The Geometry of Econometric Models: A Proposal

Xinyuan Lyu

November 30, 2025

Abstract

The integration of Machine Learning into economics has generated distinct methodological lineages, ranging from semi-parametric causal inference to deep learning-based structural estimation. However, the field lacks a unified meta-theory to rigorously compare these approaches and analyze their out-of-distribution generalization properties. This proposal establishes a geometric framework for economic modeling, conceptualizing estimation strategies as contraction operators on a metric model space. We introduce the Golden Set—the intersection of optimal models across all admissible contexts—as a robust geometric definition of invariant truth. Within this framework, we derive a novel decomposition of generalization error that explicitly accounts for context shifts. Furthermore, we formalize the epistemic tension between *Conservatism* (prioritizing robustness via partial identification and weaker assumptions) and *Activism* (prioritizing precision via stronger assumptions), deriving an optimal selection criterion based on in-sample suboptimality. Finally, we demonstrate the practical utility of these geometric principles through Economics-Informed Neural Networks (EINNs), which implement robust, set-valued identification.

1 Introduction

Quantitative modeling serves as the epistemological backbone of modern economic inquiry and quantitative marketing. Whether characterizing consumer heterogeneity in demand systems, pricing complex financial assets, or designing counterfactual policies in dynamic environments, the discipline relies on the ability to map observable data to underlying structural primitives. Traditionally, this mapping was constrained by the need for analytical tractability and parsimony, limiting researchers to sparse, parametric specifications. However, the recent integration of Machine Learning and Deep Learning has fundamentally reshaped this landscape, offering flexible, high-dimensional, and data-driven representations that challenge established methodological paradigms.

The assimilation of these algorithmic tools into economics has been rapid yet fragmented, evolving through several distinct intellectual lineages. Initially, the discourse focused on delineating the legitimate scope of ML regarding causal inference. Mullainathan and Spiess (2017) and Kleinberg et al. (2015) articulated a fundamental dichotomy between parameter estimation and outcome prediction, legitimizing "black-box" models for a specific class of "Prediction Policy Problems" distinct from traditional causal questions. This perspective was broadened by Athey (2018), who characterized ML as a "General Purpose Technology" capable of reshaping the research production function itself.

As the field moved beyond this initial demarcation, a central debate emerged regarding the nature of economic signals in high-dimensional settings. While early approaches relied on "sparsity" assumptions to apply Lasso-type regularization (Belloni et al., 2014), recent work has increasingly challenged this paradigm. Giannone et al. (2021) argue for the "illusion of sparsity" in macroeconomic contexts, advocating for dense modeling approaches that are theoretically congruent with

Deep Learning. This "dense" signal hypothesis has found strong empirical support in asset pricing, where Gu et al. (2020) demonstrate that neural networks significantly outperform linear factor models. Furthermore, the definition of economic data itself has expanded, with Gentzkow et al. (2019) formalizing frameworks to treat unstructured text as high-dimensional economic signals, albeit introducing unique non-classical measurement errors (Wei and Malik (2025)).

To reconcile these flexible representations with the rigor of economic inference, distinct methodological families have emerged. The *Semi-Parametric* lineage, exemplified by the Double/Debiased Machine Learning (DML) framework (Chernozhukov et al., 2018), utilizes Neyman Orthogonality to immunize estimation against nuisance bias. This lineage has matured to include rigorous asymptotic bounds for model with deep neural networks (Farrell et al., 2021, 2025) and tree-based inference for heterogeneous treatment effects via Causal Forests (Wager and Athey, 2018; Athey et al., 2019).

Parallel to estimation, a *Structural* lineage has leveraged DL to break the curse of dimensionality in dynamic equilibrium models. By exploiting the implicit regularization of stochastic gradient descent, researchers have successfully approximated value functions in high-dimensional state spaces (Fernández-Villaverde, 2025), solved nonlinear HJB equations in continuous-time finance (Wu et al., 2025), and handled complex lifetime reward maximization problems (Maliar et al., 2021). More radical departures from tradition include the *Direct Estimation* lineage, which bypasses intractable likelihoods by learning inverse maps via simulation (Wei and Jiang, 2024, 2025) or adversarial formulations (Kaji et al., 2023; Lewis and Syrgkanis, 2018) to use the DNN estimate the parameters of structual models.

Finally, in domains such as quantitative marketing and industrial organization, a *Representation* lineage has sought to embed economic axioms directly into neural architectures. This includes enforcing permutation invariance for multi-product demand (Singh et al., 2025), hard-coding Independence of Irrelevant Alternatives (IIA) into network connectivity (Wang et al., 2020), approximating Random Utility Models via RUMnets (Aouad and Désir, 2022), or reframing causal inference as matrix completion (Athey et al., 2021).

Despite this vibrant methodological pluralism, the literature lacks a unified meta-theoretical framework to rigorously compare these diverse strategies. We face a "Tower of Babel" scenario: a DML estimator, a structural HJB solver, and a representation learning architecture are typically analyzed using disjoint mathematical vocabularies—asymptotic statistics, numerical analysis, and approximation theory, respectively. Crucially, we lack a coherent language to discuss the trade-offs between *model complexity*, *structural assumptions*, and *cross-context robustness*. When a researcher chooses a highly structured "activist" model over a flexible "conservative" one, what exactly is being traded in terms of geometric distance to the truth? How does a model trained in one market (context) generalize to another when the underlying data-generating process shifts?

This proposal addresses these questions by proposing a unified geometric theory of economic modeling. We abstract away from specific algorithms to treat modeling strategies as geometric contraction operators on a metric space. Our framework introduces the concept of the **Golden Set**—the intersection of optimal models across all admissible contexts—as a robust geometric definition of "truth."

Within this framework, we make three primary contributions. First, we provide a formal decomposition of out-of-distribution generalization error into statistical, limited-information, and approximation components. Second, we formalize the tension between two dominant research paradigms: *Conservatism* (prioritizing the inclusion of invariant structure, akin to partial identification) and *Activism* (prioritizing precision via strong assumptions, akin to structural point estimation). We derive a criterion based on "in-sample suboptimality" to optimally select between these paradigms. Finally, we demonstrate the practical utility of this theory by introducing Economics-Informed Neural Networks (EINNs), a computational framework that operationalizes our geometric principles to achieve robust, set-valued identification.

The remainder of the proposal is organized as follows. Section 2 establishes the geometric foundations, defining the model space, contexts, and the common geometry assumption that underpins our analysis. Section 3 introduces the Golden Set, characterizing the geometric nature of invariant truth and its relationship to structural invariance. Section 4 formalizes modeling strategies as contraction operators composed of feasibility and optimality steps. Section 5 presents our main theoretical result: a geometric decomposition of generalization error into statistical, approximation, and limited-information components. Section 6 analyzes the strategic trade-off between Conservatism and Activism, deriving the optimal selection criterion based on in-sample suboptimality and contextual conditioning. Finally, we conclude with a discussion of computational implementations via Economics-Informed Neural Networks.

2 The Geometric Foundations

In this section we endow the abstract architecture of models, worlds, and metrics with an explicit geometric structure. The central idea is that the space of models carries a *common* underlying geometry, while each context—a pair of world and evaluation functional—induces its own distortion of this base geometry. Optimality and generalization can then be studied as stability properties of sublevel sets under changes of context.

2.1 The Model Space as a Base Metric Space

We begin by specifying the primitive geometric structure on the space of models. Throughout, we write

$$(\mathcal{M}, d_{\mathcal{M}})$$

for a metric space, where $d_{\mathcal{M}}$ represents the *base geometry* of the model space.

Definition 2.1 (Model Space with Base Geometry)**.** The *model space* is a nonempty set $\mathcal{M}$ equipped with a metric

$$d_{\mathcal{M}} : \mathcal{M} \times \mathcal{M} \to [0, \infty)$$

satisfying, for all $m_a, m_b, m_c \in \mathcal{M}$:

(i) $d_{\mathcal{M}}(m_a, m_b) \geq 0$ and $d_{\mathcal{M}}(m_a, m_b) = 0$ if and only if $m_a = m_b$;

(ii) $d_{\mathcal{M}}(m_a, m_b) = d_{\mathcal{M}}(m_b, m_a)$;

(iii) $d_{\mathcal{M}}(m_a, m_c) \leq d_{\mathcal{M}}(m_a, m_b) + d_{\mathcal{M}}(m_b, m_c)$.

We refer to $(\mathcal{M}, d_{\mathcal{M}})$ as the *base metric space.*

Remark 2.1. The metric $d_{\mathcal{M}}$ is intended to encode the intrinsic notion of "closeness" between models, independent of any particular world or evaluation functional. No linear or parametric structure is assumed: $\mathcal{M}$ need not be a vector space, and models need not be described by a fixed finite-dimensional parameter. This allows the framework to accommodate both parametric and nonparametric representations, including neural-network parameterizations and implicit models.

In applications it is often convenient to impose standard regularity conditions on the base geometry; for instance, one may assume that $(\mathcal{M}, d_{\mathcal{M}})$ is complete and separable. We will not need these additional properties at this stage, and only require them when explicitly stated later.

2.2 Contexts and Contextual Metrics

The performance of a model is always evaluated *in context*: performance depends on both the state of the world and the chosen evaluation criterion. We formalize this by introducing a context space and the associated contextual metrics.

Definition 2.2 (Contexts and Evaluation Functionals)**.** A *context* is a pair

$$C = (\omega, \Phi),$$

where $\omega \in \mathcal{W}$ is a world and $\Phi : \mathcal{M} \times \mathcal{W} \to \mathcal{V}$ is an evaluation functional. The *context space* is

$$\mathcal{C} := \mathcal{W} \times \{\Phi\},$$

where each admissible evaluation functional Φ is regarded as a fixed element of the second component. For a given context $C = (\omega, \Phi)$ we define the induced performance map

$$\Phi_C : \mathcal{M} \to \mathcal{V}, \qquad \Phi_C(m) := \Phi(m, \omega).$$

Given a context $C = (\omega, \Phi)$, differences in performance are measured by a context-dependent *regret distance*.

