



A Predictive Control Architecture for Information Technology Program Portfolio Risk Management Using Causal Machine Learning

Ezeokechukwu Chiemere Victor^{1*}, Funmilayo Abibat Sanusi²

¹ Department of Computer Science, School of Computing and Engineering Sciences, Babcock University, Ilishan-Remo 121103, Ogun State, Nigeria

² Department of Software Engineering, School of Computing and Engineering Sciences, Babcock University, Ilishan-Remo 121103, Ogun State, Nigeria

Corresponding Author Email: ezeokechukwu0273@pg.babcock.edu.ng

Copyright: ©2026 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/jesa.590921>

ABSTRACT

Received: 14 July 2026

Revised: 7 September 2026

Accepted: 22 September 2026

Available online: 30 September 2026

Keywords:

predictive control, portfolio governance, causal inference, feature attribution, schedule risk, inverse probability weighting

Governance offices overseeing multi-project information technology (IT) portfolios generally have no principled way to decide which interventions to apply, to which investments, or when in the reporting cycle to act. We treat this as a control problem. This paper sets out a predictive control architecture that casts IT portfolio schedule delay risk management as a discrete-time state feedback problem: portfolio risk is the controlled variable, governance interventions are the control inputs, and the gain matrix comes from causally estimated treatment effects rather than from associational feature attribution. Drawing on 5,729 federal IT investment records from the United States Office of Management and Budget IT Dashboard, we estimate the causal effect of three interventions using inverse probability weighting (IPW), with a documented adjustment set grounded in domain knowledge and 500-iteration bootstrap confidence intervals. All three effects turn out to be protective and statistically significant. When we compare these causal estimates against SHapley Additive exPlanations (SHAP) attribution from an Extreme Gradient Boosting (XGBoost) state estimator, a real gap appears: SHAP gets the strongest lever right but flips the ranking of the two weaker ones. A Monte Carlo simulation of the governance controller, run across a wide range of resource constraints, shows this attribution gap barely affects aggregate portfolio outcomes, provided the dominant intervention is correctly identified by both methods. That is a conditional result, and it tells us something concrete about when the choice of attribution method actually matters in practice. Throughout, we are explicit that the underlying data is cross-sectional, and we treat portfolio dynamics as a stated modeling assumption, tested through sensitivity analysis, rather than something estimated from data that cannot support it.

1. INTRODUCTION

There is something formally unusual about how most IT program portfolios are governed. At each reporting interval, performance data comes in, risk scores get generated, and governance teams have to decide which investments need escalation, which interventions to apply, and how to spread limited oversight resources across what can be dozens of concurrent programs. Strip away the terminology and this is a control problem in the classical sense. A reference trajectory exists: every investment on schedule and within budget. The current state is observable through performance metrics and model-generated risk scores. Control inputs exist too, in the form of governance interventions that can alter the trajectory of an at-risk investment. Yet, as far as we are aware, this framing has not shown up in the project management or IT governance literature.

What governance offices do looks more like proportional threshold control: a risk score crosses a line, and the investment gets escalated. We ask what a more deliberate

version of that decision problem would look like, and what it would take to build it responsibly.

One question sits at the center of this: explainable machine learning tools, SHapley Additive exPlanations (SHAP) in particular, are now routinely used to tell governance teams which project characteristics drive predicted risk. It is easy to treat those attribution scores as a ready-made intervention guide. We test that assumption directly, and the answer turns out to be partly yes and partly no. SHAP measures how a feature associates with a model's prediction. It does not measure what happens if you actually intervene on that feature. The two line up only when there is no confounding, which is rarely true of observational governance data. We estimate the real causal effect of three governance interventions using inverse probability weighting (IPW), set those effects against SHAP attribution computed from the same data, and follow the consequence of any mismatch all the way through to an actual governance decision, using a Model Predictive Control (MPC) architecture as the mechanism that makes that decision.

What we find is more specific, and more useful, than a flat verdict on whether SHAP can be trusted. SHAP correctly picks out the single most causally effective intervention in this dataset, cost variance control enforcement. It gets the other two wrong relative to each other, understating how protective a defined iteration cadence really is while overstating incremental development adoption. Run the numbers through a Monte Carlo simulation of the governance controller, though, and this particular mistake barely moves the outcome, because governance effort concentrates on the lever both methods agree matters most. That does not mean attribution method never matters. It means we now have a specific, testable condition under which it matters less than you might expect, and, by the same logic, a condition under which it would matter far more.

