

# Admissible Portfolio Optimization: Information Constraints, Conditional Efficient Frontiers, and the Price of Causal Identification

Alejandro Rodríguez Domínguez\*

September 2026

## Abstract

Mean–variance portfolio choice takes the conditioning information as given and optimizes over weights, so that two errors about the information, using variables not available at the decision time and treating common variation as idiosyncratic, pass into the portfolio unseen. This paper makes the conditioning information a decision variable subject to hard admissibility constraints declared in advance: availability, a no-arbitrage-preserving enlargement of the decision filtration, statistical separation and, for interventional claims, invariance across declared regimes. A lexicographic admissibility order in which no risk–return quantity enters selects an optimal information class, and the classical portfolio problem is then solved inside it. We prove existence, invariance under recodings, a value-of-admissible-information theorem whose failure conditions show that selection by decision loss admits look-ahead information whenever present, and that the two-stage solution and the joint envelope are the minimal elements of two orders on one set. Under exact separation the diversifiable part of risk is a property of the admissible class and is interventionally stable only under interventional admissibility; requiring causal identification has an explicit oracle price traded against search complexity. The estimator is consistent with second-order regret. On market data with 127 candidate drivers the condition is attainable on individual equities, where enlarging the search reduces the defect at a measurable rate, and not on pre-diversified portfolios, where it is nearly invariant to the search because the residual dependence is the common factor itself; where it fails the resulting covariance still yields minimum-variance portfolios that match shrinkage at markedly lower turnover, while discarding a measurable part of their risk, which an exact decomposition attributes to the residual share, the breadth of the portfolio and the average residual correlation.

**Keywords:** portfolio choice; information selection; conditional efficient frontier; causal identification; decision-focused learning; oracle inequalities.

**JEL:** C44, C58, G11. **MSC 2020:** 91G10, 62C05, 90C29.

## 1 Introduction

A mean–variance investor chooses portfolio weights once expected returns and covariances have been specified relative to some information [45, 47]. The information is an input. The theory of

---

\*Quantitative Analysis and Artificial Intelligence Department, Miralta Finance Bank S.A., Madrid, Spain; and Department of Computer Science, University of Reading, Reading, United Kingdom. Email: [arodriguez@miraltabank.com](mailto:arodriguez@miraltabank.com). The views expressed are those of the author and do not necessarily reflect those of the institutions with which he is affiliated.

conditioning information [29, 30] shows how much it matters for the frontier the investor perceives, but in practice it is a factor model, a set of signals or the output of a variable-selection procedure, adopted before the optimizer is called and never itself optimized. Two errors about this input, using variables that were not legitimately available when the decision was taken and treating common variation across assets as if it were idiosyncratic, cause a backtest to overstate what a portfolio can deliver [20, 31, 48]. Neither is visible once the moments have been estimated, and neither is corrected by covariance shrinkage [39] or by training the forecasting model on the decision loss [25], since both take the information as given.

The paper makes the conditioning information a decision variable, subject to hard constraints,<sup>1</sup> and the primitive of portfolio construction becomes an admissible information–portfolio pair, chosen in two stages that do not interact. A decision contract, written before any return is examined, declares the claim the portfolio is meant to support and the conditions a candidate information set must satisfy: availability at the decision time, an enlargement of the decision filtration that preserves the absence of arbitrage, removal of the common dependence between assets that mean–variance choice treats as absent, and, when the claim concerns interventions, invariance of the conditional return law across the declared regimes. Candidates that violate a required condition are infeasible. Among the admissible ones a lexicographic admissibility order, in which no risk–return quantity enters, selects an optimal information class; only then, inside that class, is the classical problem solved and its frontier, tangency portfolio and capital allocation line constructed. The order is not a utility, because chronology, publicity, separation and identification are propositions with truth values rather than attributes with rates of substitution.

Markowitz introduced diversification through the portfolio problem; the paper introduces admissibility, and causality as its interventional case, through the same problem, and the two notions are connected. Diversification concerns dependence between returns, dependence is by the common-cause principle [53] attributable to shared drivers, and the admissible information class determines which drivers may explain it. Conditional on admissible information, covariation splits into a response to common drivers and an idiosyncratic residual, only the residual is diversifiable, and the split is a property of the information class rather than of the asset universe.

The contributions are the following. The two-stage problem has solutions and the selected class is invariant to recodings of the same information. The value of information is monotone only under an embedding hypothesis that portfolio constraints can break, with the sharp consequence that any rule selecting information by decision loss, decision-focused selection included, admits look-ahead information whenever the candidate set contains it. The two-stage solution and the joint envelope over all admissible representations are the minimal elements of two orders on one set of admissible pairs, differing only in whether risk may compare portfolios across representations. Under exact separation the diversifiable part of risk is determined by the admissible class and is invariant to interventions only under interventional admissibility, the portfolio form of the principle of common drivers. Requiring causal identification carries an oracle price bounded by an explicit distance in conditional moments and traded against representation-search complexity in an oracle inequality. The two-stage estimator is consistent, the regret of the estimated portfolio is second order in estimation error, and information selection separates across decision nodes. Controlled experiments verify the identities and rates. Field experiments on six asset universes with 127 economic and financial candidate drivers find that the separator condition is attainable on individual equities, where selection halves the maximal residual cross-asset correlation and the admissible class is non-empty in a sixth of the windows, and unattainable on industry, size and value or momentum

---

<sup>1</sup>Admissibility is used throughout in the sense of eligibility for a declared claim, not in Wald’s decision-theoretic sense of non-dominance; an admissible representation here may be dominated in decision value by an inadmissible one.

portfolios, whose residual dependence is the common factor that diversification has already isolated and is nearly invariant to the size of the candidate search; where the condition fails, the conditional covariance still yields minimum-variance portfolios whose volatility is indistinguishable from linear shrinkage and lower than the sample and principal-component estimators, at a third less turnover and lower leverage than any of them, while the diagonal-residual restriction it imposes discards about a seventh of the predicted variance of the portfolios it produces, moving a risk model that is otherwise calibrated outside its confidence band.

Population dominance of the admissible problem over any fixed representation is a population result, and Section 4 gives its finite-sample counterpart; causal restriction can lower static decision value because it solves a stronger problem, and a lower conditional loss never licenses a causal interpretation; and of the admissibility conditions only the mandate scope is a declaration, the others being certified from provenance, from the enlarged filtration or from returns.

## 2 Literature review

The paper sits at the intersection of four literatures, and in each its place is the same: they take as given the object this paper makes a decision variable.

Mean–variance theory [45, 47] chooses weights given a mean vector and a covariance matrix; its conditional extensions compute those moments relative to information available at the decision time. That the conditioning information reshapes the frontier is the classical result of Hansen and Richard [29]; Ferson and Siegel [27, 30] characterized the portfolios that use a given conditioning set efficiently, and Kandel and Stambaugh [38] measured the value of predictive information from the allocation side. In all of this the conditioning set is supplied by the model; the frontier they characterize is, for a fixed admissible class, the second-stage object of this paper, whose question is the prior one of which conditioning set may be used at all, and Section 3.2 places the classical objects next to their counterparts here. The estimation side of the literature documents that optimized portfolios are fragile to errors in means [8, 20, 48] and that shrinkage and constraints stabilize them [36, 39], while approximate factor structures [18, 58] justify treating residual risk as diversifiable. The present paper is concerned with a question prior to estimation, namely on which conditioning information the split between systematic and diversifiable risk presupposed by these estimators holds, and it shows that when the split fails the resulting error has a sign.

Two orderings of information enter the analysis. Blackwell’s comparison of experiments [11, 12, 59] says that finer information weakly improves every decision problem because coarser policies remain available; Reichenbach’s common-cause principle [53] attributes dependence to a shared cause and removes it by conditioning. The paper uses the second as the separator condition, with residual dependence after conditioning as the statistical defect of a candidate, and the first as the benchmark against which admissibility truncates the value of information: Blackwell’s theorem presupposes that finer information is eligible and coarser policies implementable, and both fail here, the first for look-ahead and non-public information and the second when constraints depend on the representation.

Decision-focused learning and integrated estimation–optimization address the mismatch between the loss used to estimate a model and the loss of the decision it feeds: the smart predict-then-optimize formulation [25], its survey [44], the comparison with two-step estimation [26], and portfolio implementations [3, 15, 22, 41, 43, 62]. These methods decide how the primitives of a given representation are estimated; they do not decide whether the representation is eligible for the claim attached to it, and the paper proves that they cannot, since selection by decision loss admits look-ahead information whenever the candidate set contains it. From model selection [7, 13, 46, 60]

the paper takes the approximation–estimation logic of oracle inequalities and applies it to a nesting fixed by the claim; the backtesting problem of [31] is the case the chronological condition excludes.

Causal structure entered portfolio construction through the principle that the common causal drivers of a portfolio’s constituents are the drivers relevant to its diversification [55]; causal-network representations have since been applied to factor design and timing [34]. Pearl’s hierarchy [6, 50] grades claims, structural causal models are formalized in [52], and invariant causal prediction [14, 32, 51] identifies causal predictors by the stability of the conditional law across environments. The paper makes the grading and the invariance criterion part of the decision contract and adds the portfolio consequence, that the diversifiable component of risk is determined by the admissible class and is interventionally stable only under interventional admissibility, at a price it bounds. Coherent and spectral risk measures [2, 4], the expected-shortfall program of [57] and mean–risk theory [9, 42, 49, 54] supply alternatives to variance for the second stage, orthogonal to the information contract of the first; Appendix B shows that changing the risk contract inside a fixed class changes the portfolio without conferring dominance.

The rest of the paper is organized as follows. Section 3 first states the results the paper uses, the conditional frontier in its two parametrizations, Blackwell’s comparison of experiments and the consequences of causal separation, and then develops the theory. It defines contracts, claim levels and admissibility conditions, the admissible class and its order, proves the value-of-admissible-information theorem with its look-ahead corollary, states the portfolio problem, relates the two-stage solution to the envelope, develops the conditional frontier, the admissible tangent portfolio and capital allocation line, and derives the diversification result, the causal price and the perturbation bound. Section 4 contains the finite-sample theory, the implementation, the controlled experiments and the public-data experiment. Section 5 concludes. Proofs are in the appendix; configuration, robustness runs and a side experiment on risk contracts are in Appendix B.

### 3 Admissible information and admissible portfolio choice

This section states, with citations, the mathematics on which the paper rests: the conditional mean–variance frontier, Blackwell’s comparison of experiments, and the consequences of causal separation, which are proved here so that the paper is self-contained; every later result invokes one of the statements below.