Definition 2.3 (Contextual Regret Metric)**.** For a context $C = (\omega, \Phi)$, the associated *contextual regret metric* is the map

$$\Delta_C : \mathcal{M} \times \mathcal{M} \to [0, \infty), \qquad \Delta_C(m_a, m_b) := \big|\Phi_C(m_a) - \Phi_C(m_b)\big|.$$

By construction, each Δ_C is a pseudometric on $\mathcal{M}$. In many applications it is natural to assume that Φ_C is injective (up to observational equivalence), in which case Δ_C is a genuine metric. For the purposes of this section, it suffices to note that each context supplies its own way of measuring distances between models, based on their performance in that context.

2.3 The Common Geometry Assumption

The central geometric requirement of our framework is that all context-dependent regret metrics arise as controlled distortions of a common underlying geometry on $\mathcal{M}$.

Assumption 2.1 (Common Geometry)**.** There exists a base metric $d_{\mathcal{M}}$ on $\mathcal{M}$ such that for every context $C \in \mathcal{C}$ there exist finite constants $0 < \alpha_C \le \beta_C < \infty$ satisfying

$$\alpha_C \, d_{\mathcal{M}}(m_a, m_b) \;\le\; \Delta_C(m_a, m_b) \;\le\; \beta_C \, d_{\mathcal{M}}(m_a, m_b) \qquad \text{for all } m_a, m_b \in \mathcal{M}.$$

Remark 2.2 (Interpretation)**.** Assumption 2.1 states that each contextual metric Δ_C is *bi-Lipschitz equivalent* to the same base geometry $d_{\mathcal{M}}$. Different contexts may stretch or compress distances, but only within bounded factors: models close under $d_{\mathcal{M}}$ cannot be assigned arbitrarily large regret in any context, and models far apart cannot be made arbitrarily close. Thus all contextual metrics share the same topology and differ only by controlled geometric distortions.

The constants α_C and β_C describe the *conditioning* of context C with respect to the base geometry and provide a natural measure of geometric amplification or attenuation when mapping $d_{\mathcal{M}}$ into Δ_C.

Proposition 2.1 (Cross-Context Comparison)**.** Suppose Assumption 2.1 holds. Let $C, C' \in \mathcal{C}$ be two contexts with distortion constants (α_C, β_C) and $(\alpha_{C'}, \beta_{C'})$. Then for all $m_a, m_b \in \mathcal{M}$,

$$\Delta_{C'}(m_a, m_b) \ \leq \ \kappa(C \to C')\, \Delta_C(m_a, m_b), \qquad \kappa(C \to C') := \frac{\beta_{C'}}{\alpha_C}.$$

Proof. By Assumption 2.1,

$$\Delta_{C'}(m_a, m_b) \leq \beta_{C'}\, d_{\mathcal{M}}(m_a, m_b), \qquad \Delta_C(m_a, m_b) \geq \alpha_C\, d_{\mathcal{M}}(m_a, m_b).$$

If $\Delta_C(m_a, m_b) = 0$ the claim is trivial. Otherwise,

$$\Delta_{C'}(m_a, m_b) \leq \frac{\beta_{C'}}{\alpha_C}\, \Delta_C(m_a, m_b),$$

as required. □

Remark 2.3. The quantity $\kappa(C \to C')$ functions as a *contextual condition number*: it measures the maximal possible amplification of regret distances when transporting performance comparisons from context C to context C'. In later sections, these condition numbers determine how estimation errors and approximation errors—naturally measured in their native contexts—propagate to new or out-of-distribution contexts.

2.4 Optimal Disks as Sublevel Sets

Within a fixed context, the notion of optimality is encoded by the sublevel sets of the performance map Φ_C. These sets provide the fundamental geometric objects upon which our subsequent analysis is built.

Definition 2.4 (ϵ-Optimal Sets and Optimal Disks)**.** For a context $C = (\omega, \Phi)$ and tolerance $\epsilon \geq 0$, the associated *ϵ-optimal set* is

$$S_\epsilon(C) := \big\{ m \in \mathcal{M} : \Phi_C(m) \leq \inf_{m' \in \mathcal{M}} \Phi_C(m') + \epsilon \big\}.$$

The set $S_0(C)$ is called the *optimal disk* of context C.

Remark 2.4. The terminology "disk" is deliberately geometric: each context selects a region of the model space consisting of models that are exactly optimal (for $\epsilon = 0$) or approximately optimal (for $\epsilon > 0$) under that world–metric pair. As the context varies over the admissible collection $\mathcal{C}$, these disks move through $\mathcal{M}$, changing in location and shape. The global structure formed by this family of disks encodes how context-dependent optimality is distributed across the model space.

The sublevel sets $\{S_\epsilon(C)\}$ inherit regularity from the evaluation functional Φ_C. For example, if Φ_C is lower semicontinuous with respect to $d_{\mathcal{M}}$, then each $S_\epsilon(C)$ is closed in the base metric topology. Additional structure can be imposed when needed; for instance, if Φ_C is convex on a linearly-structured $\mathcal{M}$, then $S_\epsilon(C)$ is convex.

2.5 Set-Valued Geometry on the Model Space

To study identification, generalization, and robustness, we require a notion of distance not only between individual models, but also between sets of models. This leads us naturally to the Hausdorff metric on subsets of $\mathcal{M}$.

Definition 2.5 (Hausdorff Distance)**.** Let $\mathcal{K}(\mathcal{M})$ denote the collection of all nonempty closed subsets of $\mathcal{M}$. For $A, B \in \mathcal{K}(\mathcal{M})$, the *Hausdorff distance* induced by $d_{\mathcal{M}}$ is

$$d_H(A,B) := \max\Big\{ \sup_{a\in A}\inf_{b\in B} d_{\mathcal{M}}(a,b), \ \ \sup_{b\in B}\inf_{a\in A} d_{\mathcal{M}}(a,b) \Big\}.$$

Remark 2.5. The Hausdorff distance d_H lifts the base geometry $d_{\mathcal{M}}$ from points to closed sets. In particular, we will use d_H to quantify:

- the convergence of estimated sets to identified sets;
- the stability of optimal disks $S_\epsilon(C)$ under changes of context C;
- the diameter and shape of intersections of optimal disks, such as the Golden Set introduced in the next section.

In this way, the geometric structure on $\mathcal{M}$ induces a corresponding geometric structure on families of models.

The ingredients assembled in this section—a base metric on the model space, a family of contextual distortions, optimal disks as sublevel sets, and the Hausdorff geometry on closed subsets—provide the foundational apparatus for our subsequent analysis. In the next section, we exploit this apparatus to introduce a unified geometric notion of truth, the *Golden Set*, and to study its stability properties across contexts.

3 A Unified Geometric Notion of Truth: The Golden Set

The geometric foundations introduced in previous section permit a unified and structurally transparent notion of "truth" within a world–model system. Rather than presuming the existence of a single, point-valued true model, we show that the correct object is a *geometric fixed set*: the set of models that remain optimal across all admissible contexts. This invariant set—the *Golden Set*—is the natural generalization of structural invariance, stable mechanisms, and context-independent optimality.

3.1 Optimal Disks Across Contexts

Each context $C \in \mathcal{C}$ induces an optimal disk (Definition 2.4)

$$S_0(C) = \arg\min_{m\in\mathcal{M}} \Phi_C(m),$$

possibly containing multiple models. As C varies over the admissible context space $\mathcal{C} = \mathcal{W} \times \{\Phi\}$, the collection

$$\big\{S_0(C) : C \in \mathcal{C}\big\}$$

forms a family of moving geometric objects in $\mathcal{M}$. These disks change position and shape as the world ω, the metric Φ, or both vary.

Remark 3.1 (Geometric Interpretation)**.** Each $S_0(C)$ is the locus of models that are exactly optimal under the performance map Φ_C. The family $\{S_0(C)\}$ thus represents all context-dependent optimality relations: every context selects a "best" region in the model space, and the collection of such regions encodes how optimality shifts across worlds and evaluation criteria.

3.2 The Golden Set

The key insight is that, while individual optimal disks vary across contexts, the *intersection* of all optimal disks yields a canonical object that captures context-invariant optimality.

Definition 3.1 (Golden Set)**.** The *Golden Set* is the set of models that are optimal in every admissible context:

$$M_{\text{real}} := \bigcap_{C \in \mathcal{C}} S_0(C) \;\subseteq\; \mathcal{M}.$$

Remark 3.2 (Conceptual Meaning)**.** An element $m \in M_{\text{real}}$ is simultaneously optimal in all worlds and under all evaluation metrics. Thus M_{real} represents the strongest possible notion of *invariance* in the model space: models in M_{real} are those whose performance remains optimal regardless of context.

We emphasize that M_{real} accommodates multi-valued truth. Even if no single model is universally optimal, the Golden Set may be a nondegenerate geometric object (a region, manifold, or convex body) whose diameter measures the degree of permissible heterogeneity among universally optimal models.

3.3 Golden-Set Interpretation of Invariance

The definition of M_{real} unifies several concepts that arise in economics, statistics, and machine learning:

- *Structural invariance*: If a model family contains a true mechanism that does not change across environments, the set of models that encode this mechanism is precisely the intersection of environment-specific optimal disks.
- *Stable features in domain generalization*: Invariant predictors across environments correspond to elements lying in the intersection of all environment-wise risk minimizers.
- *Robust decision theory*: A decision rule that minimizes regret across all admissible states of the world lies in M_{real}.
- *Partial identification*: If truth is not point-identified, the identified set corresponds to a subset of M_{real} for an appropriate restriction of the context space.

Thus the Golden Set simultaneously generalizes “truth” in structural econometrics, “invariance” in machine learning, and “robust optimality” in decision theory.

3.4 Non-Vacuity and Stability of the Golden Set

The definition of M_{real} raises two fundamental questions:

(i) Under what conditions is M_{real} nonempty?

(ii) When is M_{real} stable under perturbations of the context space $\mathcal{C}$?

Both questions admit clean geometric answers when the sublevel sets $S_0(C)$ satisfy appropriate regularity conditions.