This paper contributes four things. First, a formally specified MPC architecture for IT portfolio governance. Second, causal estimates of governance intervention effects via IPW, built on an adjustment set we document in full and bootstrap uncertainty we report rather than assume. Third, a direct, same-data comparison between associational and causal effect sizes that shows both where they diverge and where that divergence does, and does not, change a governance decision. Fourth, an honest accounting of what cross-sectional data can and cannot support: portfolio dynamics here are a stated assumption, tested through sensitivity analysis, not a claim the data itself could ever justify.

Section 2 covers related work. Section 3 lays out the proposed methodology: control architecture, mathematical formulation, and causal gain estimation. Section 4 describes the experimental setup. Section 5 gives results. Section 6 discusses implications and limitations. Section 7 concludes.

2. RELATED WORK

Explainable machine learning shows up throughout the

project and schedule delay prediction literature, and SHAP has become the default attribution method for gradient-boosted ensembles largely because of its grounding in cooperative game theory [1, 2]. A separate literature, mostly unconnected to the first, handles causal inference for policy and treatment effect estimation, where IPW is a long-established estimator [3]. What neither literature does is show, in one applied setting, exactly where these two quantities pull apart, or build a decision architecture that actually runs on causally estimated effect sizes rather than associational ones. MPC itself is a mature tool in industrial process and systems engineering [4, 5], but its use for organizational governance decisions, as opposed to physical systems, is largely untested. We bring these three threads together in a single domain, IT portfolio governance, where each has obvious relevance on its own but has not previously been combined with the other two. A related literature on portfolio-level risk covers project interdependency and portfolio risk frameworks [6, 7], and a parallel explainable AI literature proposes multi-module frameworks for project risk transparency [8]; neither folds causally estimated intervention effects into a formal control architecture of the kind we build here.

3. PROPOSED METHODOLOGY

3.1 Problem formulation

We treat the portfolio governance decision as a discrete-time state feedback control problem. Portfolio-level schedule delay risk is the controlled variable. Governance interventions available at each reporting period make up the control input vector. A state estimator produces the current risk estimate from observed performance data, and a control law picks the intervention sequence expected to minimize cumulative portfolio risk within the governance office's resource constraints.

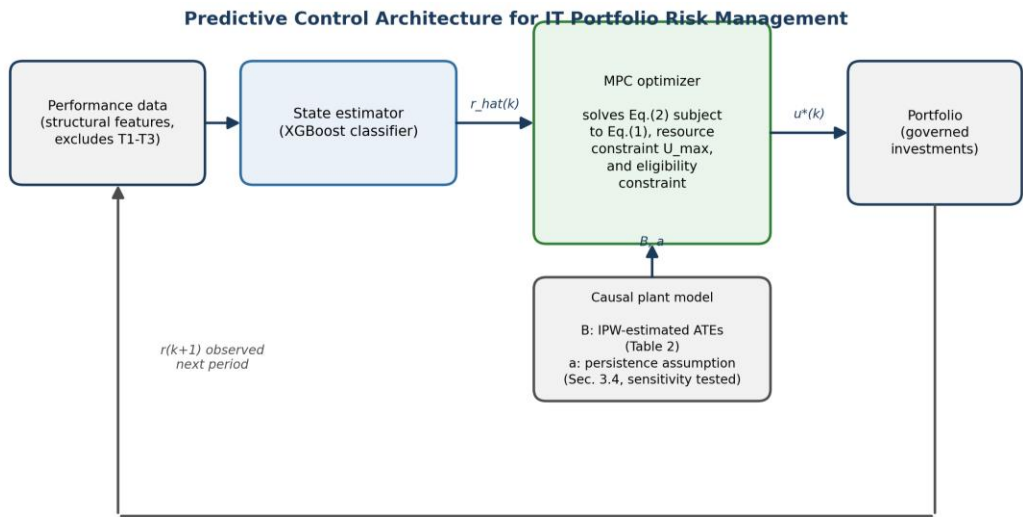


Figure 1. Predictive control architecture

Note: $r_hat(k)$: estimated risk state at period k . $u^*(k)$: optimal intervention vector. B : control gain matrix from inverse probability weighting (IPW) estimated average treatment effects (ATEs, Section 3.8). a : risk persistence parameter (Section 3.4). U_max : per-period resource constraint.