#### 3.1 Setting and notation

Fix a finite asset family  $A$  with  $|A| = n$ , a decision node  $t$ , a horizon  $h > 0$ , a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  and baseline operative information  $\mathcal{G}_t^0 \subseteq \mathcal{F}$ . Let  $R_{t,h} \in L^2(\mathbb{P}; \mathbb{R}^n)$  be the vector of horizon returns, written  $R$  when the node and horizon are fixed. The reference probability measure is  $\mathbb{P}$ , and all equalities between  $\sigma$ -algebras, conditional laws and conditional moments are modulo  $\mathbb{P}$ -null sets. The space  $\mathbb{R}^n$  carries the Euclidean topology;  $\succ 0$  and  $\succeq 0$  denote positive definiteness and semidefiniteness;  $\mathbf{1}$  is the vector of ones. Upper hemicontinuity of a correspondence is in the sense of [1, Def. 17.2]. Lower semicontinuity of extended-real functions is in the sense of  $\liminf_{x' \rightarrow x} f(x') \geq f(x)$ . Calligraphic letters  $\mathcal{A}, \mathcal{H}, \mathcal{W}, \mathcal{D}$  denote classes of representations, candidate sets, feasible portfolio sets and decision contracts;  $H, \tilde{H}, K$  denote representations,  $w$  a portfolio and  $m$  a target return. For a representation  $H$ ,  $\mu_H := \mathbb{E}[R \mid H]$  and  $\Sigma_H := \text{Cov}(R \mid H)$  are its conditional mean and covariance. A star marks optimal objects:  $\mathcal{A}_t^*$  is the Stage-I solution set,  $H_t^*$  a representative of it,  $\rho_H^*(m)$  the minimum conditional risk at target  $m$  under  $H$ , and  $\rho_t^*(m) := \rho_{H_t^*}^*(m)$  when the representative is fixed. Hats denote estimators.

### 3.2 The conditional mean–variance frontier

Given  $\mu_H \in \mathbb{R}^n$  and  $\Sigma_H \succ 0$ , the mean–variance problem with full investment is  $\min\{w^\top \Sigma_H w : \mathbf{1}^\top w = 1, \mu_H^\top w = m\}$  [45]. Its solution is classical [47].

**Proposition 3.1** (Conditional frontier; [47]). *Let  $a = \mathbf{1}^\top \Sigma_H^{-1} \mathbf{1}$ ,  $b = \mathbf{1}^\top \Sigma_H^{-1} \mu_H$ ,  $c = \mu_H^\top \Sigma_H^{-1} \mu_H$  and  $d = ac - b^2 > 0$ . For every target  $m$  the minimum-variance portfolio with  $\mathbf{1}^\top w = 1$  and  $\mu_H^\top w = m$  is*

$$w_H(m) = \frac{c - bm}{d} \Sigma_H^{-1} \mathbf{1} + \frac{am - b}{d} \Sigma_H^{-1} \mu_H, \quad \rho_H^*(m) = w_H(m)^\top \Sigma_H w_H(m) = \frac{am^2 - 2bm + c}{d}. \quad (1)$$

The minimum-variance portfolio is  $w_H(b/a) = \Sigma_H^{-1} \mathbf{1}/a$  with variance  $1/a$ ; the frontier  $m \mapsto \rho_H^*(m)$  is a convex parabola, and its efficient branch is  $m \geq b/a$ . The utility form  $\min\{\frac{\gamma}{2} w^\top \Sigma_H w - \mu_H^\top w : \mathbf{1}^\top w = 1\}$  has the unique minimizer  $w_H(m(\gamma))$  with  $m(\gamma) = b/a + d/(a\gamma)$  (the budget multiplier is  $(b - \gamma)/a$  and the mean of the minimizer is  $(c - b(b - \gamma)/a)/\gamma$ ), so every  $\gamma > 0$  selects one point of the efficient branch and every efficient target  $m > b/a$  is selected by  $\gamma = d/(a(m - b/a))$ ; the two parametrizations describe the same frontier. If a riskless rate  $r_f < b/a$  is available, the tangency portfolio is  $w_H^{\text{tan}} = \Sigma_H^{-1}(\mu_H - r_f \mathbf{1})/(b - ar_f)$  and the capital allocation line is  $\mu = r_f + \sqrt{c - 2br_f + ar_f^2} \sigma$ .

The proposition is invoked in Section 3.8 with  $H$  the admissibility-optimal class; nothing in it depends on how  $H$  was chosen. The sensitivity of  $w_H(m)$  to errors in  $\mu_H$  is the classical finding of [20, 48] and reappears in the finite-sample experiments. Figure 1 places the classical objects of this proposition next to their counterparts in the admissible problem, and Table 1 lists the correspondence.

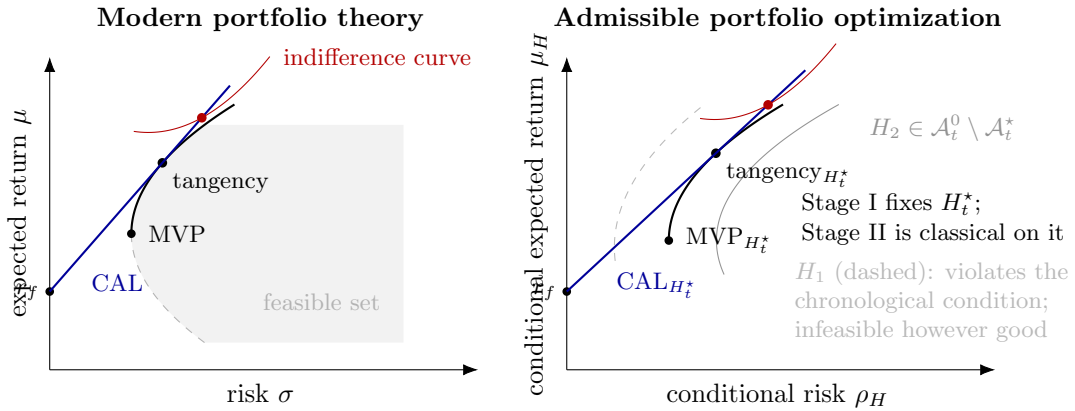


Figure 1: Modern portfolio theory and its admissible counterpart. Left: the classical picture, in which the feasible set (shaded), the minimum-variance portfolio (MVP), the efficient frontier, the capital allocation line (CAL) through the tangency portfolio and the investor’s indifference curve are defined relative to one unquestioned set of moments. Right: every one of those objects reappears, relative to the conditional moments of the admissibility-optimal information  $H_t^*$ . A candidate that violates an admissibility condition ( $H_1$ ) is infeasible whatever its frontier, and a candidate that satisfies the conditions but is not admissibility-optimal ( $H_2$ ) is not compared on risk and return. Preferences enter after the information has been fixed.

### 3.3 Comparison of information

For sub- $\sigma$ -algebras  $\sigma(H_1) \subseteq \sigma(H_2)$  every  $\sigma(H_1)$ -measurable policy is  $\sigma(H_2)$ -measurable. For statistical experiments the corresponding order is Blackwell's [11, 12, 59]: an experiment is more informative than another, in the sense of a no larger minimal Bayes risk in every decision problem with bounded loss and a fixed action space, if and only if the other is a garbling of it. Both provisos, bounded loss and a fixed action space, are replaced below by integrability and by an embedding of policies, and Section 3.8 shows that in portfolio problems the embedding can fail.

### 3.4 Causal separation, response and residual

The statistical condition on which the admissibility theory rests is that a set of drivers renders asset returns conditionally independent over the horizon. Its consequences for conditional covariances are proved here.

**Definition 3.2** (Separator). Let  $Z$  be a vector of  $k$  drivers with path  $Z_{[t,t+h]}$  over the horizon, let  $H^+ := \sigma(\mathcal{G}_t^0, Z_{[t,t+h]})$  be the horizon-closed information it generates and  $H^- := \sigma(\mathcal{G}_t^0, Z_t)$  its decision-time component. The drivers are a *separator* for  $A$  if the components of  $R$  are mutually independent conditional on  $H^+$ . A separator is *least* if no separator generates a strictly smaller  $\sigma$ -algebra.

**Proposition 3.3** (Tower decomposition). *For any drivers  $Z$ , with  $\xi := \mathbb{E}[R \mid H^+] - \mathbb{E}[R \mid H^-]$  the horizon innovation of the conditional mean,*

$$\text{Cov}(R \mid H^-) = \mathbb{E}[\text{Cov}(R \mid H^+) \mid H^-] + \text{Cov}(\xi \mid H^-). \quad (2)$$

*If  $Z$  is a separator, the first term, denoted  $\Omega$ , is diagonal. If moreover  $\mathbb{E}[R \mid H^+]$  is affine in a  $k$ -dimensional statistic  $\zeta$  of the driver path,  $\mathbb{E}[R \mid H^+] = \alpha + B\zeta$ , then  $\text{Cov}(\xi \mid H^-) = BVB^\top$  with  $V = \text{Cov}(\zeta \mid H^-)$ ;  $B$  is the response operator and  $V$  the innovation covariance. The first term is idiosyncratic and the second systematic.*

The proof is in Appendix A.

Every object in (2) is invariant under invertible bimeasurable recodings of the driver coordinates, since such a recoding leaves  $\sigma(Z)$  and hence every conditional expectation unchanged; in the affine case  $B$  transforms by the Jacobian and  $BVB^\top$  is unchanged.

Interventional invariance follows from a structural model. If the drivers and returns are embedded in a structural causal model in which  $Z$  is a cause of  $R$ , and an intervention acts on variables that are not descendants of  $Z$  and leaves the mechanism from  $Z$  to  $R$  unchanged, then the conditional law of  $R$  given  $H^+$  is the same under the observational and the post-intervention law; this is the modularity of structural causal models [50, Ch. 1 and 3], [52, Sec. 6.3]. Every object in (2) is then invariant to the intervention. The paper does not assume such a model; it makes invariance across a declared set of regimes the definition of the causal admissibility condition (Definition 3.6), and this remark records when that condition follows from a structural model.

The screening defect used in Section 3 to grade candidate representations is the departure from Definition 3.2: for a candidate  $H$  it is the maximal absolute conditional correlation between components of  $R$  after conditioning on  $H$ , equal to zero exactly for separators.

### 3.5 Common drivers and diversification

**Proposition 3.4** (Common drivers and diversification, [55]). *Let the constituents of a portfolio have causal drivers in the sense of a declared structural model, and call a driver common if it is*

a cause of at least two constituents. The drivers that are optimal for the portfolio’s diversification dynamics, in the sense that conditioning on them removes the dependence between constituents that spreading weights is meant to average away, are the common drivers; conditioning on drivers that are not common leaves that dependence in place. In the notation of Proposition 3.3, conditioning on the common drivers is what places their innovations in the systematic term  $BVB^\top$  and leaves the residual  $\Omega$  diagonal.

The principle is the causal origin of the diversification result of Section 3.12. What a portfolio can diversify away is what remains after conditioning on the common causes, and the residual  $\Omega$  of Proposition 3.3 is the formal object that remainder denotes.