Proposition 3.1 (Finite-Intersection Sufficiency)**.** Suppose that for every finite subset $\{C_1, \ldots, C_k\} \subset \mathcal{C}$, the intersection

$$S_0(C_1) \cap \cdots \cap S_0(C_k)$$

is nonempty. Then the Golden Set M_{real} is nonempty.

Remark 3.3. Proposition 3.1 is a Helly-type condition: nonemptiness of all finite intersections suffices for nonemptiness of the infinite intersection. Such conditions are natural when the optimal disks $\{S_0(C)\}$ enjoy geometric regularity such as convexity or closedness.

In addition to nonemptiness, the *diameter* of M_{real} plays a central role in identification and generalization.

Definition 3.2 (Diameter of the Golden Set)**.** The geometric diameter of M_{real} is

$$\operatorname{diam}(M_{\text{real}}) := \sup_{m_a, m_b \in M_{\text{real}}} d_{\mathcal{M}}(m_a, m_b).$$

Large diameter indicates substantial ambiguity in universally optimal models, whereas a small diameter corresponds to near point-identification of the context-invariant structure.

Proposition 3.2 (Stability Under Context Perturbations)**.** Suppose Assumption 2.1 holds and that Φ_C is Lipschitz in the context variable C with respect to a metric $d_{\mathcal{C}}$ on the context space. Then for any two contexts $C, C' \in \mathcal{C}$,

$$d_H\big(S_0(C), S_0(C')\big) \;\le\; L\, d_{\mathcal{C}}(C, C'),$$

for some constant $L < \infty$ independent of C, C'. Consequently, if $\mathcal{C}$ is compact under $d_{\mathcal{C}}$ then M_{real} has finite diameter.

Remark 3.4. Proposition 3.2 shows that optimal disks vary *continuously* with the context. Thus the Golden Set inherits geometric regularity from the continuity of Φ_C in C. This result is the geometric foundation of both identification (stability with respect to sampling noise) and out-of-distribution generalization (stability with respect to environment shifts).

3.5 Golden Set as the Invariant Core of Modeling

The Golden Set provides a conceptually and mathematically unified notion of truth across all admissible contexts:

- It is the unique subset of $\mathcal{M}$ that is immune to context variation.
- It encapsulates all invariant causal or structural components of the model space.
- It characterizes the maximal robust set of models whose performance is optimal across all environments.

In this way, the Golden Set serves as the geometric anchor for the entire theoretical framework. Subsequent sections build upon this construction: modeling strategies are interpreted as operators that contract the model space toward context-specific disks, identification corresponds to convergence toward the Golden Set, and regret analysis decomposes distance to the Golden Set along geodesic paths in the model space under multiple metrics.

4 Modeling Strategies as Geometric Contraction Operators

The geometric framework developed so far suggests that modeling is not a process of constructing a single representation but rather a systematic elimination of models that violate feasibility, optimality, or rationality requirements. These eliminations take the mathematical form of *geometric contractions* of $(\mathcal{M}, d_{\mathcal{M}})$, producing (possibly set-valued) outputs that represent the admissible models under a given strategy in a given context.

4.1 Feasibility Constraints as First-Stage Contractions

The first form of contraction arises from structural, disciplinary, institutional, or general assumption restrictions.

Definition 4.1 (Feasibility Operator)**.** A *feasibility operator* is a correspondence

$$\mathcal{F}_{\mathcal{S}} : \mathcal{C} \longrightarrow 2^{\mathcal{M}} \setminus \{\emptyset\},$$

such that for each context $C = (\omega, \Phi)$, the set $\mathcal{F}_{\mathcal{S}}(C)$ consists of all models deemed feasible under the strategy $\mathcal{S}$. We assume $\mathcal{F}_{\mathcal{S}}(C)$ is closed in the base metric $d_{\mathcal{M}}$.

Remark 4.1. Feasibility restrictions include equilibrium conditions, regularity constraints, shape constraints, and economic or scientific assumptions. The feasibility operator performs a first-stage geometric contraction:

$$(\mathcal{M}, d_{\mathcal{M}}) \longrightarrow (\mathcal{F}_{\mathcal{S}}(C), d_{\mathcal{M}}).$$

4.2 Optimality Contraction Within the Feasible Set

Optimality must be understood as *constrained optimality*. The unconstrained optimal disk $S_0(C)$ represents the global minimizers of the evaluation functional on the entire model space. However, feasibility restrictions may exclude these global minimizers.

Thus optimality must be defined *within* the feasible region.

Definition 4.2 (Optimality Operator)**.** Given a feasibility operator $\mathcal{F}_{\mathcal{S}}$, the associated *optimality operator* is the correspondence

$$\mathcal{O}_{\mathcal{S}} : \mathcal{C} \longrightarrow 2^{\mathcal{M}},$$

defined by

$$\mathcal{O}_{\mathcal{S}}(C) := \arg\min_{m \in \mathcal{F}_{\mathcal{S}}(C)} \Phi_C(m),$$

Remark 4.2. In general, the unconstrained optimal disk $S_0(C)$ may fail to intersect the feasible set:

$$S_0(C) \cap \mathcal{F}_{\mathcal{S}}(C) = \emptyset.$$

This is not pathological. It simply means that the strategy imposes restrictions strong enough that global minima may lie outside the admissible region. The constrained optimum may still perform well across other contexts and need not be far from the Golden Set.

Under mild regularity—compactness of $\mathcal{F}_{\mathcal{S}}(C)$ and continuity of Φ_C—the optimality operator always produces a nonempty closed set.

4.3 The Strategy Operator

Definition 4.3 (Strategy Operator)**.** The *strategy operator* is the correspondence

$$\mathcal{T}_{\mathcal{S}} : \mathcal{C} \longrightarrow 2^{\mathcal{M}},$$

defined by

$$\mathcal{T}_{\mathcal{S}}(C) := \mathcal{O}_{\mathcal{S}}(C).$$

Remark 4.3 (Two-Stage Contraction Structure)**.** Modeling consists of:

$$\text{(i) feasibility contraction} \quad \to \quad \text{(ii) constrained optimality contraction.}$$

Only after feasibility has collapsed the model space to a closed admissible region does optimality select the best-performing elements of this region for the given context.

4.4 Convergence and Estimation as Geometric Limits

Let $d_\omega \in \mathcal{D}$ denote a sample drawn from the training world ω.

Definition 4.4 (Sample and Population Estimator Sets)**.**

$$\widehat{M} := \mathcal{T}_{\mathcal{S}}(d_\omega, \Phi), \qquad M_\infty := \mathcal{T}_{\mathcal{S}}(\omega, \Phi).$$

Remark 4.4. The sample-based set $\widehat{M}$ is a random contraction of $\mathcal{M}$ induced by the finite sample d_ω, while M_∞ is the corresponding population-limit contraction under the true world ω_{old}. Geometric convergence of the estimator means

$$d_H(\widehat{M}, M_\infty) \to 0$$

in an appropriate probabilistic sense as the sample size grows.

Definition 4.5 (Geometric Estimation Regimes)**.** We say that the estimation problem is:

- *point-estimating* if $\operatorname{diam}(M_\infty) = 0$;
- *set-estimating* otherwise, i.e. if $\operatorname{diam}(M_\infty) > 0$.

4.5 Relation to the Golden Set and Multi-Context Optimality

The Golden Set consists of models optimal in *every* context:

$$M_{\text{real}} := \bigcap_{C \in \mathcal{C}} S_0(C).$$

A strategy may or may not include M_{real} in its feasible sets. Strategies that do are "Golden-compatible":

Definition 4.6 (Golden-Compatible Strategy)**.** A strategy $\mathcal{S}$ is *Golden-compatible* if

$$M_{\text{real}} \subseteq \mathcal{F}_{\mathcal{S}}(C) \qquad \text{for all } C \in \mathcal{C}.$$

Proposition 4.1. If $\mathcal{S}$ is Golden-compatible and constrained minimizers exist, then $M_{\text{real}} \subseteq \mathcal{T}_{\mathcal{S}}(C)$ for all C.

Remark 4.5. Even when the feasible set excludes the Golden Set, the resulting constrained-optimal models may still perform well across multiple contexts. A key geometric object is the *multi-context ϵ-optimal set*, defined by

$$S_\epsilon(\mathcal{K}) := \bigcap_{C \in \mathcal{K}} S_\epsilon(C),$$

for a finite (or compact) collection of contexts $\mathcal{K}$. A constrained-optimal solution may belong to such an intersection even when it excludes the Golden Set, yielding robustness across multiple contexts despite feasibility-induced misspecification.

In summary, modeling strategies are geometric contraction operators, consisting of feasibility contraction followed by constrained optimality contraction. They generate set-valued outputs whose geometry describes identification, generalization, and robustness. Golden compatibility is not required; even feasibility-restricted models may lie in multi-context ϵ-optimal regions and thus exhibit strong cross-context performance.

5 Regret Decomposition and Cross-Context Generalization

The geometric framework enables a rigorous decomposition of out-of-distribution generalization error. We study the performance gap when a model trained in one context is deployed in another, revealing fundamental trade-offs between statistical precision, context alignment, and approximation capability.

5.1 Problem Formulation

Consider a strategy $\mathcal{S}$ applied to data $d_{\omega_{\text{old}}} \in \mathcal{D}$ sampled from world $\omega_{\text{old}} \in \mathcal{W}$ with evaluation functional Φ_{old}. Let $C_{\text{old}} = (\omega_{\text{old}}, \Phi_{\text{old}})$ denote the training context and $C_{\text{new}} = (\omega_{\text{new}}, \Phi_{\text{new}})$ the deployment context.

Definition 5.1 (Estimator and Benchmark Sets)**.** The *sample estimator set* produced by strategy $\mathcal{S}$ is

$$\widehat{M} := \mathcal{T}_{\mathcal{S}}(d_{\omega_{\text{old}}}, \Phi_{\text{old}}).$$

The *unconstrained optimal set* in the new context is

$$M^*_{\text{new}} := S_0(C_{\text{new}}).$$

Performance is measured by the worst-case distance from estimated to optimal models under the new context metric.