3.2 Governance interventions

Three interventions serve as control inputs, labeled T1 through T3. T1 is adopting an incremental development

methodology for investments not already using it. T2 is enforcing cost variance control, keeping absolute cost variance inside a 5 percent band. T3 is adopting a defined iteration cadence. A fourth comparison, T4, major versus standard

investment classification, is handled separately, as a structural comparison rather than a control input, since a natural classification variable has no business sitting in the same input vector as interventions someone can actually choose to apply.

3.3 System components

Three components make up the architecture. The state estimator is an Extreme Gradient Boosting (XGBoost) classifier [9] trained only on structural investment characteristics, with T1 through T3 deliberately left out, producing a probabilistic risk estimate for each investment. The plant model is a state space representation whose control gain matrix carries the IPW-estimated causal effect of each intervention. The controller runs a finite-horizon optimization over that plant model, constrained by how many interventions can actually be applied per period. Figure 1 lays out the resulting architecture.

3.4 State space representation

For a portfolio of N investments, we represent the risk dynamics as:

$$r(k+1) = a \cdot r(k) + Bu(k) + d(k)$$

Here $r(k)$ is the risk state vector at period k , $u(k)$ is the binary intervention vector, B is the control gain matrix built from the IPW-estimated ATEs in Section 3.8, $d(k)$ is a stochastic disturbance term, and a is scalar persistence parameter.

The persistence parameter a needs its own discussion. Our dataset is cross-sectional: every investment appears once, with no genuine period-over-period record of its own risk trajectory. So a cannot be, and is not, pulled from longitudinal autocorrelation, because that data simply is not there. We treat it instead as a stated modeling assumption and report results across a sensitivity range in Section 5 ($a = 0.70, 0.80, 0.90$) rather than commit to one number. Investments are modeled as decoupled, a diagonal state transition, which also means no off-diagonal coupling terms are derived from SHAP attribution. Given our own finding that SHAP does not reliably capture causal structure, using it to build part of the state matrix would have undercut the paper's own argument.

3.5 Control problem

At each period, the controller solves:

$$\min J = \mathbb{E}[r(t)Qr(t) + u(t)Ru(t)] \quad (2)$$

Subject to the dynamics in Eq. (1), a resource constraint capping applied interventions per period at a fixed budget $U_{\max}$, and an eligibility rule barring an intervention already present for a given investment from being reapplied. Here Q and R are weighting matrices that penalize portfolio risk and intervention effort, respectively, and the sum runs over the prediction horizon.

3.6 Why attribution scores are not control gains

SHAP values measure how sensitive a trained model's output is to a feature within the data it was trained on. That is not the same as measuring how the outcome actually responds if you intervene on that feature. The two line up only without

confounding, a condition that rarely holds when interventions get adopted selectively by agencies or investments that already differ systematically in other ways [10]. Recent formal work goes further and proves that certain SHAP score definitions can assign misleading importance under identifiable conditions [11]. Feature dependence adds another layer: correlated or interacting features can distort marginal attribution even when confounding is not the issue [12, 13].

3.7 Adjustment sets and identification

For each intervention, we chose a confounder set based on domain knowledge, variables plausibly affecting both treatment adoption and outcome, observable independent of the treatment decision, and not sitting on the causal pathway between treatment and outcome. Table 1 lists the treatment, outcome, and adjustment set used for each intervention. We left investment status at the time of reporting out of the adjustment set on purpose, since it is measured at the same time as the outcome and risks acting as a collider rather than a genuine confounder [14].

Table 1. Treatments, outcome, and adjustment sets

Intervention	Outcome	Adjustment Set (Confounders)
T1: incremental development adoption	At-risk indicator	Agency code; log planned cost; software indicator; major/standard classification
T2: cost variance control enforcement	At-risk indicator	Agency code; log planned cost; software indicator; major/standard classification
T3: defined iteration cadence adoption	At-risk indicator	Agency code; log planned cost; software indicator; major/standard classification

3.8 Inverse probability weighting estimation and results

We estimated propensity scores with logistic regression, class-weight balanced, on the documented adjustment set, following standard propensity score practice for adjusting confounding in observational data [15, 16]. Average treatment effects (ATEs) came from IPW, with 95 percent confidence intervals built from 500 bootstrap iterations [17]. We also computed E-values [18] to gauge how robust these effects are to unmeasured confounding. Table 2 lays out the full results alongside the mean absolute SHAP value for each intervention, from a model trained separately, specifically for this comparison.