The results above are the ones the paper uses. The remainder of the section introduces the objects that are new: the decision contract with its claim levels and admissibility conditions, the admissible class and the order that ranks it, the value of admissible information, and the portfolio problem inside the selected class.

### 3.6 Decision contracts, claim levels and admissibility conditions

**Definition 3.5** (Decision contract). A decision contract  $\mathcal{D}$  specifies, before any data are examined, the asset family  $A$ , the decision node  $t$  and horizon  $h$ , the baseline information  $\mathcal{G}_t^0$ , a declared class  $\mathcal{H}$  of candidate information representations, the feasible portfolio set  $\mathcal{W}(H) \subseteq \mathbb{R}^n$  attached to each candidate  $H$ , the claim level of Definition 3.6 with its identification requirements, and the decision loss  $\ell$ , with  $J(w; H) = \mathbb{E}[\ell(w, R_{t,h}; H)]$ . A representation has a decision-time component  $H^-$ , with respect to which portfolios are measurable, and may have a horizon-closed component  $H^+$  used only for certification.

**Definition 3.6** (Claim levels). A decision contract declares the level of the causal hierarchy at which its conclusions are to be read [6, 50]. A *conditional-law claim* ( $\mathcal{C} = c$ ) asserts properties of the conditional law  $\mathcal{L}(R_{t,h} \mid H)$  under the observational regime only; it requires the financial, chronological and statistical admissibility conditions and no causal semantics. An *interventional claim* ( $\mathcal{C} = i$ ) asserts in addition that the conditional primitives remain valid under a declared set  $\mathcal{E}$  of interventions, transports or counterfactual regimes, which must contain the regimes about which the claim is made. Identification at this level is defined by invariance: writing  $\mathcal{L}^e$  for the law under regime  $e \in \mathcal{E}$  and  $\mathcal{L}^o$  for the observational law, the causal admissibility condition is satisfied if and only if  $\mathcal{L}^e(R_{t,h} \mid H) = \mathcal{L}^o(R_{t,h} \mid H)$  for every  $e \in \mathcal{E}$ . Because the interventional level adds a requirement to the conditional-law level,  $\mathcal{A}_t(i) \subseteq \mathcal{A}_t(c)$  for a fixed candidate set.

The grading is the portfolio analogue of the association–intervention–counterfactual ladder [50], and it is part of the contract because it determines which conditions a candidate must satisfy. Identification is the invariance criterion of [51], testable from data collected under the regimes in  $\mathcal{E}$  and coinciding with graphical identification when  $\mathcal{E}$  consists of interventions on non-descendants of  $H$  in a structural model [50, Ch. 3]. If the claim concerns interventions on the drivers themselves,  $\mathcal{E}$  must contain them, and invariance under them is a statement about the absence of confounding between  $H$  and  $R$ . Invariance does not select a minimal set, since every superset of an invariant conditioning set is invariant; minimality enters through the dimension coordinate of the admissibility order. The requirement that  $\mathcal{E}$  contain the regimes of interest keeps the condition from being vacuous.

**Definition 3.7** (Admissibility conditions). A representation  $H \in \mathcal{H}$  passes the *financial admissibility condition* if (i) its decision-time component is measurable with respect to information released no later than the decision node, a property of data provenance certified from release timestamps

and point-in-time vintages; (ii) the enlargement of the baseline filtration by  $H^-$  preserves the semimartingale property of discounted prices and the absence of free lunches with vanishing risk [24], for which Jacod’s condition on the conditional laws of the added information is sufficient [5, 35]; and (iii) the enlargement lies within the mandate and legal scope declared in the contract. Condition (i) concerns timing and (ii) publicity: for information that is public before  $t$  the enlargement stays inside the market filtration and (ii) holds automatically, whereas (ii) is binding when a variable passes the timestamp check but is available to the decision maker alone, as for an insider; the *statistical admissibility condition* if it satisfies the declared separator or sufficiency condition; the *causal admissibility condition* if, for an interventional claim, it carries identified intervention semantics in the sense of Definition 3.6; and the *implementation admissibility condition* if the representation can be implemented under the declared trading technology. We write  $G_F(H)$ ,  $G_S(H)$ ,  $G_C(H)$  and  $G_I(H)$  for the corresponding indicators, equal to one when the condition is satisfied. The contract determines which admissibility conditions are required, and the *admissible class*  $\mathcal{A}_t$  is the set of candidates in  $\mathcal{H}$  satisfying every required admissibility condition.

Causal admissibility is a constraint rather than a score. For an interventional claim, representations without identified intervention semantics are infeasible; for a conditional-law claim they are not excluded for lacking them.

The conditions differ in how they are established. Chronological availability is a deterministic check on provenance. The enlargement condition is the classical question of the theory of enlargement of filtrations: an initial enlargement violating Jacod’s condition can destroy the semimartingale property and create arbitrage, the mechanism by which an insider’s expected utility becomes infinite [5], so the condition is a hypothesis on the enlarged filtration, verifiable within a model. Implementation feasibility is verifiable from exchange rules and liquidity data; the statistical and causal conditions are certified from returns, with sampling error quantified in Section 4. Only the mandate scope is a declaration, with the same status as the set of admissible strategies in [24], and its misspecification is model risk rather than estimation error.

Three candidates illustrate the parts. A published factor return passes (i) by its release calendar and (ii) automatically, being in the market filtration. A proprietary order-flow signal computed before  $t$  passes (i) and is the case for which (ii) is written. A realized horizon return fails (i) on the calendar and (ii) in the extreme form in which the informed agent’s utility is unbounded.

**Definition 3.8** (Decision equivalence). Two admissible representations  $H$  and  $\tilde{H}$  are *decision-equivalent*,  $H \sim_{\mathcal{D}} \tilde{H}$ , if the feasible sets  $\mathcal{W}(H)$  and  $\mathcal{W}(\tilde{H})$  coincide and every population decision primitive entering  $J$  coincides almost surely, so that  $J(\cdot; H)$  and  $J(\cdot; \tilde{H})$  coincide as functions on the common feasible set. In the exact conditional-law case, equality of the generated  $\sigma$ -algebras modulo null sets together with identical implementation constraints is sufficient.

**Definition 3.9** (Information order and policy embedding). For decision-time information objects write  $H_1 \preceq H_2$  when  $\sigma(H_1) \subseteq \sigma(H_2)$  modulo  $\mathbb{P}$ -null sets; when the objects are statistical experiments rather than sub- $\sigma$ -algebras,  $\preceq$  denotes the declared Blackwell (garbling) order [12]. For  $H \in \mathcal{A}_t$  let  $\Pi(H)$  denote the set of feasible policies, i.e.  $\sigma(H^-)$ -measurable random vectors  $\pi$  with  $\pi \in \mathcal{W}(H)$  almost surely, and write  $J(\pi; H) = \mathbb{E}[\ell(\pi, R_{t,h}; H)]$ . For  $H_1 \preceq H_2$ , say that  $H_2$  embeds the  $H_1$  decision problem if there is a map  $\iota : \Pi(H_1) \rightarrow \Pi(H_2)$  with  $\ell(\iota\pi, R_{t,h}; H_2) \leq \ell(\pi, R_{t,h}; H_1)$  almost surely for every  $\pi \in \Pi(H_1)$ . When the loss does not depend on the representation the inequality is an equality for  $\iota$  the identity; when it does, as for the conditional mean–variance loss, the identity map is still an embedding whenever  $\sigma(H_1^-) \subseteq \sigma(H_2^-)$  and  $\mathcal{W}(H_1) \subseteq \mathcal{W}(H_2)$ , by the law of total variance (Appendix A).

The order  $\preceq$  compares information content and  $\sim_{\mathcal{D}}$  identifies representations indistinguishable for the declared decision. Embedding is stronger than inclusion of weight vectors when constraints or costs depend on the representation, and its failure is one of the two ways in which the monotonicity of information value fails (Corollary 3.14).

### 3.7 The admissible class and the admissibility order

Fix a decision node  $t$ , the asset family, horizon, baseline information, candidate set  $\mathcal{H}_t$ , claim level, tolerances and portfolio constraints.

**Definition 3.10** (Admissible class and admissibility order). The admissible class  $\mathcal{A}_t$  is the set of candidates in  $\mathcal{H}_t$  that satisfy every admissibility condition the contract requires. To each  $H \in \mathcal{A}_t$  are attached four nonnegative quantities: the separator or sufficiency defect  $\varepsilon_t(H)$ ; the causal-response defect  $\delta_t(H)$ , identically zero under a conditional-law contract; the information complexity  $\kappa_t(H)$ ; and a predeclared implementation burden  $\iota_t(H)$ . The first two are scientific quantities, the last two tie-breaks.

- (a) (*Exact order.*)  $H_1 \prec_{\mathcal{A}} H_2$  if the first of  $\varepsilon_t, \delta_t, \kappa_t, \iota_t$  at which  $H_1$  and  $H_2$  differ is smaller for  $H_1$ .
- (b) (*Order at tolerance  $\bar{\varepsilon}$ .*) Let  $\mathcal{A}_t(\bar{\varepsilon})$  be the candidates in  $\mathcal{A}_t$  with  $\varepsilon_t(H) \leq \bar{\varepsilon}$ . On  $\mathcal{A}_t(\bar{\varepsilon})$ ,  $H_1 \prec_{\mathcal{A}}^{\bar{\varepsilon}} H_2$  if the first of  $\delta_t, \kappa_t, \iota_t, \varepsilon_t$  at which they differ is smaller for  $H_1$ : within tolerance the defect is a satisfied constraint, and it re-enters only as the last tie-break.

The Stage-I solution is the set of minimal elements,

$$\mathcal{A}_t^* = \text{Min}_{\prec_{\mathcal{A}}} \mathcal{A}_t, \quad \mathcal{A}_t^*(\bar{\varepsilon}) = \text{Min}_{\prec_{\mathcal{A}}^{\bar{\varepsilon}}} \mathcal{A}_t(\bar{\varepsilon}), \quad (3)$$

The two orders coincide at  $\bar{\varepsilon} = 0$  when an exact separator exists, since then (a) and (b) both select the exact separators of least dimension.

No return target, Sharpe ratio, realized utility or point of the Stage-II frontier enters any of the four quantities; Figure 2 illustrates.

The lexicographic order is chosen over a penalty because chronology, publicity, separation and identification are propositions with truth values, not attributes with rates of substitution: a penalty would let a large enough Sharpe ratio purchase a false proposition, and a lexicographic order admits no continuous utility representation [28]. On a finite candidate set it has a real representation (Section 4), and every result holds for each fixed pre-declared tie-break.

In the exact population problem a least separator, when it exists, has  $\varepsilon_t = 0$  and is selected by information minimality; in finite samples the statistical condition is applied with a pre-declared tolerance calibrated to sampling error (Section 4). Causal validity is never exchanged against portfolio performance.