Definition 5.2 (Worst-Case Regret)**.** The *worst-case regret* of estimator set $\widehat{M}$ relative to benchmark M^*_{new} is

$$\mathcal{R}(\widehat{M}, M^*_{\text{new}} \mid C_{\text{new}}) := \sup_{m \in \widehat{M}} \inf_{m^* \in M^*_{\text{new}}} \Delta_{C_{\text{new}}}(m, m^*),$$

where $\Delta_{C_{\text{new}}}(m, m^*) = |\Phi_{C_{\text{new}}}(m) - \Phi_{C_{\text{new}}}(m^*)|$ is the contextual regret metric.

5.2 The Mediating Sets

Regret decomposes along a geometric path through two intermediate objects.

Definition 5.3 (Population and Constrained-Optimal Sets)**.** The *population limit set* under strategy $\mathcal{S}$ in the old context is

$$M_\infty := \mathcal{T}_\mathcal{S}(\omega_{\text{old}}, \Phi_{\text{old}}).$$

The *constrained-optimal set* when strategy $\mathcal{S}$ is applied directly to the new context is

$$M_{\text{med}} := \mathcal{T}_\mathcal{S}(C_{\text{new}}) = \mathcal{O}_\mathcal{S}(C_{\text{new}}).$$

These sets induce the decomposition path

$$\widehat{M} \longrightarrow M_\infty \longrightarrow M_{\text{med}} \longrightarrow M^*_{\text{new}}.$$

5.3 Main Decomposition Theorem

Theorem 5.1 (Geometric Regret Decomposition)**.** Suppose Assumption 2.1 holds. Then

$$\begin{aligned}
\mathcal{R}(\widehat{M}, M^*_{\text{new}} \mid C_{\text{new}}) \le \beta_{C_{\text{new}}} \cdot d_H(\widehat{M}, M_\infty) \quad &(1)\\
+ \beta_{C_{\text{new}}} \cdot d_H(M_\infty, M_{\text{med}}) \quad &(2)\\
+ \sup_{m \in M_{\text{med}}} \inf_{m^* \in M^*_{\text{new}}} \Delta_{C_{\text{new}}}(m, m^*), \quad &(3)
\end{aligned}$$

where $\beta_{C_{\text{new}}}$ is the upper distortion constant from Assumption 2.1.

Proof. Let $\hat{m} \in \widehat{M}$ be arbitrary. By properties of Hausdorff distance, for any $\varepsilon > 0$ there exist $m_\infty \in M_\infty$ and $m_{\text{med}} \in M_{\text{med}}$ satisfying

$$d_\mathcal{M}(\hat{m}, m_\infty) \le d_H(\widehat{M}, M_\infty) + \varepsilon, \quad d_\mathcal{M}(m_\infty, m_{\text{med}}) \le d_H(M_\infty, M_{\text{med}}) + \varepsilon.$$

Applying Assumption 2.1,

$$\begin{aligned}
\Delta_{C_{\text{new}}}(\hat{m}, m_\infty) &\le \beta_{C_{\text{new}}} \cdot d_\mathcal{M}(\hat{m}, m_\infty) \le \beta_{C_{\text{new}}}(d_H(\widehat{M}, M_\infty) + \varepsilon),\\
\Delta_{C_{\text{new}}}(m_\infty, m_{\text{med}}) &\le \beta_{C_{\text{new}}} \cdot d_\mathcal{M}(m_\infty, m_{\text{med}}) \le \beta_{C_{\text{new}}}(d_H(M_\infty, M_{\text{med}}) + \varepsilon).
\end{aligned}$$

By the triangle inequality for $\Delta_{C_{\text{new}}}$,

$$\begin{aligned}
\inf_{m^* \in M^*_{\text{new}}} \Delta_{C_{\text{new}}}(\hat{m}, m^*) \le \Delta_{C_{\text{new}}}(\hat{m}, m_\infty) + \Delta_{C_{\text{new}}}(m_\infty, m_{\text{med}})\\
+ \inf_{m^* \in M^*_{\text{new}}} \Delta_{C_{\text{new}}}(m_{\text{med}}, m^*).
\end{aligned}$$

Since $m_{\text{med}} \in M_{\text{med}}$, we have

$$\inf_{m^* \in M^*_{\text{new}}} \Delta_{C_{\text{new}}}(m_{\text{med}}, m^*) \le \sup_{m \in M_{\text{med}}} \inf_{m^* \in M^*_{\text{new}}} \Delta_{C_{\text{new}}}(m, m^*).$$

Taking $\sup_{\hat{m} \in \widehat{M}}$, then $\varepsilon \to 0$, yields the claim. □

5.4 Interpretation of Error Components

We refer to the three terms in (1)–(3) as statistical, limited information, and approximation errors respectively.

Statistical Error $\beta_{C_{\text{new}}} \cdot d_H(\widehat{M}, M_\infty)$ quantifies sampling variability. This term converges to zero as sample size increases, at a rate determined by the complexity of $\mathcal{T}_\mathcal{S}$ and properties of the data-generating process. The factor $\beta_{C_{\text{new}}}$ translates base-metric convergence into regret convergence under the new context.

Limited Information Error $\beta_{C_{\text{new}}} \cdot d_H(M_\infty, M_{\text{med}})$ captures context misalignment. Even with infinite data from the old world, the resulting population set may differ from the set optimal under the new context. This term reflects the irreducible out-of-distribution gap and cannot be eliminated by increasing sample size alone.

Approximation Error The final term measures the penalty from feasibility constraints imposed by strategy $\mathcal{S}$. If $M^*_{\text{new}} \subseteq \mathcal{F}_\mathcal{S}(C_{\text{new}})$, this term vanishes. Otherwise, it quantifies the price of model misspecification or regularization.

5.5 The Complexity Trade-off

The decomposition reveals opposing effects of model complexity on the three error terms.

Proposition 5.1 (Complexity-Dependent Error Rates)**.** Consider a parametric family of strategies $\{\mathcal{S}_\lambda\}_{\lambda>0}$ satisfying

$$\lambda < \lambda' \implies \mathcal{F}_{\mathcal{S}_\lambda}(C) \subseteq \mathcal{F}_{\mathcal{S}_{\lambda'}}(C)$$

for all $C \in \mathcal{C}$. Under standard regularity conditions:

(i) Statistical error can be analyzed by statistical theory;

(ii) Limited information error $d_H(M_{\infty,\lambda}, M_{\text{med},\lambda})$ is weakly increasing in λ;

(iii) Approximation error is weakly decreasing in λ, with $\lim_{\lambda\to\infty}$ equal to zero if $\bigcup_\lambda \mathcal{F}_{\mathcal{S}_\lambda}(C) = \mathcal{M}$.

The optimal complexity λ^* balances statistical stability against approximation capability, with the balance point depending on sample size, signal strength, and context similarity.

5.6 Context Alignment and Information Transfer

The limited information error depends critically on the geometric relationship between old and new contexts.

Definition 5.4 (Context Proximity)**.** The *normalized context distance* between C_{old} and C_{new} relative to the Golden Set M_{real} is

$$\rho(C_{\text{old}}, C_{\text{new}}) := \frac{d_H(S_0(C_{\text{old}}), S_0(C_{\text{new}}))}{\max\{d_H(S_0(C_{\text{old}}), M_{\text{real}}), d_H(S_0(C_{\text{new}}), M_{\text{real}})\}}.$$

When $\rho \approx 0$, the optimal disks nearly coincide, indicating high information transfer. When $\rho \approx 1$, the contexts select maximally separated regions of model space.

Proposition 5.2 (Information Transfer Bound)**.** Suppose strategy $\mathcal{S}$ is Golden-compatible. Then

$$d_H(M_\infty, M_{\text{med}}) \le \rho(C_{\text{old}}, C_{\text{new}}) \cdot \max\{d_H(S_0(C_{\text{old}}), M_{\text{real}}), d_H(S_0(C_{\text{new}}), M_{\text{real}})\}.$$

Proof. Since $\mathcal{S}$ is Golden-compatible, $M_{\text{real}} \subseteq \mathcal{F}_{\mathcal{S}}(C)$ for all C. Therefore $M_\infty \subseteq S_0(C_{\text{old}})$ and $M_{\text{med}} \subseteq S_0(C_{\text{new}})$. The claim follows from the triangle inequality for Hausdorff distance and the definition of ρ. □

Remark 5.1. Context alignment is partially exogenous—determined by how nature distributes optimal models across worlds. However, strategic choices can reduce ρ: incorporating invariant constraints into $\mathcal{F}_{\mathcal{S}}$ shrinks feasible sets toward M_{real}, while domain adaptation techniques effectively modify the training context to better approximate the deployment context.

5.7 Regret Amplification via Conditioning

Theorem 5.1 can be reformulated using the contextual condition numbers introduced in Proposition 2.1.

Corollary 5.1 (Regret Bound via Conditioning)**.** Under Assumption 2.1,

$$\begin{aligned}\mathcal{R}(\widehat{M}, M^*_{\text{new}} \mid C_{\text{new}}) \le{}& \kappa(C_{\text{old}} \to C_{\text{new}}) \cdot \mathcal{R}(\widehat{M}, M_\infty \mid C_{\text{old}}) \\ &+ \beta_{C_{\text{new}}} \cdot d_H(M_\infty, M_{\text{med}}) \\ &+ \sup_{m \in M_{\text{med}}} \inf_{m^* \in M^*_{\text{new}}} \Delta_{C_{\text{new}}}(m, m^*),\end{aligned}$$

where $\kappa(C_{\text{old}} \to C_{\text{new}}) = \beta_{C_{\text{new}}}/\alpha_{C_{\text{old}}}$. Furthermore, if strategy $\mathcal{S}$ is Golden-compatible, then by Proposition 5.2,

$$\begin{aligned}\mathcal{R}(\widehat{M}, M^*_{\text{new}} \mid C_{\text{new}}) \le{}& \kappa(C_{\text{old}} \to C_{\text{new}}) \cdot \mathcal{R}(\widehat{M}, M_\infty \mid C_{\text{old}}) \\ &+ \beta_{C_{\text{new}}} \cdot \rho(C_{\text{old}}, C_{\text{new}}) \cdot \max\{d_H(S_0(C_{\text{old}}), M_{\text{real}}), d_H(S_0(C_{\text{new}}), M_{\text{real}})\} \\ &+ \sup_{m \in M_{\text{med}}} \inf_{m^* \in M^*_{\text{new}}} \Delta_{C_{\text{new}}}(m, m^*).\end{aligned}$$

Context pairs with small condition numbers enable robust generalization: in-sample estimation errors are transported to out-of-sample contexts with limited amplification. Conversely, large κ indicates that small training errors may lead to large deployment regret.