Table 2. Causal average treatment effects (ATEs), bootstrap confidence intervals, E-values, and SHapley Additive exPlanations (SHAP) comparison

Int.	IPW ATE	95% Bootstrap CI	Sig.?	E-value (CI bound)	SHAP mean
T1	-0.0660	(-0.0970, -0.0350)	Yes	1.906	0.1845
T2	-0.1084	(-0.1351, -0.0850)	Yes	2.976	0.4077
T3	-0.0678	(-0.1104, -0.0259)	Yes	1.674	0.0816
T4	-0.0069	(-0.0293, +0.0161)	No	n/a	n/a

All three interventions show protective effects that are statistically significant. T2, cost variance control, has both the largest causal effect and the highest SHAP importance. T1 and T3 sit close together in true causal magnitude, -0.0660 and -0.0678, yet SHAP rates T1 higher (0.1845 against 0.0816), flipping their actual order. E-values between 1.67 and 2.98 mean a fairly strong unmeasured confounder would be needed to explain these effects away entirely; we report that as a robustness bound, not as proof there is no such confounder. Because these ATEs come from cross-sectional data, we treat each one as a fixed, constant gain applied at every period of the simulation described in Section 4.2, rather than as a dynamically re-estimated effect. That is a modeling simplification, stated here explicitly, and it carries the same caveat as the persistence parameter α in Section 3.4: it is a stated assumption, not a claim this dataset could support empirically.

4. EXPERIMENTAL SETUP

4.1 Dataset and state estimator

Our dataset holds 5,729 federal IT investment records from the United States Office of Management and Budget IT Dashboard [19]. The binary at-risk label comes from the schedule variance indicator. The state estimator, an XGBoost classifier, is trained only on structural features, agency, planned and projected cost, software classification, investment scale, concurrent investment count, and deliberately leaves out T1 through T3 so the state estimate does not circularly depend on the interventions we are trying to evaluate.

4.2 Monte Carlo simulation

We evaluated the MPC architecture with a Monte Carlo simulation over a 12-period horizon, on a synthetic 30-investment portfolio, initial risk states drawn from a distribution calibrated to the real test-set prevalence, and 100 independent trials per configuration. Four strategies went head to head: no intervention; a naive rule-based greedy policy that targets the highest-risk investments in a fixed default intervention order, included as a simple, realistic baseline; SHAP-guided allocation; and causal MPC allocation. We reran the whole comparison across a sweep of resource

constraints, U_{max} from 2 to 30 of 30 investments per period, to check whether the findings hold up regardless of how much governance capacity is assumed.

5. RESULTS

Table 3 and Figure 2 report terminal portfolio risk reduction relative to no intervention, across the resource-constraint sweep, at persistence $\alpha = 0.80$.

At every resource level we tested, from tightly constrained (2 of 30 investments per period) to fully open (30 of 30), the gap between SHAP-guided and causal MPC allocation stays under 0.02 percentage points, which is noise. That held across the whole sensitivity range for the persistence parameter α too. The reason is simple: T2 has both the largest true causal effect and the highest SHAP importance, so both strategies reach for it first whenever it is available. They only disagree on which of T1 or T3 comes second, and those two are close enough in true effect that getting the order wrong barely registers in the aggregate. Table 3 also shows reduction falling as the budget grows. This follows from the eligibility rule and the dynamics in Eq. (1): each investment can receive each intervention only once, and most already have it, so the pool of eligible interventions is small. A loose budget spends that pool in the first periods, and the resulting effects decay at rate α before the terminal period, whereas a tight budget spreads the same interventions across the horizon, so more of their effect remains at period 12.

