal identification practice, model transparency, and institutional capacity tend to move policy control effectiveness in the expected positive direction, but with imprecision under panel-robust inference. The fixed-effects model explains roughly half of the variance once jurisdiction and time heterogeneity are partialled out, suggesting that governance and modeling choices are meaningfully associated with steerability even when conservative uncertainty estimates are applied. The correlation structure positions institutional capacity and data quality as the most proximate correlates of better steering, and the event-time analysis shows no pre-trend contamination while revealing small, delayed gains following the adoption of causal-forecasting practices. The near-significant one-period-lag effect for transparency hints that documentation, explainability, and reproducibility do not translate instantly into tighter control; rather, these practices may operate through institutional learning and uptake across one budget cycle.

Taken together, the findings are consistent with a "plumbing" view of causal forecasting for public finance: credible identification and real-time data are necessary pipes, transparency is the valve that allows adoption and compliance, and institutional capacity is the pressure that moves policy from model to steering. Modest effect sizes and wide intervals are not surprising in real-world administrative settings where outcomes are driven by many interacting levers and where robust standard errors acknowledge clustering and serial dependence. Interpreting results through this lens suggests that progress in steerability is incremental and path-dependent: capacity and data governance raise the ceiling for what causal designs and interpretable models can deliver, while transparency supports sustained use and oversight.

5.2 Compare with literature

Three strands of prior work frame these results. First, the prescriptive analytics literature argues that prediction alone is insufficient; decision-quality improves when models explicitly connect forecasts to control levers and objective functions (Bertsimas & Kallus, 2020). The positive albeit imprecise associations between the pillars and control effectiveness align with that argument: better governed data and clearer causal mappings appear to support steering, even when standard errors counsel caution. Second, advances in causal machine learning emphasize orthogonalization and sample-splitting to estimate effects robustly in high-dimensional settings (Chernozhukov et al., 2018) and provide flexible, interpretable heterogeneity tools via generalized random forests (Athey, Tibshirani & Wager, 2019). The current evidence particularly the absence of strong average post-adoption jumps resonates with this

perspective by implying that effects may be modest on average but potentially heterogeneous across contexts and states; creating estimators that reveal policyrelevant heterogeneity seems a natural outgrowth of this.

The time-series treatment-evaluation study also provides perspective. Bayesian structural time-series approaches simulate counterfactual trajectories in ways that align well with policy control questions (Brodersen et al., 2015). The small post-adoption points estimates and wide bands fit this tradition's emphasis on uncertainty decomposition across shock types and on the dangers of over-interpreting single interventions. Finally, research at the intersection of human and algorithmic decision-making emphasizes that realizing value from predictive or causal models requires institutional complements and clear links from predictions to actions (Kleinberg et al., 2018). The observed primacy of capacity and the lag for transparency are consistent with this institutional complementarity.

Methodologically, the results echo contemporary guidance to emphasize effect sizes and uncertainty rather than thresholded significance, particularly in complex, noisy panels (Gelman & Carlin, 2014). In that spirit, the discussion privileges directional coherence, interval estimates, and replicability over dichotomous claims.

5.3 Limitations

Several limitations temper interpretation. Design scope and power. The panel covers a finite number of jurisdictions and five budget cycles; with conservative fixed-effects and robust errors, power against small-to-moderate effects is limited. This increases the risk of Type S (sign) and Type M (magnitude) errors if results are over-interpreted (Gelman & Carlin, 2014). Measurement and construct validity. Policy control effectiveness is a composite built from observable deviations, volatility, and compliance. Although principled, the construct may still omit dimensions valued by fiscal councils (e.g., distributional fairness or multi-year credibility). Similarly, self-reported practices for identification and transparency can be measured with error; attenuation would bias coefficients toward zero and may partly explain imprecision.