5.8 Optimal Strategy Selection

Given sample size n, context pair $(C_{\text{old}}, C_{\text{new}})$, and strategy class $\mathfrak{S}$, the optimal strategy solves

$$\mathcal{S}^* \in \arg\min_{\mathcal{S} \in \mathfrak{S}} \mathbb{E}_{d_{\omega_{\text{old}}} \sim P^n_{\omega_{\text{old}}}} \big[\mathcal{R}(\widehat{M}_{\mathcal{S}}, M^*_{\text{new}} \mid C_{\text{new}})\big].$$

By Theorem 5.1, this reduces to minimizing a weighted sum of estimation complexity, context sensitivity, and approximation error. The weights determined by $\beta_{C_{\text{new}}}$ and $\kappa(C_{\text{old}} \to C_{\text{new}})$, which reflect the geometric properties of the context pair, making optimal strategy selection intrinsically problem-dependent. When $\rho(C_{\text{old}}, C_{\text{new}})$ is small and $\kappa(C_{\text{old}} \to C_{\text{new}})$ is moderate, complex strategies (large λ) are favored: approximation gains dominate statistical costs. When contexts are poorly aligned or ill-conditioned, simple strategies provide robustness at the cost of approximation capability. This formalization recovers classical bias-variance-approximation intuitions within a unified geometric framework.

6 Conservatism versus Activism: Strategic Paradigms for Out-of-Distribution Deployment

The regret decomposition in the previous section assumes that the strategy operator $\mathcal{T}_{\mathcal{S}}$ is given exogenously. In practice, however, the modeler must *choose* among multiple candidate strategies, each encoding different trade-offs between constraint strength, estimation complexity, and cross-context robustness. This section introduces a fundamental classification of modeling strategies based on their relationship to the Golden Set, revealing a deep tension between two epistemic stances: *conservatism*, which prioritizes guaranteed inclusion of context-invariant structure at the cost of larger uncertainty sets, and *activism*, which pursues precision through stronger assumptions while risking exclusion of invariant models.

We establish rigorous performance bounds for both paradigms and derive a quantitative criterion for strategy selection. The key insight is that the *in-sample suboptimality* of a strategy provides an observable diagnostic signal for Golden Set exclusion, enabling data-driven navigation of the conservatism-activism trade-off.

6.1 Golden-Compatibility as a Design Principle

Recall from Definition 3.1 that the Golden Set $M_{\text{real}} = \bigcap_{C \in \mathcal{C}} S_0(C)$ is the subset of models achieving optimality in every admissible context. A strategy's relationship to M_{real} determines its cross-context robustness guarantees.

Definition 6.1 (Golden-Compatible Strategy)**.** A strategy $\mathcal{S}$ is *Golden-compatible* if its feasibility operator satisfies

$$M_{\text{real}} \subseteq \mathcal{F}_{\mathcal{S}}(C) \quad \text{for all } C \in \mathcal{C}.$$

A strategy that is not Golden-compatible is called *Golden-exclusive*.

Remark 6.1. Golden-compatibility ensures that feasibility constraints cannot inadvertently exclude universally optimal models. This property is not automatically satisfied: stronger economic or parametric assumptions may impose restrictions that, while plausible, conflict with true context-invariant structure.

Golden-compatibility has immediate geometric consequences for the population limit sets.

Proposition 6.1 (Golden-Compatible Population Sets)**.** If strategy $\mathcal{S}$ is Golden-compatible, then for any context C,

$$M_{\text{real}} \subseteq \mathcal{T}_{\mathcal{S}}(C) = \mathcal{O}_{\mathcal{S}}(C).$$

In particular, $M_{\text{real}} \subseteq M_\infty := \mathcal{T}_{\mathcal{S}}(\omega_{\text{old}}, \Phi_{\text{old}})$.

Proof. Since $M_{\text{real}} \subseteq S_0(C)$ for all C and $M_{\text{real}} \subseteq \mathcal{F}_{\mathcal{S}}(C)$ by Golden-compatibility, every element of M_{real} is both feasible under $\mathcal{S}$ and globally optimal in context C. Therefore M_{real} must be contained in the constrained-optimal set $\mathcal{T}_{\mathcal{S}}(C)$. □

6.2 The Conservative Paradigm

Definition 6.2 (Conservative Strategy)**.** A strategy $\mathcal{S}_{\text{cons}}$ is *conservative* if:

(i) It is Golden-compatible (Definition 6.1);

(ii) Its feasibility constraints impose only restrictions that are *certain* to hold for all models in M_{real}.

We denote the population limit set under conservative strategy applied to the old context by

$$M_{\infty,\text{cons}} := \mathcal{T}_{\mathcal{S}_{\text{cons}}}(\omega_{\text{old}}, \Phi_{\text{old}}).$$

Remark 6.2 (Epistemic Interpretation)**.** Conservatism embodies a *principle of epistemic caution*: impose only those restrictions for which violation would contradict fundamental structural properties known to hold universally. The resulting feasible set $\mathcal{F}_{\mathcal{S}_{\text{cons}}}(C)$ is typically large, reflecting agnosticism about context-specific features. This stance echoes Manski's advocacy for "incredible certitude" in identification—committing only to what can be established with minimal assumptions.

Definition 6.3 (Conservative Set Diameter)**.** The *conservative set diameter* is

$$R_{\text{cons}} := \operatorname{diam}(M_{\infty,\text{cons}}) = \sup_{m,m' \in M_{\infty,\text{cons}}} d_{\mathcal{M}}(m, m').$$

The following theorem establishes the performance guarantee for conservative strategies.

Theorem 6.1 (Conservative Regret Bound)**.** Suppose strategy $\mathcal{S}_{\text{cons}}$ is conservative. Then for any deployment context $C_{\text{new}} = (\omega_{\text{new}}, \Phi_{\text{new}})$,

$$\mathcal{R}(M_{\infty,\text{cons}}, M^*_{\text{new}} \mid C_{\text{new}}) \leq \beta_{C_{\text{new}}} \cdot R_{\text{cons}},$$

where $M^*_{\text{new}} = S_0(C_{\text{new}})$ is the unconstrained optimal set in the new context.

Proof. By Proposition 6.1, $M_{\text{real}} \subseteq M_{\infty,\text{cons}}$. Let $m \in M_{\infty,\text{cons}}$ be arbitrary. Since M_{real} is nonempty (under standard compactness conditions), there exists $g \in M_{\text{real}}$ such that

$$d_{\mathcal{M}}(m, g) \leq \operatorname{diam}(M_{\infty,\text{cons}}) = R_{\text{cons}}.$$

By definition of the Golden Set, $g \in S_0(C_{\text{new}}) = M^*_{\text{new}}$. Therefore,

$$\inf_{m^* \in M^*_{\text{new}}} d_{\mathcal{M}}(m, m^*) \leq d_{\mathcal{M}}(m, g) \leq R_{\text{cons}}.$$

Applying Assumption 2.1 (upper bound),

$$\inf_{m^* \in M^*_{\text{new}}} \Delta_{C_{\text{new}}}(m, m^*) \leq \beta_{C_{\text{new}}} \cdot d_{\mathcal{M}}(m, M^*_{\text{new}}) \leq \beta_{C_{\text{new}}} \cdot R_{\text{cons}}.$$

Taking supremum over $m \in M_{\infty,\text{cons}}$ yields the claim. □

Corollary 6.1 (Uniform Cross-Context Guarantee)**.** Under the conservative paradigm, the regret bound holds *uniformly* for all deployment contexts:

$$\sup_{C \in \mathcal{C}} \mathcal{R}(M_{\infty,\text{cons}}, S_0(C) \mid C) \leq \sup_{C \in \mathcal{C}} \beta_C \cdot R_{\text{cons}}.$$

Remark 6.3. The conservative bound depends solely on the diameter of the population set and does not require knowledge of the specific deployment context. This robustness is purchased at the cost of potentially large R_{cons}: weaker assumptions yield larger feasible sets and hence greater residual uncertainty.

6.3 The Activist Paradigm

Definition 6.4 (Activist Strategy)**.** A strategy $\mathcal{S}_{\text{act}}$ is *activist* if it imposes stronger feasibility constraints than any conservative strategy, potentially excluding the Golden Set. Formally, there exists a conservative strategy $\mathcal{S}_{\text{cons}}$ such that

$$\mathcal{F}_{\mathcal{S}_{\text{act}}}(C) \subsetneq \mathcal{F}_{\mathcal{S}_{\text{cons}}}(C)$$

for all $C \in \mathcal{C}$, with the possibility that $M_{\text{real}} \not\subseteq \mathcal{F}_{\mathcal{S}_{\text{act}}}(C)$ for some (or all) C.

We denote

$$M_{\infty,\text{act}} := \mathcal{T}_{\mathcal{S}_{\text{act}}}(\omega_{\text{old}}, \Phi_{\text{old}}).$$

Remark 6.4 (Epistemic Interpretation)**.** Activism reflects a willingness to incorporate stronger theoretical or empirical assumptions—assumptions that may be well-motivated but not guaranteed to hold across all contexts. This stance is characteristic of structural econometrics, where economic theory guides model specification even when such restrictions cannot be definitively verified. Activism pursues *precision through commitment*, accepting the risk that imposed structure may conflict with invariant features.