Table 3. Monte Carlo terminal risk reduction across resource constraint levels (100 trials per cell)

Budget (of 30)	No Intervention	Greedy	SHAP-Guided	Causal MPC
2 (7%)	0.1056	16.1% red.	16.0% red.	16.0% red.
6 (20%)	0.1056	5.5% red.	5.5% red.	5.5% red.
10 (33%)	0.1056	4.4% red.	4.4% red.	4.4% red.
20 (67%)	0.1056	3.8% red.	3.8% red.	3.8% red.
30 (100%)	0.1056	3.8% red.	3.8% red.	3.8% red.

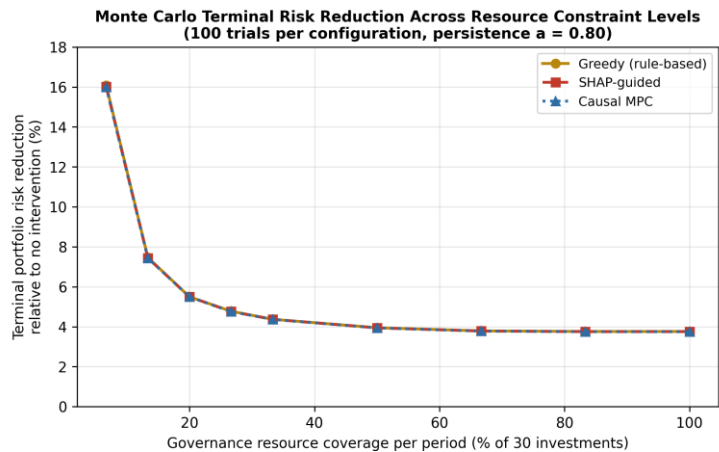


Figure 2. Terminal portfolio risk reduction relative to no intervention across the full resource constraint sweep. The greedy, SHapley Additive exPlanations (SHAP) guided, and causal Model Predictive Control (MPC) strategies are visually indistinguishable at every tested resource level

None of this contradicts the attribution gap documented in Section 3.8; it just marks out where that gap actually bites. The SHAP-versus-causal ranking error is real, and it shows up clearly at the level of individual intervention attribution. Its effect on aggregate governance outcomes here is small, because the mistake lands on a secondary lever, not the dominant one.

On the held-out test set of 860 records (84 at-risk), the state estimator posted a receiver operating characteristic area under the curve of 0.8134 and a Matthews correlation coefficient of 0.3361, with precision at 0.271 and recall at 0.690 [20]. That precision-recall balance reflects how imbalanced the underlying data is (9.8 percent at-risk prevalence), and it means a fair number of flagged investments will turn out not to be at risk at all, a real limitation for how this output should be used in practice, and one we return to in Section 6.

6. DISCUSSION AND LIMITATIONS

Three limitations are worth stating plainly. First, the data is cross-sectional, so the persistence parameter α is a modeling assumption tested through sensitivity analysis, not something we measured; anyone applying this architecture to genuinely longitudinal data should estimate portfolio dynamics directly rather than borrow the assumption used here. Second, the causal estimates depend on there being no unmeasured confounding beyond the documented adjustment set; the E-values we report quantify that risk without eliminating it, and our conclusions describe the federal IT portfolio in this dataset rather than any universal claim. Third, given the state estimator's precision-recall profile, a real deployment would generate a meaningful share of false positive risk flags, a cost that has to be weighed against the resources spent chasing them down.

The practical takeaway is conditional, not absolute. Governance offices leaning on SHAP-based attribution to decide where to intervene can expect that attribution to hold up well when it points squarely at one dominant lever, but should be more careful about SHAP-based rankings among secondary interventions with similar effect sizes, since that is exactly where this study finds attribution and true causal effect pulling apart.

7. CONCLUSIONS

We set out a predictive control architecture for IT portfolio governance, parameterized with causally estimated rather than associational intervention effects. Using real IPW estimation, a documented adjustment set, and bootstrap uncertainty quantification, we found three governance interventions with statistically significant protective effects. Set against SHAP-based attribution, the picture that emerges is a real but conditional divergence: SHAP gets the dominant lever right and the secondary ones wrong, and that divergence has almost no bearing on aggregate governance outcomes precisely when the dominant lever is the one correctly identified, a result that held across every resource constraint we tested. The natural next step is extending this architecture to genuinely longitudinal portfolio data, where the persistence assumption used here could give way to a directly estimated dynamic model.