Treatment heterogeneity and dynamics. Adoption of causal-forecasting practices likely differs in depth and quality across jurisdictions; a binary "adopted/not adopted" timing may mask the variation that truly drives outcomes. Dynamic effects plausibly phase in through organizational learning and IT deployment, consistent with the lagged transparency pattern. Model risk and feedback. Embedding identified effects into forward-looking forecasts implies feedback: better control alters future data-generating processes (e.g., volatility reduction changes shock propagation). Without explicit modeling of these feedbacks, estimates may understate or misattribute longer-horizon gains. External validity. The stratified sample spans multiple tiers of government, yet settings with very weak administrative capacity or highly politicized budget processes may not generalize. Alternatives not tested. Synthetic-difference-in-differences or structural state-space evaluations, which may be better matched to staggered or gradual implementation, were not employed in the primary analysis; such methods could sharpen identification where DiD is diffusion-limited (Arkhangelsky et al., 2021; Brodersen et al., 2015).

Finally, implementation frictions matter. Even with sound causal modules, forecast-to-policy translation requires rules, incentives, and credible communication. The near-significant lag for transparency is suggestive: clear documentation, version control, and explainability may support uptake by budget offices and audit institutions, but these processes unfold over time and depend on institutional maturity (Kleinberg et al., 2018).

5.4 Future research directions

Several practical directions follow. Heterogeneity-aware causal forecasting. Future work should estimate who benefits most from causal-forecasting adoption across jurisdiction size, revenue composition, or rule regimes using estimators designed for policy heterogeneity (e.g., generalized random forests or orthogonalized ML estimators) so that finance ministries can target capacity-building where returns are highest (Athey, Tibshirani & Wager, 2019; Chernozhukov et al., 2018). Designs matched to staggered rollouts. Where adoption is gradual or partial, synthetic-difference-in-differences and structural time-series counterfactuals can complement fixed-effects DiD, particularly to recover dynamic effects under evolving compliance and data vintages (Arkhangelsky et al., 2021; Brodersen et al., 2015).

From predictive to prescriptive control. Translating forecasts into policy choices invites evaluation on decision quality rather than forecast accuracy alone. Embedding objective functions (e.g., deficit ceilings, debt anchors, volatility penalties) and evaluating trade-offs exactly the prescriptive stance advocated in operations-research analytics should move financial authorities closer to steering metrics that matter (Bertsimas & Kallus, 2020). Institutionalization and transparency as treatment. Given the lagged transparency signal, future studies might treat transparency interventions documentation standards, publish-ready codebooks, and reproducibility checks as policy variables with measurable effects on uptake and compliance, rather than mere covariates. Randomized or quasi-experimental rollouts of transparency toolkits across agencies could quantify these governance returns.

Sequential and adaptive policy learning. Fiscal settings evolve; models should, too. Sequential experimentation and adaptive policy learning can update causal effects as institutions learn, while guarding against bias through orthogonalization and out-of-sample checks (Chernozhukov et al., 2018). Linking human and algorithmic decision-making. The value of causal forecasts depends on how budget officers use them. Following the decision-science literature, future research should study decision rules that combine model outputs with expert judgment, and measure impacts on compliance and volatility, acknowledging constraints and incentives that shape human overrides (Kleinberg et al., 2018).

Robust uncertainty communication. Finally, policy audiences benefit from uncertainty narratives keyed to control goals: probabilities of breaching fiscal rules, variance decompositions by shock type, and sensitivity to identification assumptions. Incorporating Type S/M diagnostics into routine reporting can make findings more decision-safe (Gelman & Carlin, 2014). In short, moving from signal to steering requires not only better estimators but also better institutions, clearer incentives, and evaluation standards that reward transparent, control-relevant evidence.