**Proposition 3.11** (Canonicity of Stage I). *Suppose the required admissibility conditions and the four quantities  $\varepsilon_t, \delta_t, \kappa_t, \iota_t$  are constant on information-equivalence classes. Then (a)  $\mathcal{A}_t^*$  is a union of information-equivalence classes and is invariant to invertible measurable recodings of candidate variables; (b) if an exact least admissible separator  $H_t^*$  exists, satisfies the required admissibility conditions and has finite implementation burden, then  $[H_t^*] \subseteq \mathcal{A}_t^*$ .*

The proof is in Appendix A.

**Theorem 3.12** (Existence of the two solution sets). *If the candidate set is finite,  $\mathcal{A}_t^*$  is nonempty, and for every  $H \in \mathcal{A}_t^*$  with  $\mathcal{W}(H)$  nonempty and compact and  $\rho_H, \mu_H$  continuous, the conditional efficient set*

$$\mathcal{E}_H(m) = \arg \min \{ \rho_H(w) : w \in \mathcal{W}(H), \mu_H(w) \geq m \} \quad (4)$$

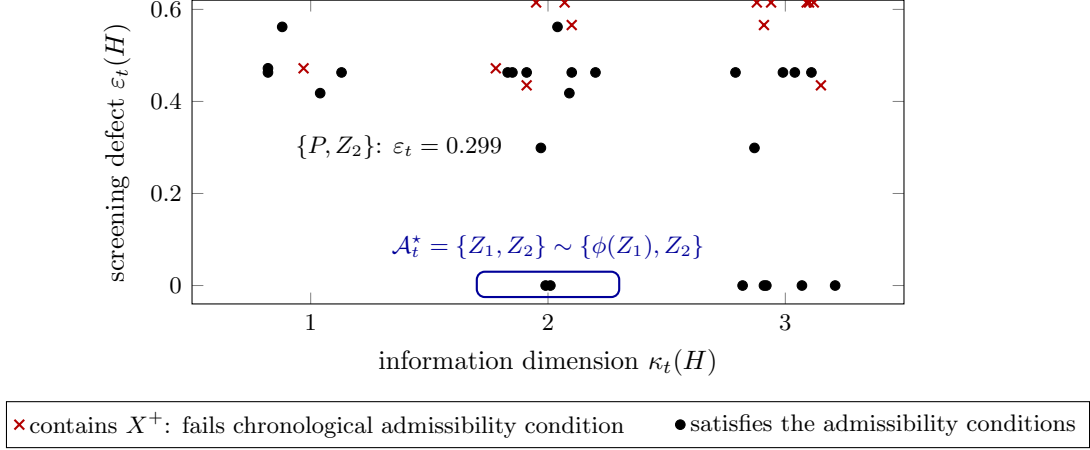


Figure 2: Stage I is an information problem, shown on the 41 candidates of experiment E1 (population values). Crosses are the 16 candidates containing the future variable  $X^+$ ; they include the best predictors and are infeasible. Among the 25 admissible candidates the lexicographic order first minimizes the screening defect, which selects the exact separators, and then the dimension, which selects the class of  $\{Z_1, Z_2\}$  and its recoding over the three-dimensional exact separators. No risk–return quantity enters.

is nonempty and compact for every feasible target  $m$ . The same holds when  $\mathcal{A}_t/\sim_{\mathcal{D}}$  is compact and the four admissibility quantities are lower semicontinuous on it.

The proof is in Appendix A; for a finite candidate set the first claim is that a finite set with a total preorder has minimal elements, and the second is Weierstrass’ theorem.

Because neither the four quantities nor  $\mathcal{A}_t$  depend on  $m$  or on any portfolio, the set  $\mathcal{A}_t^*$  is the same for every feasible target, so a change of target moves the portfolio inside the fixed class and never reselects information.

When  $k \geq 2$  no single direction is an exact separator, so  $\varepsilon_t(H_\theta) > 0$  on the whole sphere and Stage I selects the directions of least residual dependence; enlarging the class to  $\sigma(\Theta^\top Z)$  for  $\Theta$  on the Stiefel manifold of  $k$ -frames recovers  $\varepsilon_t = 0$  at  $\Theta$  spanning  $\mathbb{R}^k$ , and the same argument applies.

### 3.8 Value of admissible information

That more information cannot hurt is the Blackwell order (Section 3.3), a statement about decision problems with a fixed feasible policy space, not an unconditional property of portfolio representations. It requires the richer information to remain eligible and the poorer-information policies to remain implementable.

**Theorem 3.13** (Value of admissible information). *Let  $H_1, H_2 \in \mathcal{A}_t$  with  $H_1 \preceq H_2$  and suppose  $H_2$  embeds the  $H_1$  decision problem (Definition 3.9). Then*

$$V(H_2) \leq V(H_1). \quad (5)$$

*If moreover  $V(H_1)$  is attained at some  $\pi_1 \in \Pi(H_1)$  and there exists  $\pi_2 \in \Pi(H_2)$  with*

$$\ell(\pi_2, R_{t,h}; H_2) \leq \ell(\iota\pi_1, R_{t,h}; H_2) \text{ a.s.}, \quad \mathbb{P}\{\ell(\pi_2, R_{t,h}; H_2) < \ell(\iota\pi_1, R_{t,h}; H_2)\} > 0,$$

*then  $V(H_2) < V(H_1)$ .*

The proof is in Appendix A.

**Corollary 3.14** (The envelope need not choose maximal information). *Let  $H_1 \preceq H_2$ . If  $H_2 \notin \mathcal{A}_t$ , or if  $H_2 \in \mathcal{A}_t$  but no embedding  $\iota$  exists, Theorem 3.13 is inapplicable and  $V(H_2) \leq V(H_1)$  may fail. In particular maximal raw information need not solve the envelope problem even when it is more informative in the Blackwell order.*

The proof is in Appendix A.

Enlarging a financial filtration can violate the financial condition, a future variable improves prediction while failing chronology, and a richer observational representation need not carry interventional semantics: admissibility truncates the information order.

The failure of chronology has a sharper consequence, because decision loss is decreasing in information.

**Proposition 3.15** (Decision-loss selection under look-ahead contamination). *Let  $H$  be a representation with decision-time component  $H^-$ , and let  $X^+$  be a random variable measurable with respect to the horizon-closed information and not  $\sigma(H^-)$ -measurable. Let  $H^\sharp$  be the representation with decision-time component  $\sigma(H^-, X^+)$  and  $\mathcal{W}(H^\sharp) = \mathcal{W}(H)$ . Then  $V(H^\sharp) \leq V(H)$ ; and for the quadratic loss  $J(w; \cdot) = \frac{\gamma}{2} \text{Var}(w^\top R \mid \cdot) - \mathbb{E}[w^\top R \mid \cdot]$  with  $\mathcal{W}$  containing a neighbourhood of the optimizer, the inequality is strict whenever  $\mathbb{E}[R \mid H^\sharp] \neq \mathbb{E}[R \mid H^-]$  on a set of positive probability. Consequently, on any candidate set that contains  $H$  and  $H^\sharp$ , every rule that selects by minimizing  $V$ , the envelope problem included, selects a candidate containing  $X^+$ ; decision-loss selection is consistent only on candidate sets from which the chronological admissibility condition has removed every such candidate.*

The proof is in Appendix A.

The proposition is the population form of the fact, measured in Section 4.6, that the look-ahead representation attains the lowest in-sample loss in every replication: the chronological condition cannot be replaced by a loss criterion, because the criterion is how look-ahead enters.

The information problem and the portfolio problem must be kept separate. The first selects a class by admissibility, and only after that class is fixed does the second construct the efficient set conditional on it (Figure 3). The conditional portfolio problem is used in two parametrizations of one frontier, the utility form with risk aversion  $\gamma$  and the target form with return target  $m$ , related by Proposition 3.1.

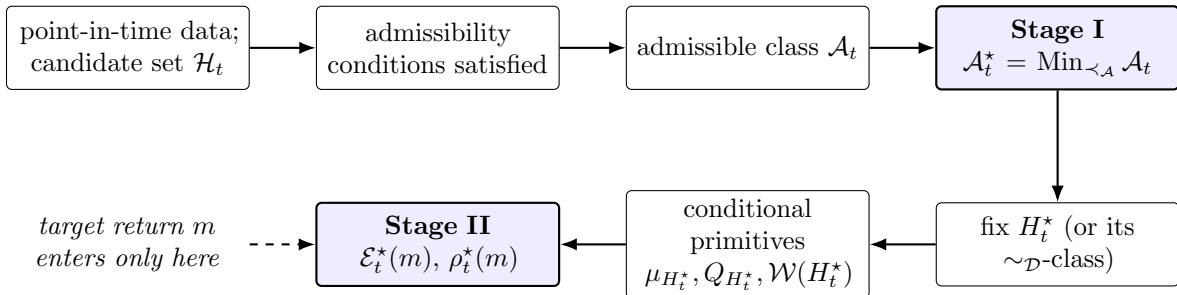


Figure 3: The two-solution architecture. Candidates violating a required admissibility condition are infeasible, not penalized. Stage I ranks the survivors by the admissibility order alone and fixes the minimal class. Stage II constructs the conditional efficient set and frontier inside that class; portfolio performance never relaxes a violated admissibility condition.

### 3.9 The portfolio problem: envelope and two-stage solution

In conditional form, Markowitz portfolio choice reads

$$w_H^* \in \arg \min_{w \in \mathcal{W}(H)} \left\{ \frac{\gamma}{2} w^\top \Sigma_H w - \mu_H^\top w + c_H(w) \right\}, \quad (6)$$

where  $H$  is the information relative to which  $\mu_H, \Sigma_H$  are defined and  $c_H$  a cost term; the representation  $H$  is an input. The admissible problem makes it a variable.

**Definition 3.16** (Admissible envelope problem). Let  $J : \{(H, w) : H \in \mathcal{A}_t, w \in \mathcal{W}(H)\} \rightarrow \mathbb{R} \cup \{+\infty\}$  be the population decision loss. The admissible envelope problem is

$$V^* = \inf_{H \in \mathcal{A}_t} \inf_{w \in \mathcal{W}(H)} J(w; H), \quad (7)$$

and any minimizer  $(H^*, w^*)$  is an envelope-optimal pair. The conditional value function is  $V(H) = \inf_{w \in \mathcal{W}(H)} J(w; H)$ , so that  $V^* = \inf_{H \in \mathcal{A}_t} V(H)$ . Problem (7) ranks admissible representations by decision loss. The name APO is reserved for the two-stage problem of Definition 3.19, in which admissible representations are ranked by admissibility and loss enters only inside the selected class; (7) is its envelope in the sense of Proposition 3.20, and it is the object to which the finite-sample bounds of Section 4 apply.