ACKNOWLEDGMENT

The authors thank the Department of Computer Science, Babcock University, for institutional support during the conduct of this research.

REFERENCES

- [1] Lundberg, S.M., Lee, S.I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30: 4765-4774.
- [2] Lundberg, S.M., Erion, G., Chen, et al. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence*, 2(1): 56-67. <https://doi.org/10.1038/s42256-019-0138-9>
- [3] Imbens, G.W., Rubin, D.B. (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences*. Cambridge University Press, Cambridge.
- [4] Camacho, E.F., Bordons, C. (2007). *Model Predictive Control*. Springer, London.
- [5] Mayne, D.Q., Rawlings, J.B., Rao, C.V., Scokaert, P.O.M. (2000). Constrained model predictive control: Stability and optimality. *Automatica*, 36(6): 789-814. [https://doi.org/10.1016/S0005-1098\(99\)00214-9](https://doi.org/10.1016/S0005-1098(99)00214-9)
- [6] Bai, L., Shi, H., Kang, S., Zhang, B. (2023). Project portfolio risk analysis with the consideration of project interdependencies. *Engineering, Construction and Architectural Management*, 30(2): 647-670. <https://doi.org/10.1108/ECAM-06-2021-0555>
- [7] Micán, C., Fernandes, G., Araújo, M. (2024). Towards a comprehensive framework for risk assessment of organizational development project portfolios. *International Journal of Information Systems and Project Management*, 12(3): 50-69. <https://doi.org/10.12821/ijispm120303>
- [8] Badhon, B., Chakraborty, R.K., Anavatti, S.G., Vanhoucke, M. (2025). A multi-module explainable artificial intelligence framework for project risk management: Enhancing transparency in decision-making. *Engineering Applications of Artificial Intelligence*, 148: 110427. <https://doi.org/10.1016/j.engappai.2025.110427>
- [9] Chen, T., Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, USA, pp. 785-794. <https://doi.org/10.1145/2939672.2939785>
- [10] Janzing, D., Minorics, L., Blöbaum, P. (2020). Feature relevance quantification in explainable AI: A causal problem. In *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, PMLR, 108: 2907-2916.
- [11] Huang, X., Marques-Silva, J. (2024). On the failings of Shapley values for explainability. *International Journal of Approximate Reasoning*, 171: 109112. <https://doi.org/10.1016/j.ijar.2023.109112>
- [12] Herbringer, J., Wright, M.N., Nagler, T., Bischl, B., Casalicchio, G. (2024). Decomposing global feature effects based on feature interactions. *Journal of Machine Learning Research*, 25(381): 1-65.
- [13] Aas, K., Jullum, M., Løland, A. (2021). Explaining individual predictions when features are dependent: More accurate approximations to Shapley values.

- Artificial Intelligence, 298: 103502.
<https://doi.org/10.1016/j.artint.2021.103502>
- [14] Cole, S.R., Platt, R.W., Schisterman, E.F., et al. (2010). Illustrating bias due to conditioning on a collider. *International Journal of Epidemiology*, 39(2): 417-420. <https://doi.org/10.1093/ije/dyp334>
- [15] 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>
- [16] Austin, P.C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research*, 46(3): 399-424. <https://doi.org/10.1080/00273171.2011.568786>
- [17] Efron, B., Tibshirani, R.J. (1993). *An Introduction to the Bootstrap*. Chapman & Hall, New York.
- [18] VanderWeele, T.J., Ding, P. (2017). Sensitivity analysis in observational research: Introducing the E-value. *Annals of Internal Medicine*, 167(4): 268-274. <https://doi.org/10.7326/M16-2607>
- [19] Office of Management and Budget. IT Dashboard. <https://www.itdashboard.gov/>.
- [20] Chicco, D., Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, 21: 6. <https://doi.org/10.1186/s12864-019-6413-7>

NOMENCLATURE

$r(k)$	portfolio risk state vector at period k
$u(k)$	binary governance intervention vector
B	control gain matrix (IPW-estimated ATEs)
a	risk persistence parameter (modeling assumption)
$d(k)$	stochastic disturbance term
ATE	average treatment effect
IPW	inverse probability weighting
MPC	model predictive control
SHAP	SHapley Additive exPlanations
XGBoost	Extreme Gradient Boosting