The critical consequence of activism is that the population set $M_{\infty,\text{act}}$ may fail to include the Golden Set, introducing a new source of error.

Definition 6.5 (In-Sample Suboptimality)**.** The *in-sample suboptimality* of strategy $\mathcal{S}$ in context C_{old} is

$$\epsilon_{\text{old}} := \inf_{m \in \mathcal{T}_{\mathcal{S}}(C_{\text{old}})} \Phi_{C_{\text{old}}}(m) - \inf_{m' \in S_0(C_{\text{old}})} \Phi_{C_{\text{old}}}(m').$$

Remark 6.5. The quantity ϵ_{old} measures how much performance the strategy sacrifices in the training context due to feasibility constraints. If $\epsilon_{\text{old}} = 0$, the constrained optimum achieves the unconstrained global optimum, indicating that $S_0(C_{\text{old}}) \cap \mathcal{F}_{\mathcal{S}}(C_{\text{old}}) \neq \emptyset$. If $\epsilon_{\text{old}} > 0$, the feasibility constraints exclude the global optimum.

The following lemma connects in-sample suboptimality to geometric distance from the optimal set.

Lemma 6.1 (Suboptimality-Distance Relationship)**.** Under Assumption 2.1, if $M_{\infty,\text{act}} \subseteq S_{\epsilon_{\text{old}}}(C_{\text{old}})$ (the ϵ_{old}-suboptimal set), then

$$d_H(M_{\infty,\text{act}}, S_0(C_{\text{old}})) \leq \frac{\epsilon_{\text{old}}}{\alpha_{C_{\text{old}}}},$$

where $\alpha_{C_{\text{old}}}$ is the lower distortion constant from Assumption 2.1.

Proof. By definition of ϵ_{old}, for any $m \in M_{\infty,\text{act}}$ there exists $m^* \in S_0(C_{\text{old}})$ such that

$$\Delta_{C_{\text{old}}}(m, m^*) = |\Phi_{C_{\text{old}}}(m) - \Phi_{C_{\text{old}}}(m^*)| \leq \epsilon_{\text{old}}.$$

Applying the lower bound in Assumption 2.1,

$$\alpha_{C_{\text{old}}} \cdot d_{\mathcal{M}}(m, m^*) \leq \Delta_{C_{\text{old}}}(m, m^*) \leq \epsilon_{\text{old}}.$$

Therefore $d_{\mathcal{M}}(m, S_0(C_{\text{old}})) \leq \epsilon_{\text{old}}/\alpha_{C_{\text{old}}}$. Taking supremum over $m \in M_{\infty,\text{act}}$ yields the Hausdorff distance bound. □

We now establish the activist regret bound, which necessarily involves both the activist set's intrinsic diameter and its deviation from the Golden Set.

Definition 6.6 (Activist Set Diameter)**.** The *activist set diameter* is

$$R_{\text{act}} := \text{diam}(M_{\infty,\text{act}}).$$

By construction of stronger constraints, $R_{\text{act}} \leq R_{\text{cons}}$.

Theorem 6.2 (Activist Regret Bound)**.** *Suppose strategy $\mathcal{S}_{\text{act}}$ is activist with in-sample suboptimality ϵ_{old} in context $C_{\text{old}} = (\omega_{\text{old}}, \Phi_{\text{old}})$. Then for deployment context $C_{\text{new}} = (\omega_{\text{new}}, \Phi_{\text{new}})$,*

$$\mathcal{R}(M_{\infty,\text{act}}, M^*_{\text{new}} \mid C_{\text{new}}) \leq \kappa(C_{\text{old}} \to C_{\text{new}}) \cdot \epsilon_{\text{old}} + \beta_{C_{\text{new}}} \cdot R_{\text{act}},$$

where $\kappa(C_{\text{old}} \to C_{\text{new}}) := \beta_{C_{\text{new}}}/\alpha_{C_{\text{old}}}$ is the contextual condition number from Proposition 2.1.

Proof. Let $m \in M_{\infty,\text{act}}$ be arbitrary. We construct a path from m to M^*_{new} via the Golden Set.

Step 1: Distance to Golden Set in old context. By definition of the Golden Set, $M_{\text{real}} \subseteq S_0(C_{\text{old}})$. Since $M_{\infty,\text{act}} \subseteq S_{\epsilon_{\text{old}}}(C_{\text{old}})$ by Definition 6.5, there exists $g \in M_{\text{real}}$ such that

$$\Delta_{C_{\text{old}}}(m, g) \leq \sup_{m' \in M_{\infty,\text{act}}} \inf_{g' \in S_0(C_{\text{old}})} \Delta_{C_{\text{old}}}(m', g') \leq \epsilon_{\text{old}}.$$

By Assumption 2.1 (lower bound),

$$d_{\mathcal{M}}(m, g) \leq \frac{\Delta_{C_{\text{old}}}(m, g)}{\alpha_{C_{\text{old}}}} \leq \frac{\epsilon_{\text{old}}}{\alpha_{C_{\text{old}}}}.$$

Step 2: Path through activist set diameter. For any $m \in M_{\infty,\text{act}}$, by triangle inequality and definition of diameter,

$$d_{\mathcal{M}}(m, g) \leq \sup_{m' \in M_{\infty,\text{act}}} d_{\mathcal{M}}(m', g) \leq \frac{\epsilon_{\text{old}}}{\alpha_{C_{\text{old}}}} + R_{\text{act}}.$$

The second inequality accounts for the worst-case path: from m to the element of $M_{\infty,\text{act}}$ closest to g, then to g.

Step 3: Golden Set membership in new context. By definition, $g \in M_{\text{real}} \subseteq S_0(C_{\text{new}}) = M^*_{\text{new}}$. Therefore,

$$\inf_{m^* \in M^*_{\text{new}}} d_{\mathcal{M}}(m, m^*) \leq d_{\mathcal{M}}(m, g) \leq \frac{\epsilon_{\text{old}}}{\alpha_{C_{\text{old}}}} + R_{\text{act}}.$$

Step 4: Translation to new context regret. Applying Assumption 2.1 (upper bound),

$$\inf_{m^* \in M^*_{\text{new}}} \Delta_{C_{\text{new}}}(m, m^*) \leq \beta_{C_{\text{new}}} \cdot d_{\mathcal{M}}(m, M^*_{\text{new}}) \leq \beta_{C_{\text{new}}} \left(\frac{\epsilon_{\text{old}}}{\alpha_{C_{\text{old}}}} + R_{\text{act}} \right).$$

Recognizing $\kappa(C_{\text{old}} \to C_{\text{new}}) = \beta_{C_{\text{new}}}/\alpha_{C_{\text{old}}}$ and taking supremum over $m \in M_{\infty,\text{act}}$ yields

$$\mathcal{R}(M_{\infty,\text{act}}, M^*_{\text{new}} \mid C_{\text{new}}) \leq \kappa \cdot \epsilon_{\text{old}} + \beta_{C_{\text{new}}} \cdot R_{\text{act}}.$$

□

Remark 6.6 (Interpreting the Two Terms)**.** The activist bound decomposes into:

- **Amplified suboptimality**: $\kappa \cdot \epsilon_{\text{old}}$ measures the consequence of Golden Set exclusion, with amplification governed by the condition number. If $\epsilon_{\text{old}} = 0$, the strategy achieves global optimality in the old context, suggesting (but not proving) Golden-compatibility.
- **Intrinsic uncertainty**: $\beta_{C_{\text{new}}} \cdot R_{\text{act}}$ reflects residual ambiguity within the (smaller) activist set, analogous to the conservative term but with $R_{\text{act}} \ll R_{\text{cons}}$.

6.4 The Conservatism-Activism Trade-off

Theorems 6.1 and 6.2 yield an explicit criterion for paradigm selection.

Theorem 6.3 (Optimal Paradigm Selection)**.** Suppose both conservative strategy $\mathcal{S}_{\text{cons}}$ and activist strategy $\mathcal{S}_{\text{act}}$ are available. The activist strategy achieves lower worst-case regret if and only if

$$\kappa(C_{\text{old}} \to C_{\text{new}}) \cdot \epsilon_{\text{old}} + \beta_{C_{\text{new}}} \cdot R_{\text{act}} < \beta_{C_{\text{new}}} \cdot R_{\text{cons}},$$

which simplifies to

$$\epsilon_{\text{old}} < \frac{\beta_{C_{\text{new}}}}{\kappa(C_{\text{old}} \to C_{\text{new}})} \cdot (R_{\text{cons}} - R_{\text{act}}).$$

Proof. Direct comparison of the bounds in Theorems 6.1 and 6.2. □

Remark 6.7 (Trade-off Structure)**.** The inequality reveals four key factors governing paradigm choice:

(i) **Constraint gain**: $R_{\text{cons}} - R_{\text{act}}$ quantifies the reduction in set diameter achieved by stronger assumptions. Larger gains favor activism.

(ii) **In-sample performance**: ϵ_{old} is the directly observable diagnostic. Small values suggest Golden-compatibility, licensing activist deployment.

(iii) **Context conditioning**: κ measures how adversely estimation errors in the old context propagate to the new context. Well-conditioned problems (small κ) tolerate activism; ill-conditioned settings demand conservatism.

(iv) **New context geometry**: $\beta_{C_{\text{new}}}$ weights all terms. Contexts with large β magnify all uncertainties proportionally.

The following corollary formalizes the diagnostic interpretation of in-sample suboptimality.

Corollary 6.2 (Zero-Suboptimality Test)**.** If $\epsilon_{\text{old}} = 0$, then $M_{\text{real}} \cap M_{\infty,\text{act}} \neq \emptyset$, and the activist bound reduces to

$$\mathcal{R}(M_{\infty,\text{act}}, M^*_{\text{new}} \mid C_{\text{new}}) \leq \beta_{C_{\text{new}}} \cdot R_{\text{act}}.$$

In this case, activism strictly dominates conservatism if $R_{\text{act}} < R_{\text{cons}}$.