**Example 3.17** (A representation switch that a fixed-representation optimizer cannot express). Consider two assets under full investment and two admissible information states with conditional moments

$$\Sigma_{H_1} = \begin{pmatrix} 1 & 0.8 \\ 0.8 & 1 \end{pmatrix}, \mu_{H_1} = \begin{pmatrix} 0.12 \\ 0.08 \end{pmatrix}; \quad \Sigma_{H_2} = \begin{pmatrix} 1 & 0.1 \\ 0.1 & 1 \end{pmatrix}, \mu_{H_2} = \begin{pmatrix} 0.09 \\ 0.10 \end{pmatrix}.$$

For  $\gamma = 2$  and  $c_H \equiv 0$ , problem (6) returns  $w_{H_1}^* = (0.55, 0.45)$  with  $J = 0.799$ , and  $w_{H_2}^* = (0.497, 0.503)$  with  $J = 0.455$ . A Markowitz program produces either allocation but cannot compare them, since each conditional law is one of its inputs. The envelope problem (7) compares  $V(H_1) = 0.799$  with  $V(H_2) = 0.455$  and selects  $(H_2, w_{H_2}^*)$  although  $H_1$  carries the larger conditional mean. APO proper (Definition 3.19) compares  $H_1$  and  $H_2$  by admissibility and constructs only the winner's frontier; Proposition 3.20 gives the exact relation between the two.

Two facts about (7) are immediate. If  $H \sim_{\mathcal{D}} \tilde{H}$  then  $V(H) = V(\tilde{H})$ , because the value depends on  $H$  only through the primitives that define decision equivalence, so  $V$  is well defined on  $\mathcal{A}_t / \sim_{\mathcal{D}}$  and the problem may be posed on the quotient; and  $V^* \leq V(H_0)$  for every  $H_0 \in \mathcal{A}_t$ , since the fixed-representation problem is one point of the admissible one.

**Theorem 3.18** (Existence). *If the candidate set is finite and each  $\mathcal{W}(H)$  is compact with  $J(\cdot; H)$  lower semicontinuous, the envelope problem (7) has a minimizer. More generally a minimizer exists when  $\mathcal{A}_t / \sim_{\mathcal{D}}$  is compact,  $H \mapsto \mathcal{W}(H)$  is compact-valued and upper hemicontinuous and  $J$  is lower semicontinuous on its graph.*

The proof is in Appendix A.

The hypotheses of the general clause are verified on a continuous class of driver directions in Appendix A (Proposition A.1); elsewhere in the paper the candidate set is finite. For a finite candidate set nothing beyond compactness of each feasible set is needed; for continuous classes the topology on the quotient and the continuity of  $H \mapsto \mathcal{W}(H)$  are modeling requirements.

**Definition 3.19** (APO as a partial order on admissible pairs). Fix a feasible target  $m$  and write  $\Gamma_t(m) = \{(H, w) : H \in \mathcal{A}_t, w \in \mathcal{W}(H), \mu_H(w) \geq m\}$  for the admissible pairs at that target. Define the strict partial order  $\prec_{\text{APO}}$  on  $\Gamma_t(m)$  by

$$(H', w') \prec_{\text{APO}} (H, w) \iff H' \prec_{\mathcal{A}} H \text{ or } [H' = H \text{ and } \rho_H(w') < \rho_H(w)]. \quad (8)$$

Admissible portfolio optimization at target  $m$  is the problem of finding the  $\prec_{\text{APO}}$ -minimal elements of  $\Gamma_t(m)$ ; its solution set is denoted  $\text{APO}_t(m)$ .

The relation is the lexicographic order on the pair (admissibility of  $H$ , risk of  $w$ ) with the second coordinate compared only for the same representation, hence a strict partial order; it compares representations by admissibility and portfolios by risk only within a representation, and two portfolios attached to different representations are incomparable. The requirement that loss must not select information is expressed by this incomparability.

**Proposition 3.20** (Two-stage solution and envelope as two orders on the same set). *For every feasible target  $m$ ,*

$$\text{APO}_t(m) = \bigcup_{H \in \mathcal{A}_t^*} \{H\} \times \mathcal{E}_H(m), \quad (9)$$

*so the two-stage construction of this section computes the  $\prec_{\text{APO}}$ -minimal pairs. If  $\prec_{\text{APO}}$  is replaced by the total preorder  $(H', w') \preceq_{\text{env}} (H, w) \iff \rho_{H'}(w') \leq \rho_H(w)$  on  $\Gamma_t(m)$ , the minimal elements are the pairs attaining the joint envelope (11); with  $\rho$  replaced by  $J$  and the target constraint dropped, they are the minimizers of the envelope problem (7).*

The proof is in Appendix A.

The two concepts differ only in whether risk may compare portfolios across representations. No admissibility order reproduces the envelope, since portfolios attached to different representations are incomparable under  $\prec_{\text{APO}}$ ; the envelope is the total preorder, appropriate when the representation is a free modeling choice, whereas APO is appropriate whenever the representation carries a claim.

### 3.10 The conditional efficient frontier

Assume first that Stage I identifies a single information-equivalence class and choose any representative  $H_t^* \in \mathcal{A}_t^*$ . Conditional on it the portfolio problem is classical in form; in the quadratic case with full investment its solution is Proposition 3.1 applied to  $\mu_{H_t^*}, \Sigma_{H_t^*}$ .

Once the class is fixed, the second solution set is the classical one. For each feasible target  $m$ ,

$$\mathcal{E}_t^*(m) = \arg \min_{w \in \mathcal{W}(H_t^*) : \mu_{H_t^*}(w) \geq m} \rho_{H_t^*}(w), \quad \rho_t^*(m) = \min_{w \in \mathcal{W}(H_t^*) : \mu_{H_t^*}(w) \geq m} \rho_{H_t^*}(w). \quad (10)$$

The correspondence  $m \mapsto \mathcal{E}_t^*(m)$  is the second solution set of APO and the graph of  $m \mapsto \rho_t^*(m)$  is its efficient frontier. Transaction and holding costs belong to this stage; the cost term  $c_H$  of (6) may be carried in the risk functional or in the feasible set, in which case the frontier is net of costs; costs never enter the first stage.

**Proposition 3.21** (Convexity of the conditional frontier). *Let  $\rho_H(w) = w^\top \Sigma_H w$  with  $\Sigma_H \succ 0$ ,  $\mu_H(w) = \mu_H^\top w$ , and let  $\mathcal{W}(H_t^*)$  be defined by finitely many affine equalities and convex inequalities. Then for every feasible  $m$  the Stage-II problem (10) is a convex program with a unique solution, and  $m \mapsto \rho_t^*(m)$  is convex and nondecreasing on the set of feasible targets.*

The proof is in Appendix A.

The first solution is generally set-valued. For  $\mathcal{A}_t^*$  with several non-equivalent elements APO returns the correspondence  $m \mapsto \{\rho_H^*(m) : H \in \mathcal{A}_t^*\}$ ; it is single-valued when all elements share the feasible set and the conditional mean and risk functionals, and equal admissibility records alone do not imply this (Example 3.17 with equal defects gives equal records and different frontiers).

When Stage I ties, a predeclared implementation criterion may select one class before portfolio optimization; performance is never used to redefine admissibility.

The nested problem

$$\rho^{\text{joint}}(m) = \inf_{H \in \mathcal{A}_t} \rho_H^*(m) \quad (11)$$

is the  $\preceq_{\text{env}}$  problem of Proposition 3.20 and encodes a different contract, in which the representation may change with the target return. Since  $H_t^* \in \mathcal{A}_t$ , the envelope is never above the two-stage frontier,  $\rho^{\text{joint}}(m) \leq \rho_{H_t^*}^*(m)$ , with equality if and only if some admissibility-optimal class attains the target-specific minimum over the whole admissible class; and as the pointwise infimum of convex functions it is in general neither convex nor attached to a single class, whereas the two-stage frontier is convex by Proposition 3.21. Under APO,

$$\rho^{\text{two}}(m) = \rho_{H_t^*}^*(m), \quad H_t^* \in \mathcal{A}_t \text{ fixed before } m \text{ is varied.} \quad (12)$$

A strict inequality between the envelope and the two-stage frontier does not show that Stage I chose badly; the envelope may switch to a representation that is admissible but not admissibility-optimal (Figure 4).

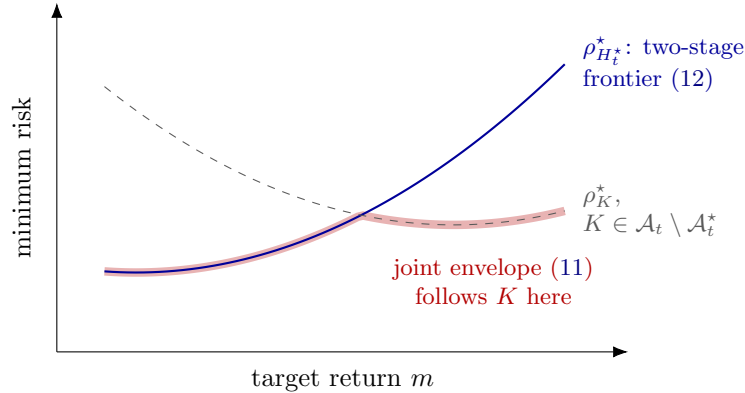


Figure 4: Two contracts. The shaded curve is the target-adaptive joint envelope  $\rho^{\text{joint}}$ , the lower envelope of the individually convex frontiers of  $H_t^*$  (solid) and  $K$  (dashed); it is not convex at the crossing and switches representation. The solid  $H_t^*$  branch is the admissibility-conditioned frontier, the canonical object whenever admissibility must be resolved before risk–return preference.

Admissibility determines which information may define the decision; it does not force variance to be the risk measure, and it does not by itself carry the classical objects, which are now recovered conditional on the fixed class.

### 3.11 Admissible tangent portfolio and capital allocation line

Let a riskless rate  $r_f$  be available and let  $H_t^*$  be the fixed class with conditional mean  $\mu_{H_t^*}$  and covariance  $\Sigma_{H_t^*} \succ 0$ .

The admissible tangent portfolio at  $t$  is the solution of  $\max\{(\mu_{H_t^*}^\top w - r_f)/\sqrt{w^\top \Sigma_{H_t^*} w} : w \in \mathcal{W}(H_t^*), \mathbf{1}^\top w = 1\}$ , and the admissible capital allocation line is the half-line from  $(0, r_f)$  through its risk-return point. Its slope  $S_t(H_t^*)$  is the admissible conditional Sharpe ratio; it is a random variable through the conditional moments, and  $\bar{S}_t(H) := \mathbb{E}[S_t(H)^2]^{1/2}$  is the unconditional Sharpe ratio of the conditional tangent strategy, the object of [29, 30].