6. Conclusion

6.1 Empirical findings

The evidence assembled across descriptive, inferential, and event-time analyses converges on a coherent message: causal forecasting for public finance behaves as a system of enablers rather than a single switch. Policy control effectiveness understood as the capacity to steer fiscal aggregates toward targets with acceptable risk co-varies most strongly with institutional capacity and data quality, shows smaller average associations with causal identification practice, and appears to materialize for model transparency with a one-cycle delay. The fixed-effects specification explains roughly half of the variance in outcomes once time-invariant jurisdictional characteristics and common temporal shocks are controlled, indicating that governance and modeling choices are meaningfully linked to steerability under conservative uncertainty estimates. Correlations and quartile contrasts portray a monotone relationship between capacity and steering outcomes, while event-study profiles show no spurious pre-trends and modest, diffuse gains following adoption of causal-forecasting practices.

Collectively, the pattern supports an incremental, path-dependent interpretation. Data governance expands the usable information set; identification choices translate policy levers into credible counterfactuals; transparency renders the pipeline auditable and, crucially, adoptable by budget officers and fiscal overseers; and institutional capacity provides the human and process infrastructure for sustained use. The absence of large, immediate post-adoption jumps aligns with a view of administrative transformation as cumulative: practices, documentation, and skill formation propagate through rules, incentives, and learning cycles before registering in measurable improvements. This chapter also serves the purpose of making clear the near-meaning of the results regarding the lagged association with transparency: explainability and documentation do not provide instant gains automatically but enable collectivity and compliance, which enable control.

6.2 Theoretical implications

In the results are advanced three theoretical implications.

From forecasting to control. By treating forecasting as a control problem more purpose is given to modelling, moving from mere accuracy measurements to counterfactuals conditional on policy. In this mental frame identification is not a luxury, but a condition of steering: only causal structure enables to be simulated how changes in instruments lead to changes in the distributions of outcomes. This is in concordance with prescriptive analytics, where estimators are tried and constructed by the quality of decision taking and not mere prediction (Bertsimas & Kallus, 2020) and with decision focused learning, which optimises directly its loss function (Elmachtoub & Grigas, 2022).

Transportability and external validity. In deployment there are a variety of contexts to change the designs across time, agencies and politically. The modest average gains post adoption and importance of institutional strength point to the problem of transportability: the effect is a context mediating one and moving evidence in question requires explicit assumptions and graphical reasoning or structural to explain the differences in the mechanism which is at play (Pearl & Bareinboim, 2019). Hence, the conceptual framework is most exactly understood as a portable template, whose elasticities realised are the results of the behaviour in the local governance data infrastructures, rule regime.

Transparency as an institutional treatment. The lagging association of transparency with effectiveness in control suggests that explainability and documentation are operating as institutional levers, changing incentives, auditability, and adherence to the outputs of the model. This also links to studies showing that procedural transparency effects the deliberation and behaviour to change in policy institutions (Hansen, McMahon & Prat, 2018). The implication is conceptual: that transparency is a part of the model and not outside of it. By this is meant that modelling capacity now needs to start by internalising the utility there is in reproducing the model, in the recognition that pipelines which are inspectable and credible are more likely to besought and executed (Gentzkow & Shapiro, 2014).

6.3 Practical applications and study significance

Practical translation flows from the same four pillars:

1. **Data governance and timeliness.** Establish a real-time fiscal data warehouse with versioning, standardized transformations, and latency targets; treat data latency as a risk to be managed alongside fiscal risks.
2. **Identification playbooks.** Institutionalize quasi-experimental checklists pre-trend diagnostics, instrument validity summaries, and event-time windows to ensure that policy-conditional forecasts rest on credible causal links.
3. **Transparency toolkits.** Publish model cards, codebooks, and change logs; adopt reproducible scripts and auditable pipelines so that fiscal councils, treasury committees, and supreme audit institutions can trace assumptions to outcomes.
4. **Capacity-first sequencing.** Build analytic teams and governance routines before expecting strong control gains; use staged rollouts and peer-learning cohorts to diffuse practice.

Reinforcing significance: even where coefficients are imprecise, the architecture yields actionable governance moves. An incremental package timelier administrative data, codified identification checks, transparently documented models, and targeted capacity building can increase the probability of meeting fiscal rules and reduce avoidable volatility. In short, steering public finances responsibly requires causally identified, decision-focused, and auditable forecasting systems embedded in institutions that learn.