Proof. If $\epsilon_{\text{old}} = 0$, then $M_{\infty,\text{act}} \cap S_0(C_{\text{old}}) \neq \emptyset$. Since $M_{\text{real}} \subseteq S_0(C_{\text{old}})$, feasibility constraints have not excluded all globally optimal models. The bound follows immediately from Theorem 6.2 with $\epsilon_{\text{old}} = 0$. □

6.5 Implications for Strategy Design

The conservative-activist dichotomy yields several actionable principles for modeling practice.

(i) **Conservative as default**: When contexts are poorly conditioned (κ large) or deployment environment is highly uncertain, default to conservative strategies.

(ii) **Activist with validation**: If in-sample suboptimality ϵ_{old} is negligible and κ is moderate, activist strategies are justified.

(iii) **Adaptive strategy**: Estimate ϵ_{old} from data. If it exceeds the threshold $\beta_{C_{\text{new}}}(R_{\text{cons}} - R_{\text{act}})/\kappa$, revert to conservatism.

If auxiliary data from multiple contexts $\{C_1, \dots, C_k\}$ are available, compute

$$\bar{\epsilon} := \frac{1}{k}\sum_{i=1}^{k}\epsilon_i,$$

where ϵ_i is in-sample suboptimality in context C_i. Consistently low $\bar{\epsilon}$ provides stronger evidence for Golden-compatibility than single-context validation. And the collection of strategies $\{\mathcal{S}_\lambda\}_{\lambda\in\Lambda}$ indexed by constraint strength forms a *robustness-precision frontier*. Conservatism and activism represent two regions on this frontier:

- Conservative region: Large R, small ϵ, guaranteed Golden-inclusion.
- Activist region: Small R, potentially positive ϵ, precision prioritized.

Intermediate strategies interpolate between these extremes, with optimal choice determined by Theorem 6.3.

6.6 Relation to Existing Methodological Debates

The conservatism-activism framework provides formal resolution to several longstanding methodological controversies:

Structural versus Reduced-Form Estimation. Structural econometrics typically adopts activist stances—imposing equilibrium conditions, functional form restrictions, and behavioral assumptions to achieve point identification. Reduced-form methods prioritize conservatism, estimating only relationships guaranteed under minimal assumptions. Our framework shows this is fundamentally a trade-off governed by ϵ_{old}, κ, and constraint gains, not an either-or dichotomy.

Model Selection in Machine Learning. The bias-variance trade-off is often framed as complexity versus generalization. We reinterpret this geometrically: simpler models (conservatism) have larger R but guaranteed coverage of invariant structure; complex models (activism) have smaller R but risk excluding transferable features. The condition $\epsilon_{\text{old}} < \beta(R_{\text{cons}} - R_{\text{act}})/\kappa$ operationalizes the classical intuition that well-fitting models generalize when the problem is "smooth."

Partial Identification versus Point Identification. Manski's advocacy for partial identification reflects a conservative stance: widen identified sets to guarantee inclusion of truth. Our bounds formalize this intuition and reveal the precise cost: regret linear in R_{cons}. Activists pursue point identification ($R_{\text{act}} \to 0$) by imposing structure, accepting positive ϵ_{old} as the price of exclusion risk.

7 Discussion and Conclusion

This proposal has established a unified geometric meta-theory for economic modeling, reinterpreting diverse methodological lineages—from semi-parametric inference to structural deep learning—as distinct contraction operators on a common metric space. By defining the **Golden Set** as the locus of context-invariant truth, we have provided a rigorous language to quantify the trade-offs between model complexity, structural assumptions, and out-of-distribution generalization.

Our analysis highlights the fundamental tension between *Conservatism* and *Activism*. While activist strategies offer precision through strong assumptions, they carry the risk of excluding the Golden Set when contexts shift. Conversely, conservative strategies prioritize robustness by identifying sets of models compatible with minimal invariant structure.

To bridge this theoretical framework with empirical practice, we propose the **Economics-Informed Neural Network (EINN)** framework. EINNs operationalize the conservative paradigm by optimizing a robust minimax objective that respects minimal economic constraints (such as equilibrium conditions or shape restrictions) while accommodating high-dimensional heterogeneity.

Detailed architectural specifications, training algorithms, and an analysis of the asymptotic properties of EINNs are provided in **Appendix A**. This computational framework demonstrates that the abstract geometric principles derived herein can guide the design of scalable, robust estimators for modern economic inquiry.

References

Ali Aouad and Antoine Désir. Representing random utility choice models with neural networks. *arXiv preprint arXiv:2207.12877*, 2022.

Susan Athey. The impact of machine learning on economics. In *The Economics of Artificial Intelligence: An Agenda*, pages 507–547. University of Chicago Press, 2018.

Susan Athey, Julie Tibshirani, and Stefan Wager. Generalized random forests. *The Annals of Statistics*, 47(2):1148–1178, 2019.

Susan Athey, Mohsen Bayati, Nikolay Doudchenko, Guido Imbens, and Khashayar Khosravi. Matrix completion methods for causal panel data models. *Journal of the American Statistical Association*, 116(536):1716–1730, 2021.

Alexandre Belloni, Victor Chernozhukov, and Christian Hansen. Inference on treatment effects after selection among high-dimensional controls. *The Review of Economic Studies*, 81(2):608–650, 2014.

Victor Chernozhukov, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21(1):C1–C68, 2018.

Max H. Farrell, Tengyuan Liang, and Sanjog Misra. Deep neural networks for estimation and inference. *Econometrica*, 89(1):181–213, 2021.

Max H. Farrell, Tengyuan Liang, and Sanjog Misra. Deep learning for individual heterogeneity. *arXiv preprint arXiv:2010.14694v3*, 2025.

Jesús Fernández-Villaverde. Deep learning for solving economic models. Working Paper 34250, National Bureau of Economic Research, 2025.

Matthew Gentzkow, Bryan Kelly, and Matt Taddy. Text as data. *Journal of Economic Literature*, 57(3):535–574, 2019.

Domenico Giannone, Michele Lenza, and Giorgio E Primiceri. Economic predictions with big data: The illusion of sparsity. *Econometrica*, 89(5):2409–2437, 2021.

Shihao Gu, Bryan Kelly, and Dacheng Xiu. Empirical asset pricing via machine learning. *The Review of Financial Studies*, 33(5):2223–2273, 2020.

Tetsuya Kaji, Elena Manresa, and Guillaume Pouliot. An adversarial approach to structural estimation. *Econometrica*, 91(6):2041–2063, 2023.

Jon Kleinberg, Jens Ludwig, Sendhil Mullainathan, and Ziad Obermeyer. Prediction policy problems. *American Economic Review*, 105(5):491–495, 2015.

Greg Lewis and Vasilis Syrgkanis. Adversarial generalized method of moments. *arXiv preprint arXiv:1803.07164*, 2018.

Lilia Maliar, Serguei Maliar, and Pablo Winant. Deep learning for solving dynamic economic models. *Journal of Monetary Economics*, 122:76–101, 2021.

Sendhil Mullainathan and Jann Spiess. Machine learning: an applied econometric approach. *Journal of Economic Perspectives*, 31(2):87–106, 2017.

Amandeep Singh, Ye Liu, and Hema Yoganarasimhan. Choice models and permutation invariance: Demand estimation in differentiated products markets. *Working Paper*, 2025.

Stefan Wager and Susan Athey. Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523):1228–1242, 2018.

Shenhao Wang, Baichuan Mo, and Jinhua Zhao. Deep neural networks for choice analysis: Architecture design with alternative-specific utility functions. *Transportation Research Part C: Emerging Technologies*, 112:234–251, 2020.

Yanhao 'Max' Wei and Zhenling Jiang. Estimating parameters of structural models using neural networks. *Marketing Science*, 2024.

Yanhao 'Max' Wei and Zhenling Jiang. Pre-training estimators for structural models: Application to consumer search. *Working Paper*, 2025.

Yanhao 'Max' Wei and Nikhil Malik. Unstructured data, econometric models, and estimation bias. *Working Paper*, 2025.

Yuntao Wu, Goutham Gopalakrishna, Jiayuan Guo, and Zissis Poulos. Deep-macrofin: Informed equilibrium neural network for continuous-time economic models. *arXiv preprint arXiv:2408.10368v4*, 2025.

A Appendix A: Economics-Informed Neural Networks as Conservative Estimators

This appendix develops a computational framework for implementing the conservative modeling paradigm introduced in Section 6. Economics-Informed Neural Networks (EINNs) provide a flexible, scalable approach to estimating the conservative identified set $M_{\infty,\text{cons}}$ while respecting both minimal economic structure and robustness to contextual uncertainty.

A.1 Motivation: Operationalizing Conservative Identification

Recall from Section 6 that the conservative identified set $M_{\infty,\text{cons}}$ satisfies Golden-compatibility ($M_{\text{real}} \subseteq \mathcal{F}_{\mathcal{S}_{\text{cons}}}(C)$ for all C) and imposes only minimal structural constraints. Theorem 6.1 guarantees regret $\mathcal{R} \leq \beta_{C_{\text{new}}} \cdot R_{\text{cons}}$, where $R_{\text{cons}} = \text{diam}(M_{\infty,\text{cons}})$.

Computational Challenge: Traditional grid-based methods for partial identification become infeasible when $\mathcal{M}$ is high-dimensional. EINNs address this via continuous optimization in parameter space.

A.2 Neural Parameterization of the Model Space

Let $\Theta \subseteq \mathbb{R}^p$ denote a parameter space. We consider models of the form

$$m_\theta = \mathfrak{M}(\theta; \{\text{NN components}\}),$$

where here $\mathfrak{M}$ is a known structural form (e.g., $m_\theta(x,t) = \sum_{j=0}^{K-1} \theta_j(x)\phi_j(t)$ for VCM) and some components (e.g., the coefficient functions $\theta_j(x)$) are parameterized as neural networks. This can be also used flexibly in any structual model that does not want to specify the parametrical forms of the structual functions such as the utility function.

Assumption A.1 (Hybrid Neural Parameterization)**.** (i) Θ is compact or regularized to ensure effective compactness;

(ii) The map $\theta \mapsto m_\theta$ is continuous under $d_{\mathcal{M}}$;

(iii) Neural components have sufficient capacity to approximate the relevant function spaces.