For  $\mathcal{W}(H_t^*) = \{w : \mathbf{1}^\top w = 1\}$  the tangent portfolio is  $w_T \propto \Sigma_{H_t^*}^{-1}(\mu_{H_t^*} - r_f \mathbf{1})$  and  $S_t(H_t^*)^2 = (\mu_{H_t^*} - r_f \mathbf{1})^\top \Sigma_{H_t^*}^{-1}(\mu_{H_t^*} - r_f \mathbf{1})$ , as in the classical case; what changes is which  $\Sigma_{H_t^*}$  is allowed to enter. Two properties distinguish it from the tangent portfolio of an arbitrary conditioning set.

**Proposition 3.22** (Admissible Sharpe ratio). *Let the feasible set be  $\{w : \mathbf{1}^\top w = 1\}$  for every admissible representation. (a) For every  $H \in \mathcal{A}_t$ ,  $\bar{S}_t(H) \leq \sup_{G \in \mathcal{A}_t} \bar{S}_t(G)$ , and the difference for  $H = H_t^*$  is the admissibility price in Sharpe units; it is zero if and only if some Stage-I-optimal class attains the maximal unconditional Sharpe ratio over the admissible class. (b) If  $H_1 \preceq H_2$  are admissible and  $H_2$  embeds the  $H_1$  decision problem, then  $\bar{S}_t(H_2) \geq \bar{S}_t(H_1)$ : the unconditional Sharpe ratio of the conditional tangent strategy is nondecreasing along the information order inside the admissible class.*

The proof is in Appendix A.

Part (b) is the admissible form of the statement that conditioning information cannot lower the maximal Sharpe ratio [29], with the same qualification as Theorem 3.13: it holds inside the admissible class and under embedding, and fails for look-ahead information because that information is not admissible.

Because  $\mu_{H_t^*}$  and  $\Sigma_{H_t^*}$  depend on the decision state, the admissible CAL is one line per state and their union is a ruled surface in (risk, state, return) space, the *admissible capital allocation surface*  $\mu = r_f + S_t(H_t^*)(z) \sigma$ , whose root-mean-square slope over the state distribution is  $\bar{S}_t(H_t^*)$ . Figure 5 draws the geometry at one state and Figure 6 the line across states and the surface: the line of the admissible class is steeper at each state than that of a representation failing separation, and the classical line is a horizontal section of the surface.

Markowitz portfolio theory is the degenerate case in which a single admissible class  $\{H_0\}$  with  $\rho_{H_0}(w) = w^\top \Sigma w$  makes Stage I vacuous and Stage II the classical program, and the trivial information object gives the unconditional problem.

### 3.12 Diversification conditional on admissible information

Suppose the conditional mean is affine in a statistic of the driver path as in Proposition 3.3, so that with  $u_H := R - \mathbb{E}[R | H^+]$  the conditional return innovation decomposes as

$$R - \mu_H = B_H \xi_H + u_H, \quad \text{Cov}(\xi_H | H) = V_H, \quad \text{Cov}(u_H, \xi_H | H) = 0, \quad \Omega_H := \text{Cov}(u_H | H) \succeq 0, \quad (13)$$

where the orthogonality of  $u_H$  and  $\xi_H$  is a consequence of the construction,  $u_H$  having zero conditional mean given  $H^+$  while  $\xi_H$  is  $H^+$ -measurable. Then  $\Sigma_H = B_H V_H B_H^\top + \Omega_H$ , and a risk functional built on the response term alone agrees with covariance risk on  $\mathcal{W}(H)$  exactly when  $w \mapsto w^\top \Omega_H w$  is constant there, in particular when  $\Omega_H = 0$ ; otherwise the two rank portfolios differently, and the difference is the residual term the rest of this subsection studies.

By Proposition 3.3 an admissible representation may induce a response operator  $B_H \in \mathbb{R}^{n \times k}$  and an innovation covariance  $V_H \in \mathbb{R}^{k \times k}$ . Define the represented response risk and the response-based objective

$$Q_H^{\text{resp}} = B_H V_H B_H^\top, \quad J^{\text{resp}}(w; H) = \frac{\gamma}{2} w^\top Q_H^{\text{resp}} w - r_H^\top w. \quad (14)$$

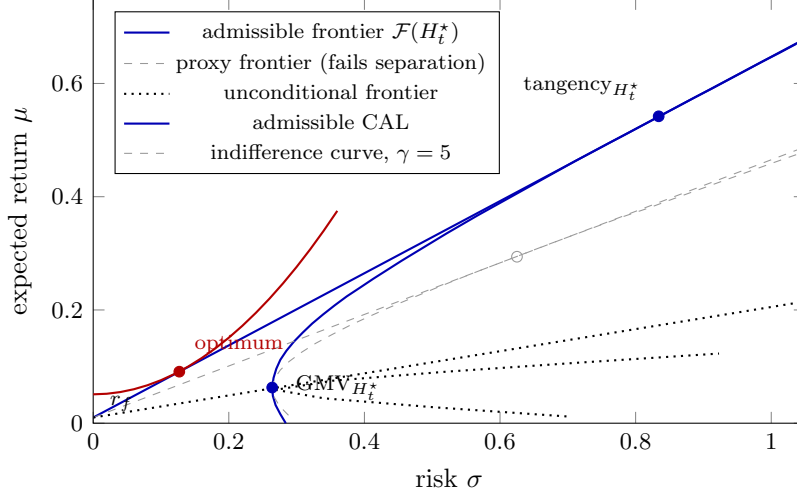


Figure 5: The mean–variance geometry of the admissible problem at one decision state ( $z = 0.3$ ), from the population moments of the E1 system with the admissibility-optimal class  $\{Z_1, Z_2\}$ ,  $r_f = 0.01$  and risk aversion  $\gamma = 5$ . The conditional efficient frontier of the admissible class carries its minimum-variance portfolio (GMV) and its tangency portfolio; the admissible capital allocation line runs from  $r_f$  through the tangency portfolio, and the investor’s indifference curve is tangent to it at the optimum. The frontier and line of a representation that fails the separator condition (the proxy, open marker at its tangency portfolio) and the unconditional frontier and line are flatter, so the same investor obtains less expected return per unit of risk; the difference is attributable to the choice of information.

Markowitz diversification acts on the quadratic form  $w \mapsto w^\top \Sigma w$ , and under an approximate factor structure the idiosyncratic part is diversified away while the factor part is not [18, 58]. Part (a) below is that argument; the content of the proposition is that the split it presupposes is a property of the conditioning class and is stable under interventions only under interventional admissibility.

**Proposition 3.23** (Diversification is a property of the admissible class). *Let (13) hold for a representation  $H$  and let  $\mathcal{W} \subseteq \{w \in \mathbb{R}^n : \mathbf{1}^\top w = 1\}$ .*

- (a) (Separation makes the residual idiosyncratic.) *If  $H$  satisfies the separator condition  $G_S$  in the exact sense of Definition 3.2, and  $\xi_H$  in (13) is the innovation of  $H^+$  relative to  $H^-$ , then  $\Omega_H$  is diagonal, and for every  $w \in \mathcal{W}$*

$$w^\top \Omega_H w \leq \left( \max_i \Omega_{H,ii} \right) \|w\|_2^2, \quad (15)$$

*so the residual risk of an equal-weight portfolio, or of any portfolio whose squared Euclidean norm is of order  $1/n$ , is of order  $\max_i \Omega_{H,ii}/n$  and vanishes along any sequence of models in which  $n \rightarrow \infty$  with  $\sup_i \Omega_{H,ii}$  uniformly bounded, the approximate-factor-structure setting of [18]. The systematic component  $w^\top B_H V_H B_H^\top w$  admits no such bound: it equals  $\beta_w^\top V_H \beta_w$  with  $\beta_w := B_H^\top w$  the portfolio’s exposure to the common drivers, and is unaffected by spreading weights among assets with equal exposures.*

- (b) (Failure of the admissibility condition misclassifies risk.) *If  $H$  fails  $G_S$ , so that  $\Omega_H$  has a nonzero off-diagonal block  $\Omega_H^{\text{off}}$ , then  $w^\top \Omega_H w$  equals  $\sum_i w_i^2 \Omega_{H,ii}$  plus  $w^\top \Omega_H^{\text{off}} w$ , and the second term does not vanish under spreading: for equal weights it equals  $n^{-2} \mathbf{1}^\top \Omega_H^{\text{off}} \mathbf{1}$ , whose order is that of the average off-diagonal residual covariance rather than  $1/n$ . Common variation not explained by  $H$  is then counted as diversifiable when it is not.*

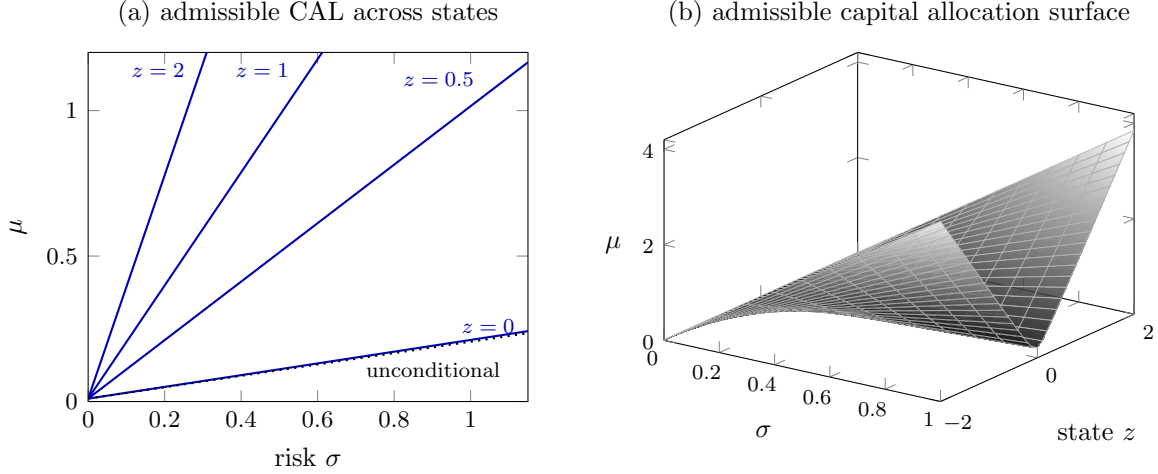


Figure 6: The admissible capital allocation line across decision states and the surface it sweeps, E1 system, class  $\{Z_1, Z_2\}$ ,  $r_f = 0.01$ . (a) The admissible CAL at four states; its slope is the conditional Sharpe ratio  $S_t(H_t^*)(z)$  and grows with the information carried by the state, while the unconditional line does not move. (b) The union of the lines over states is the admissible capital allocation surface  $\mu = r_f + S_t(H_t^*)(z) \sigma$ ; its root-mean-square slope over the state distribution is the unconditional Sharpe ratio  $\bar{S}_t(H_t^*)$  of the conditional tangent strategy, 1.90 against 1.29 for the proxy and 0.20 unconditionally, and the classical capital allocation line is one horizontal section of it.