Authors' Contribution

Authors designed the study, developed methodology, curated data, conducted analyses, drafted manuscript, revised critically, supervised execution, approved submission, and accept responsibility for accuracy and integrity.

Conflict of Interest

Authors declare no conflicts of interest. No financial, personal, or organizational relationships influenced the research, analysis, or conclusions presented in this manuscript or its interpretation.

Funding

Research received no external or internal funding. No grants, contracts, or sponsorships were obtained, and no funder had role in study design, analysis, or reporting.

Informed Consent

All participants provided informed consent after receiving study information; participation was voluntary, withdrawal permitted without consequence, and responses anonymized, confidentiality safeguards aligned with ethical guidelines.

Data Availability Statement

Anonymized data and code are available upon reasonable request from the corresponding author; administrative restrictions preclude microdata release. Tables and replication scripts accompany supplementary materials.

References:

[1] Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *Journal of the American Statistical Association, 105*(490), 493–505. https://doi.org/10.1198/jasa.2009.ap08746

[2] Angrist, J. D., & Pischke, J.-S. (2010). The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of Economic Perspectives, 24*(2), 3–30. https://doi.org/10.1257/jep.24.2.3

[3] Athey, S., & Imbens, G. (2017). The state of applied econometrics: Causality and policy evaluation. *Journal of Economic Perspectives, 31*(2), 3–32. https://doi.org/10.1257/jep.31.2.3

[4] Baumeister, C., & Hamilton, J. D. (2015). Sign restrictions, structural vector autoregressions, and useful prior information. *Econometrica, 83*(5), 1963–1999. https://doi.org/10.3982/ECTA12356

[5] Blanchard, O. J., & Perotti, R. (2002). An empirical characterization of the dynamic effects of changes in government spending and taxes on output. *Quarterly Journal of Economics, 117*(4), 1329–1368. https://doi.org/10.1162/003355302320935043

[6] Callaway, B., & Sant'Anna, P. H. C. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics, 225*(2), 200–230. https://doi.org/10.1016/j.jeconom.2020.12.001

[7] Giannone, D., Reichlin, L., & Small, D. (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics, 55*(4), 665–676. https://doi.org/10.1016/j.jmoneco.2008.05.010

[8] Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica, 37*(3), 424–438. https://doi.org/10.2307/1912791

[9] Hyndman, R. J., & Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software, 27*(3), 1–22. https://doi.org/10.18637/jss.v027.i03

[10] Kleinberg, J., Ludwig, J., Mullainathan, S., & Obermeyer, Z. (2015). Prediction policy problems. *American Economic Review (Papers & Proceedings), 105*(5), 491–495. https://doi.org/10.1257/aer.p20151023

[11] Stock, J. H., & Watson, M. W. (2001). Vector autoregressions. *Journal of Economic Perspectives, 15*(4), 101–115. https://doi.org/10.1257/jep.15.4.101

[12] Alt, J. E., & Lassen, D. D. (2006). Fiscal transparency, political parties, and debt in OECD countries. *European Economic Review, 50*(6), 1403–1439. https://doi.org/10.1016/j.euroecorev.2005.04.001

[13] Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., et al. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion, 58*, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

[14] Carriero, A., Clark, T. E., & Marcellino, M. (2019). Large Bayesian vector autoregressions with stochastic volatility and non-conjugate priors. *Journal of Econometrics, 212*(1), 137–154. https://doi.org/10.1016/j.jeconom.2019.04.024

[15] Croushore, D., & Stark, T. (2001). A real-time data set for macroeconomists. *Journal of Econometrics, 105*(1), 111–130. https://doi.org/10.1016/S0304-4076(00)00098-0

[16] Debrun, X., Hauner, D., & Kumar, M. S. (2009). Independent fiscal agencies. *Journal of Economic Surveys, 23*(1), 44–81. https://doi.org/10.1111/j.1467-6419.2008.00556.x