Remark A.1. This hybrid approach preserves economic structure (via $\mathfrak{M}$) while leveraging neural networks for flexible nonparametric components. For example, in demand estimation, $\mathfrak{M}$ might encode utility maximization constraints while allowing flexible taste heterogeneity via neural networks.

A.3 Economics Information Constraints

Here is the minimum assumption we assume or be confident that the reality must holds.

Definition A.1 (Economics Information Constraint Functional)**.** An *EIC* is a differentiable functional $\mathcal{E} : \Theta \to [0, \infty)$ with $\mathcal{E}(\theta) = 0$ iff m_θ satisfies the minimal structural requirements.

Assumption A.2 (Golden-Compatibility of EIC)**.** The constraints satisfy $M_{\text{real}} \subseteq \{m_\theta : \mathcal{E}(\theta) = 0\}$.

Examples:

1. **Moment conditions**: $\mathcal{E}(\theta) = \sum_{j=1}^{J} \left(\mathbb{E}[g_j(m_\theta, D)]\right)_+^2$
2. **Shape constraints**: $\mathcal{E}(\theta) = \int_{\mathcal{X}} \left(-\frac{\partial m_\theta(x)}{\partial x}\right)_+^2 d\mu(x)$ (monotonicity)
3. **Equilibrium conditions**: $\mathcal{E}(\theta) = \|h(m_\theta, \omega_{\text{old}})\|_2^2$ for $h(\cdot) = 0$

A.4 Conservative EINN: Constrained Optimization

Conservative identification is achieved by minimizing in-sample error subject to Golden-compatibility constraints.

The conservative EINN solves:

$$\boxed{\theta^* \in \arg\min_{\theta \in \Theta} \Phi_{C_{\text{old}}}(m_\theta) + \lambda \mathcal{E}(\theta)} \tag{4}$$

where:

- $\Phi_{C_{\text{old}}}(m_\theta)$ is the empirical loss on training context C_{old}
- $\mathcal{E}(\theta)$ enforces Golden-compatibility via structural constraints
- $\Phi_{C_{\text{old}}}(m_\theta) + \mathcal{E}(\theta)$ is the total loss
- $\lambda > 0$ is the penalty weight

A.5 Partial Identification and Set-Valued Estimation

Conservative estimation naturally yields partial identification when information is insufficient for point identification.

Definition A.2 (Conservative Identified Set)**.** The population identified set is

$$\Theta_\infty := \left\{\theta : \mathcal{E}(\theta) = 0,\ \Phi_{C_{\text{old}}}(\theta) = \inf_{\theta'} \Phi_{C_{\text{old}}}(\theta')\right\},$$

with model-space counterpart $M_\infty = \{m_\theta : \theta \in \Theta_\infty\}$.

The identification strength is $R_{\text{cons}} := \text{diam}(M_\infty)$, where $R_{\text{cons}} = 0$ indicates point identification.

Remark A.2. Multiple parameter values may achieve the same minimal loss while satisfying constraints, leading to $R_{\text{cons}} > 0$. This occurs when: (i) the model is overparameterized relative to available information, or (ii) constraints $\mathcal{E}(\theta) = 0$ define a non-degenerate manifold in Θ.

A.6 Computing Prediction Intervals

For target functional $\psi : \mathcal{M} \times \Omega \to \mathbb{R}$ and deployment world ω_{new}, compute bounds via constrained optimization:

$$\underline{\psi}(\omega_{\text{new}}) := \min_{\theta} \psi(m_\theta, \omega_{\text{new}}) \quad \text{s.t. } \mathcal{E}(\theta) \le \delta,\ \Phi_{C_{\text{old}}}(\theta) \le \Phi^* + \epsilon, \tag{5}$$

$$\overline{\psi}(\omega_{\text{new}}) := \max_{\theta} \psi(m_\theta, \omega_{\text{new}}) \quad \text{s.t. } \mathcal{E}(\theta) \le \delta,\ \Phi_{C_{\text{old}}}(\theta) \le \Phi^* + \epsilon, \tag{6}$$

where $\Phi^* := \inf_\theta \Phi_{C_{\text{old}}}(\theta)$ and $\delta, \epsilon > 0$ are small tolerances.

Conservative prediction interval: $\mathcal{I}(\omega_{\text{new}}) = [\underline{\psi}(\omega_{\text{new}}), \overline{\psi}(\omega_{\text{new}})]$.

Remark A.3. Problems (5)–(6) are solved via standard constrained optimization with multiple random restarts. The constraint $\Phi_{C_{\text{old}}}(\theta) \leq \Phi^* + \epsilon$ defines an approximate level set of well-fitting models, while $\mathcal{E}(\theta) \leq \delta$ ensures near-satisfaction of structural requirements.

A.6.1 Alternative: Diverse Ensemble

Train K models minimizing

$$\sum_{k=1}^{K}[\Phi_{C_{\text{old}}}(\theta_k) + \lambda\mathcal{E}(\theta_k)] - \gamma \cdot \frac{1}{K(K-1)} \sum_{k \neq k'} \mathbb{E}_\omega \|\psi(m_{\theta_k}, \omega) - \psi(m_{\theta_{k'}}, \omega)\|^2,$$

where the diversity term encourages distinct predictions across the identified set.

Ensemble-based prediction interval: $\mathcal{I}_{\text{ensemble}}(\omega_{\text{new}}) = [\min_k \psi(m_{\theta_k}, \omega_{\text{new}}), \max_k \psi(m_{\theta_k}, \omega_{\text{new}})]$.

Remark A.4. The diversity penalty prevents mode collapse and encourages exploration of the identified set Θ_∞. Empirically, this often provides better coverage than single-model optimization, though theoretical guarantees require that diversity does not compromise constraint satisfaction.

A.7 Theoretical Properties

Conjecture A.1 (Hausdorff Consistency)**.** Let $\widehat{M}_n(\epsilon_n)$ denote the finite-sample identified set with $\epsilon_n \to 0$ as $n \to \infty$. Under regularity conditions, $d_H(\widehat{M}_n(\epsilon_n), M_\infty) \xrightarrow{p} 0$.

Conjecture A.2 (Convergence Rates)**. (i) Parametric**: If $\Theta \subseteq \mathbb{R}^p$ with fixed $p < \infty$, then $d_H(\widehat{M}_n, M_\infty) = O_p(n^{-1/2})$. **(ii) Nonparametric**: If neural components have $p_n \to \infty$ parameters, then $d_H(\widehat{M}_n, M_\infty) = O_p(n^{-s/(2s+d)})$ for smoothness s and effective dimension d.

Conjecture A.3 (Asymptotic Coverage)**.** If $m_{\text{true}} \in M_\infty$ and C_{new} satisfies the Golden-compatibility assumption, then $\lim_{n\to\infty} \mathbb{P}(\psi(m_{\text{true}}, \omega_{\text{new}}) \in \mathcal{I}_n(\omega_{\text{new}})) = 1$.

A.8 Connection to Main Framework

Proposition A.1 (EINN Conservative Guarantee)**.** If Assumption A.2 holds and $\theta^* \in \Theta_\infty$, then for any deployment context C_{new},

$$\mathcal{R}(\{m_{\theta^*}\}, S_0(C_{\text{new}}) \mid C_{\text{new}}) \leq \beta_{C_{\text{new}}} \cdot R_{\text{cons}}.$$

Proof. By Assumption A.2, $M_{\text{real}} \subseteq \{m_\theta : \mathcal{E}(\theta) = 0\}$. Since $\theta^* \in \Theta_\infty$ satisfies $\mathcal{E}(\theta^*) = 0$, the feasible set contains M_{real}. Apply Theorem 6.1 with $M_{\infty,\text{cons}} = M_\infty$. □

Contrasting with Activist EINN: An activist approach uses stronger constraints $\mathcal{E}_{\text{strong}}$ (e.g., parametric functional forms) that may exclude M_{real}:

$$\theta^*_{\text{act}} \in \arg\min_\theta \Phi_{C_{\text{old}}}(m_\theta) + \lambda\mathcal{E}_{\text{strong}}(\theta).$$

This yields tighter predictions ($R_{\text{act}} < R_{\text{cons}}$) but potential in-sample suboptimality $\epsilon_{\text{old}} := \Phi_{C_{\text{old}}}(m_{\theta^*_{\text{act}}}) - \inf_{m \in M_{\text{real}}} \Phi_{C_{\text{old}}}(m) > 0$.

By Theorem 6.3, use activist if $\epsilon_{\text{old}} < \frac{\beta_{C_{\text{new}}}}{\kappa(C_{\text{old}} \to C_{\text{new}})}(R_{\text{cons}} - R_{\text{act}})$.

A.9 Discussion

Advantages: (i) Scalability to high-dimensional model spaces via neural parameterization, (ii) flexible constraint encoding through differentiable $\mathcal{E}$, (iii) unified framework for partial/point identification, (iv) automatic handling of set-valued estimation without grid search.

Limitations: (i) Non-convexity requires multiple restarts, (ii) hyperparameter selection (λ, γ, δ, ϵ) lacks formal guidance, (iii) finite-sample coverage guarantees remain open, (iv) constraint specification $\mathcal{E}$ requires domain expertise to ensure Golden-compatibility.

Related work: EINN extends moment inequality methods and Manski bounds to neural network settings. Unlike distributionally robust optimization, robustness derives from theory-driven constraints rather than optimization over ambiguity sets. It is very similar as physics-informed neural networks, but constraints encode economic theory (equilibrium, rationality) rather than PDEs.

A.10 Summary

EINNs operationalize conservative identification through: (i) hybrid parameterization combining structural models with neural flexibility, (ii) EIC functionals ensuring Golden-compatibility via theory-driven constraints, (iii) Balancing in-sample fit against structural requirements, (iv) set-valued estimation naturally handling partial identification, and (v) prediction intervals with expected asymptotic coverage. The framework bridges geometric regret theory and computational practice, providing a principled approach to robust model estimation under limited information.