- (c) (Interventional admissibility makes the split invariant.) *Let  $H$  be admissible for an interventional contract with declared regime set  $\mathcal{E}$ , so that  $\mathcal{L}^e(R \mid H) = \mathcal{L}^o(R \mid H)$  for every  $e \in \mathcal{E}$  by Definition 3.6. Then for every  $e \in \mathcal{E}$  the decomposition (13) and the diagonality in (a) hold under the post-intervention law with the same  $B_H$  and  $\Omega_H$ , so the diversifiable and non-diversifiable components are the same before and after the intervention.*
- (d) (What the failure costs a given portfolio.) *Let  $s_i = \Omega_{H,ii}^{1/2}$ ,  $v = w \odot s$ ,  $c_{ij} = \Omega_{H,ij}/(s_i s_j)$  the residual correlations, and  $D = (\sum_i v_i)^2 - \sum_i v_i^2$ . If  $D \neq 0$ , the variance that the diagonal-residual restriction discards satisfies the exact decomposition*

$$\frac{w^\top \Omega_H^{\text{off}} w}{w^\top \Sigma_H w} = \underbrace{\frac{\|v\|_2^2}{w^\top \Sigma_H w}}_{\text{residual share}} \times \underbrace{\left( \frac{(\sum_i v_i)^2}{\|v\|_2^2} - 1 \right)}_{\text{breadth}-1} \times \underbrace{\frac{1}{D} \sum_{i \neq j} v_i v_j c_{ij}}_{\bar{c}_w}, \quad (16)$$

where  $\bar{c}_w$  is the portfolio-weighted average residual correlation. If  $v \geq 0$  then  $|\bar{c}_w| \leq \varepsilon_t(H)$ , so the relative misallocation is at most the screening defect times the residual share times breadth minus one. The defect therefore bounds the cost but does not order it: it is a maximum over pairs, whereas the cost is governed by  $\bar{c}_w$ , a weighted average.

The proof is in Appendix A. Under a conditional-law contract part (c) has no basis: an intervention may sever the channel through which  $H$  explained the covariation, after which  $\Omega_H$  acquires off-diagonal mass and part (b) applies; E3 measures this case. The proposition is the portfolio-level form of the principle of common drivers stated in Proposition 3.4: Markowitz diversification acts on  $\Sigma$ , admissible diversification on the residual  $\Omega_{H_t^*}$  of the selected class, and causal diversification requires the class to be interventionally admissible.

Table 1: Correspondence between the objects of modern portfolio theory and their admissible counterparts.

| Modern portfolio theory                     | Admissible portfolio optimization                                                                                                                           | What changes                                                                              |
|---------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------|
| moments $\mu, \Sigma$ taken as given        | conditional moments $\mu_H, \Sigma_H$ of an admissible $H$                                                                                                  | the information is a decision variable subject to gates                                   |
| feasible set of portfolios                  | admissible pairs $\Gamma_t(m)$                                                                                                                              | pairs, not weights; inadmissible $H$ absent                                               |
| efficient frontier                          | conditional frontier of $H_t^*$ (Def. 3.10)                                                                                                                 | one frontier per class; the reported one is that of the admissibility-optimal class       |
| minimum-variance portfolio                  | conditional minimum-variance portfolio of $H_t^*$                                                                                                           | unchanged in form                                                                         |
| tangency portfolio, capital allocation line | admissible tangent portfolio and admissible CAL, one line per decision state; their union over states is the admissible capital allocation surface (Fig. 6) | the slope becomes state-dependent; its root-mean-square is the unconditional Sharpe ratio |
| indifference curve, utility                 | the same preference, applied only in Stage II                                                                                                               | preference cannot reselect information (Def. 3.19)                                        |
| —                                           | joint envelope over $\mathcal{A}_t^0$ (Sec. 3.10)                                                                                                           | a different contract, not an object of the classical theory                               |
| diversification acts on $\Sigma$            | diversification acts on the residual $\Omega_{H_t^*}$ (Prop. 3.23)                                                                                          | the diversifiable part is fixed by the class                                              |
| estimation error in $\mu, \Sigma$           | estimation error in $\mu_H, \Sigma_H$ plus selection error in $H$ (Sec. 4)                                                                                  | a second source of error with its own bound                                               |

### 3.13 Causal admissibility and its price

Causality changes the feasible information class when the claim requires interventional meaning. With the claim levels of Definition 3.6, let  $\mathcal{A}_t(c)$  denote the representations admissible for a conditional-law contract and  $\mathcal{A}_t(i) \subseteq \mathcal{A}_t(c)$  those admissible for an interventional contract, i.e. additionally carrying an identified intervention response under the declared design.

For an interventional contract,

$$V_i^* = \inf_{H \in \mathcal{A}_t(i)} \inf_{w \in \mathcal{W}(H)} J(w; H); \quad (17)$$

for a conditional-law contract the problem is posed over  $\mathcal{A}_t(c)$  with value  $V_c^*$ . For the same population loss and nested classes the causal admissibility price is

$$\Delta_{ci} := V_i^* - V_c^*. \quad (18)$$

Since  $\mathcal{A}_t(i) \subseteq \mathcal{A}_t(c)$ ,  $V_c^* \leq V_i^*$  and  $\Delta_{ci} \geq 0$ . The inequality is not evidence against causality, since the interventional problem solves a stronger semantic contract, and  $\Delta_{ci}$  is the oracle price of that contract at fixed numerical loss. Conversely a lower conditional-law loss does not license an interventional interpretation.

**Theorem 3.24** (Distance bound for the causal admissibility price). *Let  $(\mathcal{I}, d)$  be a metric space of decision-equivalence classes containing  $\mathcal{A}_t(c)$  and suppose  $V$  is  $L$ -Lipschitz on  $\mathcal{A}_t(c)$ , i.e.  $|V(H) - V(G)| \leq L d(H, G)$  for  $H, G \in \mathcal{A}_t(c)$ . If  $H_0^* \in \mathcal{A}_t(c)$  satisfies  $V(H_0^*) = V_c^*$  and  $\mathcal{A}_t(i) \neq \emptyset$ , then*

$$0 \leq \Delta_{ci} \leq L d(H_0^*, \mathcal{A}_t(i)), \quad d(H, \mathcal{A}_t(i)) := \inf_{G \in \mathcal{A}_t(i)} d(H, G). \quad (19)$$

For the quadratic loss  $J(w; H) = \frac{\gamma}{2} w^\top \Sigma_H w - \mu_H^\top w$  on a common feasible set with  $\bar{r} := \sup_{w \in \mathcal{W}} \|w\|_2 < \infty$ , the hypothesis holds with  $L = \bar{r}$  for the moment metric

$$d(H, G) := \|\mu_H - \mu_G\|_2 + \frac{\gamma}{2} \bar{r} \|\Sigma_H - \Sigma_G\|_{\text{op}}, \quad (20)$$

and, when the moments are random functions of the decision state, for its expectation over the state.

The proof is in Appendix A.

On a finite class every function is Lipschitz in the discrete metric, so the bound has content only for a decision-relevant metric such as (20).

The causal admissibility price is therefore controlled by the distance, in conditional mean and covariance, between the best predictive representation and the nearest interventionally admissible one; in E3 this distance is zero before the intervention, which is why proxy and state tie. Requiring identification removes otherwise useful representations in exchange for interventional, transport or counterfactual meaning; it is neither a general advantage nor a general prerequisite. When identification establishes an intervention-response map  $\Psi_H^{\text{do}}$ , the response operator may be taken as its derivative at the operating point,  $B_H^{\text{do}} = D\Psi_H^{\text{do}}(z_0)$ , and only then may a response-based objective be called causal-response optimization; a causal graph selected by an algorithm does not by itself supply it.

### 3.14 Perturbation of the conditional portfolio

The finite-sample results of Section 4 rest on one population fact about the quadratic Stage-II problem, stated here because it involves no sampling: the regret of a portfolio computed from perturbed conditional moments is second order in the perturbation.

**Proposition 3.25** (Regret is second order in estimation error). *Let  $J(w) = \frac{\gamma}{2} w^\top \Sigma w - \mu^\top w$  and  $\hat{J}(w) = \frac{\gamma}{2} w^\top \hat{\Sigma} w - \hat{\mu}^\top w$  on a convex feasible set  $\mathcal{W}$  with  $\sup_{\mathcal{W}} \|w\|_2 \leq \bar{r}$ , and suppose  $\Sigma \succeq \lambda I$  with  $\lambda > 0$ . Let  $w^*$  minimize  $J$  and  $\hat{w}$  minimize  $\hat{J}$  over  $\mathcal{W}$ . Then*

$$J(\hat{w}) - J(w^*) \leq \frac{2}{\gamma\lambda} \left( \gamma\bar{r} \|\hat{\Sigma} - \Sigma\|_{\text{op}} + \|\hat{\mu} - \mu\|_2 \right)^2. \quad (21)$$

In particular, if  $\hat{\mu}$  and  $\hat{\Sigma}$  are  $\sqrt{N}$ -consistent, the regret of the estimated portfolio is  $O_p(N^{-1})$ .

The proof is in Appendix A.

The frontier is first order in the moments and inherits their  $N^{-1/2}$  error; the regret at a declared risk aversion is second order and converges at  $N^{-1}$ , which is why the sensitivity of the frontier to errors in means does not carry over to the decision.

## 4 Finite-sample theory and evidence

The population statements above are about oracles. This section prices the selection of the representation from data: consistency of both stages, the regret of joint selection, the oracle inequality that trades the causal price against search complexity, and reselection across decision nodes.

## 4.1 Consistency of the two-stage estimator

Let  $\hat{\mathcal{A}}_t^0$  be produced on a certification block and the four quantities of Definition 3.10 estimated without the final evaluation block, and define  $\hat{\mathcal{A}}_t^* = \text{Min}_{\succsim_{\mathcal{A}}} \hat{\mathcal{A}}_t^0$  with  $\succsim_{\mathcal{A}}$  the admissibility order computed from the estimated quantities. For a finite class the admissibility order is a total preorder on a finite set and therefore admits a real-valued representation: there is  $s_t : \mathcal{A}_t \rightarrow \mathbb{R}$  with  $H_1 \prec_{\mathcal{A}} H_2 \iff s_t(H_1) < s_t(H_2)$ , and  $\mathcal{A}_t^* = \arg \min s_t$ . We state consistency for such a score.