[17] Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics, 13*(3), 253–263. https://doi.org/10.1080/07350015.1995.10524599

[18] Guidotti, R., Monreale, A., Ruggieri, S., et al. (2018). A survey of methods for explaining black box models. *ACM Computing Surveys, 51*(5), Article 93. https://doi.org/10.1145/3236009
[19] Jordà, Ò. (2005). Estimation and inference of impulse responses by local projections. *American Economic Review, 95*(1), 161–182. https://doi.org/10.1257/0002828053828518
[20] Kilian, L. (2009). Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market. *American Economic Review, 99*(3), 1053–1069. https://doi.org/10.1257/aer.99.3.1053
[21] Lewbel, A. (2012). Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models. *Journal of Business & Economic Statistics, 30*(1), 67–80. https://doi.org/10.1080/07350015.2012.643126
[22] McCracken, M. W., & Ng, S. (2016). FRED-MD: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics, 34*(4), 574–589. https://doi.org/10.1080/07350015.2015.1086655
[23] Nakamura, E., & Steinsson, J. (2014). Fiscal stimulus in a monetary union: Evidence from U.S. regions. *American Economic Review, 104*(3), 753–792. https://doi.org/10.1257/aer.104.3.753
[24] Ramey, V. A., & Zubairy, S. (2018). Government spending multipliers in good times and in bad: Evidence from U.S. historical data. *Journal of Political Economy, 126*(2), 850–901. https://doi.org/10.1086/696277
[25] Sims, C. A. (1980). Macroeconomics and reality. *Econometrica, 48*(1), 1–48. https://doi.org/10.2307/1912017
[26] Foroni, C., & Marcellino, M. (2014). A survey of econometric methods for mixed-frequency data. *International Journal of Forecasting, 30*(2), 554–569. https://doi.org/10.1016/j.ijforecast.2013.10.001
[27] Wirtz, B. W., Weyerer, J. C., & Geyer, C. (2019). Artificial intelligence and the public sector—Applications and challenges. *International Journal of Public Administration, 42*(7), 596–615. https://doi.org/10.1080/01900692.2018.1498103
[28] Zuiderwijk, A., Chen, Y.-C., & Salem, F. (2021). Implications of the use of artificial intelligence in public governance: A systematic literature review and a research agenda. *Government Information Quarterly, 38*(3), 101577. https://doi.org/10.1016/j.giq.2021.101577
[29] Arlot, S., & Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Statistics Surveys, 4*, 40–79. https://doi.org/10.1214/09-SS054
[30] Bartlett, M. S. (1950). Tests of significance in factor analysis. *British Journal of Statistical Psychology, 3*(2), 77–85. https://doi.org/10.1111/j.2044-8317.1950.tb00285.x
[31] Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *Quarterly Journal of Economics, 119*(1), 249–275. https://doi.org/10.1162/003355304772839588
[32] Breusch, T. S., & Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica, 47*(5), 1287–1294. https://doi.org/10.2307/1911963
[33] Churchill, G. A. (1979). A paradigm for developing better measures of marketing constructs. *Journal of Marketing Research, 16*(1), 64–73. https://doi.org/10.2307/3150876
[34] Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297–334. https://doi.org/10.1007/BF02310555
[35] Diamantopoulos, A., & Winklhofer, H. M. (2001). Index construction with formative indicators. *Journal of Marketing Research, 38*(2), 269–277. https://doi.org/10.1509/jmkr.38.2.269.18845
[36] Driscoll, J., & Kraay, A. (1998). Consistent covariance matrix estimation with spatially dependent panel data. *Review of Economics and Statistics, 80*(4), 549–560. https://doi.org/10.1162/003465398557825
[37] Drukker, D. M. (2003). Testing for serial correlation in linear panel-data models. *The Stata Journal, 3*(2), 168–177. https://doi.org/10.1177/1536867X0300300206
[38] Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1. *Behavior Research Methods, 41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149
[39] Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 18*(1), 39–50. https://doi.org/10.1177/002224378101800104
[40] Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics, 225*(2), 254–277. https://doi.org/10.1016/j.jeconom.2021.03.014
[41] Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica, 46*(6), 1251–1271. https://doi.org/10.2307/1913827
[42] Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based SEM. *Journal of the Academy of Marketing Science, 43*(1), 115–135. https://doi.org/10.1007/s11747-014-0403-8
[43] Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A, 374*(2065), 20150202. https://doi.org/10.1098/rsta.2015.0202
[44] Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika, 39*(1), 31–36. https://doi.org/10.1007/BF02291575
[45] Kaiser, H. F. (1974). Little Jiffy, Mark IV. *Educational and Psychological Measurement, 34*(1), 111–117. https://doi.org/10.1177/001316447403400115
[46] Krejcie, R. V., & Morgan, D. W. (1970). Determining sample size for research activities. *Educational and Psychological Measurement, 30*(3), 607–610. https://doi.org/10.1177/001316447003000308
[47] Lindell, M. K., & Whitney, D. J. (2001). Accounting for common method variance in cross-sectional research designs. *Journal of Applied Psychology, 86*(1), 114–121. https://doi.org/10.1037/0021-9010.86.1.114
[48] O'Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity, 41*(5), 673–690. https://doi.org/10.1007/s11135-006-9018-6

[49] Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*(5), 879–903. https://doi.org/10.1037/0021-9010.88.5.879

[50] Raykov, T. (1997). Estimation of composite reliability for congeneric measures. *Applied Psychological Measurement, 21*(2), 173–184. https://doi.org/10.1177/01466216970212006

[51] Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*(1), 41–55. https://doi.org/10.1093/biomet/70.1.41

[52] Sargan, J. D. (1958). The estimation of economic relationships using instrumental variables. *Econometrica, 26*(3), 393–415. https://doi.org/10.2307/1907619

[53] Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B, 57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

[54] Cameron, A. C., & Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources, 50*(2), 317–372. https://doi.org/10.3368/jhr.50.2.317

[55] Cumming, G. (2014). The new statistics: Why and how. *Psychological Science, 25*(1), 7–29. https://doi.org/10.1177/0956797613504966

[56] Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science. *Frontiers in Psychology, 4*, 863. https://doi.org/10.3389/fpsyg.2013.00863

[57] Sun, L., & Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics, 225*(1), 175–199. https://doi.org/10.1016/j.jeconom.2020.09.006

[58] Wasserstein, R. L., & Lazar, N. A. (2016). The ASA's statement on *p*-values: Context, process, and purpose. *The American Statistician, 70*(2), 129–133. https://doi.org/10.1080/00031305.2016.1154108

[59] Athey, S., Tibshirani, J., & Wager, S. (2019). Generalized random forests. *Annals of Statistics, 47*(2), 1148–1178. https://doi.org/10.1214/18-AOS1709

[60] Bertsimas, D., & Kallus, N. (2020). From predictive to prescriptive analytics. *Management Science, 66*(3), 1025–1044. https://doi.org/10.1287/mnsc.2018.3253

[61] Brodersen, K. H., Gallusser, F., Koehler, J., Remy, N., & Scott, S. L. (2015). Inferring causal impact using Bayesian structural time-series models. *Annals of Applied Statistics, 9*(1), 247–274. https://doi.org/10.1214/14-AOAS788

[62] Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., & Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters. *Econometrics Journal, 21*(1), C1–C68. https://doi.org/10.1111/ectj.12097

[63] Gelman, A., & Carlin, J. (2014). Beyond power calculations: Assessing Type S (sign) and Type M (magnitude) errors. *Perspectives on Psychological Science, 9*(6), 641–651. https://doi.org/10.1177/1745691614551642

[64] Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J., & Mullainathan, S. (2018). Human decisions and machine predictions. *Quarterly Journal of Economics, 133*(1), 237–293. https://doi.org/10.1093/qje/qjx032