**Proposition 4.1** (Stage-I selection consistency, finite class). *Let  $\mathcal{A}_t$  be finite, let  $s_t$  represent  $\prec_{\mathcal{A}}$  and  $\hat{s}_t$  represent  $\succsim_{\mathcal{A}}$  on  $\hat{\mathcal{A}}_t^0$ , and suppose (i)  $\mathbb{P}\{\hat{\mathcal{A}}_t^0 = \mathcal{A}_t\} \rightarrow 1$  and (ii)  $\max_{H \in \mathcal{A}_t} |\hat{s}_t(H) - s_t(H)| \xrightarrow{p} 0$ . If  $\Delta_a := \min\{s_t(H) - \min s_t : H \in \mathcal{A}_t \setminus \mathcal{A}_t^*\} > 0$ , then  $\mathbb{P}\{\hat{\mathcal{A}}_t^* \subseteq \mathcal{A}_t^*\} \rightarrow 1$ , and  $\mathbb{P}\{\hat{\mathcal{A}}_t^* = \mathcal{A}_t^*\} \rightarrow 1$  whenever  $\hat{s}_t$  takes equal values on  $\mathcal{A}_t^*$  with probability tending to one.*

The proof is in Appendix A.

**Proposition 4.2** (Stage-II frontier consistency). *Assume Stage I consistently selects a decision-equivalent class  $[H_t^*]$  and write  $\mathcal{W} := \mathcal{W}(H_t^*)$ ,  $\rho := \rho_{H_t^*}$ ,  $\mu := \mu_{H_t^*}$ . Suppose (i)  $\mathcal{W}$  is known, nonempty, convex and compact,  $\rho$  is continuous and convex, and  $\mu$  is continuous and affine on  $\mathcal{W}$ ; (ii) the estimators  $\hat{\rho}_N, \hat{\mu}_N$  are continuous on  $\mathcal{W}$  and  $\eta_N := \sup_{\mathcal{W}} |\hat{\rho}_N - \rho| + \sup_{\mathcal{W}} |\hat{\mu}_N - \mu| \xrightarrow{p} 0$ ; (iii) the target interval  $\mathcal{M} = [m_-, m_+]$  satisfies  $m_- < m_+ < \max_{\mathcal{W}} \mu$ . Let  $\hat{\rho}_N^*(m) = \min\{\hat{\rho}_N(w) : w \in \mathcal{W}, \hat{\mu}_N(w) \geq m\}$  and  $\hat{\mathcal{E}}_N(m)$  the corresponding minimizer set. Then*

$$\sup_{m \in \mathcal{M}} |\hat{\rho}_N^*(m) - \rho^*(m)| \xrightarrow{p} 0, \quad (22)$$

and every accumulation point of every sequence  $\hat{w}_N \in \hat{\mathcal{E}}_N(m)$  belongs to  $\mathcal{E}_t^*(m)$ .

The proof is in Appendix A.

Three qualifications concern the sample. If the feasible set is estimated with Hausdorff distance  $\delta_N \xrightarrow{p} 0$  and  $\rho, \mu$  are Lipschitz near  $\mathcal{W}$ , the sandwich in the proof of Proposition 4.2 holds with  $\delta_N$ -neighbourhoods and an additional term  $L_\rho \delta_N$ , and the conclusion is unchanged. All finite-sample statements are for independent observations. Returns are serially dependent; under  $\beta$ -mixing the same conclusions hold with contiguous blocks separated by a gap in place of independent observations and with the effective sample size reduced accordingly, by the blocking arguments of [61]. The certification, estimation and evaluation blocks are disjoint for a second reason: the frontier is estimated after a selection over the candidate set, and without splitting the Stage-II estimates would be subject to the selection effect and their confidence statements invalid. Splitting is the post-selection device that makes Proposition 4.2 conditional on the selected class alone.

## 4.2 The price of joint selection

Let  $\hat{J}_N(w; H)$  be an estimated loss on a selection sample and

$$(\hat{H}, \hat{w}) \in \arg \min_{H \in \hat{\mathcal{A}}_N, w \in \mathcal{W}(H)} \hat{J}_N(w; H), \quad (23)$$

where the empirical admissible set  $\hat{\mathcal{A}}_N$  is itself estimated and can make false-pass or false-fail errors. Consider first the case in which the population admissible class is known.

**Proposition 4.3** (Uniform decision-regret bound). *Assume  $\mathcal{A}$  is known and that on an event  $E_N$  of probability at least  $1 - \delta$ ,*

$$\sup_{H \in \mathcal{A}} \sup_{w \in \mathcal{W}(H)} |\hat{J}_N(w; H) - J(w; H)| \leq \varepsilon_N. \quad (24)$$

Let  $(H^*, w^*)$  minimize  $J$  and  $(\hat{H}, \hat{w})$  minimize  $\hat{J}_N$  over  $\{(H, w) : H \in \mathcal{A}, w \in \mathcal{W}(H)\}$ . Then on  $E_N$ ,

$$J(\hat{w}; \hat{H}) - J(w^*; H^*) \leq 2\varepsilon_N. \quad (25)$$

The proof is in Appendix A.

Representation choice raises the statistical burden because the uniform deviation must hold over the joint class, not over the weights of one model.

If the admissible class is itself estimated by a certification procedure independent of the selection sample, and on an event of probability at least  $1 - \alpha$  that procedure includes no inadmissible representation and retains a population-optimal class, then (25) holds over the certified class with probability at least  $1 - \alpha - \delta$  by the same chain of inequalities and the union bound; outside that event no admissibility or causal conclusion follows from decision performance.

Hence the sample is split, so that certification decides eligibility on one block and only certified representations are ranked on another; reusing the same outcomes to relax the admissibility conditions destroys the interpretation of admissibility.

### 4.3 Oracle restriction versus representation-search complexity

The generic bound does not explain when a stronger admissibility condition can help statistically. The bound reflects the approximation–estimation tradeoff of oracle inequalities for model selection [7], here for a claim-dependent admissible class.

**Theorem 4.4** (Finite-class oracle inequality). *Let  $\mathcal{A}_r = \{H_1, \dots, H_{M_r}\}$  be the admissible class for a conditional-law ( $r = c$ ) or interventional ( $r = i$ ) contract, with  $\mathcal{A}_t(i) \subseteq \mathcal{A}_t(c)$ . Suppose the inner optimized empirical values  $\hat{V}_N(H)$  are computed on data independent of condition certification and satisfy, for every fixed  $H \in \mathcal{A}_t(c)$  and every  $x > 0$ ,*

$$\mathbb{P}\{|\hat{V}_N(H) - V(H)| > x\} \leq 2 \exp\{-Nx^2/(2\sigma^2)\}. \quad (26)$$

Let  $\hat{H}_r \in \arg \min_{H \in \mathcal{A}_r} \hat{V}_N(H)$ . Then for every  $\delta \in (0, 1)$ , with probability at least  $1 - \delta$ ,

$$V(\hat{H}_r) - \min_{H \in \mathcal{A}_r} V(H) \leq 2\sigma \sqrt{\frac{2 \log(2M_r/\delta)}{N}}, \quad (27)$$

and, relative to the conditional-law oracle,

$$V(\hat{H}_i) - V_c^* \leq \Delta_{ci} + 2\sigma \sqrt{\frac{2 \log(2M_i/\delta)}{N}}. \quad (28)$$

The proof is in Appendix A.

**Corollary 4.5** (Causal admissibility as regularization). *Suppose  $M_i < M_c$ . The interventional class has larger oracle loss by  $\Delta_{ci}$  and a smaller search term, and the upper bound (28) is tighter than the conditional-law bound (27) with  $r = c$  whenever*

$$\Delta_{ci} < 2\sigma \sqrt{\frac{2}{N}} \left[ \sqrt{\log(2M_c/\delta)} - \sqrt{\log(2M_i/\delta)} \right]. \quad (29)$$

The proof is in Appendix A.

This is structural risk minimization [60] over a nesting fixed by the claim rather than tuned to the data, with the causal price as approximation term and the size of the eligible class as complexity term; it is a regularization statement, not a claim that causal representations predict better. For infinite classes  $\log M_r$  is replaced by a complexity measure and the deviation assumption by a uniform empirical-process bound [13]; when candidates are nested a chaining argument sharpens the logarithmic term.

#### 4.4 Several decision nodes

Consider a sequence of decision nodes  $t = 1, \dots, T$ , each with its own contract, candidate set and admissible class, and a portfolio process  $(w_t)$  that may be constrained intertemporally, for instance by transaction costs  $c(w_t - w_{t-1})$ . The multi-period problem is to choose an admissible pair at every node.

When the admissibility conditions and quantities at node  $t$  depend only on node- $t$  information, the multi-period Stage-I problem is the product of the one-node problems, whatever the intertemporal constraints on portfolios, because nothing in the order at node  $t$  involves a portfolio; given the selected classes, Stage II is a dynamic program coupled only through the cost  $c(w_t - w_{t-1})$ . What remains is how the *estimated* class should change between nodes.

When the problem is solved at successive nodes  $t = 1, 2, \dots$ , Stage-I reselection is itself a source of turnover. Let  $\hat{s}_t$  be the estimated score of Proposition 4.1 and  $\hat{H}_{t-1}$  the incumbent class.

The rule proposed here is a hysteresis rule: for a predeclared margin  $\eta > 0$ , set  $\hat{H}_t = \hat{H}_{t-1}$  if  $\hat{s}_t(\hat{H}_{t-1}) \leq \min_{\hat{\mathcal{A}}_t^0} \hat{s}_t + \eta$ , and  $\hat{H}_t \in \arg \min \hat{s}_t$  otherwise.

**Proposition 4.6** (Reselection only on population change). *Suppose  $\max_H |\hat{s}_t(H) - s_t(H)| \leq \delta_t$  at every node. (a) A reselection at node  $t$  implies  $s_t(\hat{H}_{t-1}) - \min s_t > \eta - 2\delta_t$ . (b) If the four population quantities are constant in  $t$ , the incumbent is population-optimal and  $\delta_t < \eta/2$  for all  $t$ , no reselection occurs. (c) Every reselection satisfies  $s_t(\hat{H}_t) \leq \min s_t + 2\delta_t$ .*

The proof is in Appendix A.

The margin  $\eta$  is in the units of the score and must exceed twice the estimation error for part (b); under a rolling protocol the incumbent is certified only on observations that post-date its fixing. A larger margin suppresses reselection driven by noise at the cost of a longer lag when the population changes, and because it is declared before any portfolio is evaluated it does not reintroduce performance into Stage I.

#### 4.5 Implementation, and comparison with mean–variance practice

The procedure, of which the public-data experiment of Section 4.7 is a worked instance, is the following.

- (1) *Declare the contract.* Fix the asset universe, horizon, rebalancing calendar, portfolio constraints, claim level with its regime set, candidate driver universe with its release calendar, tolerance  $\bar{\epsilon}$  and tie-break quantities, before looking at returns; nothing in later steps may alter them.
- (2) *Split the sample.* For each window, three disjoint blocks, certification, estimation and evaluation; the split is the post-selection device of Section 4.
- (3) *Certify.* On the certification block apply the required conditions to each candidate: chronology and publicity from the release calendar, separation by the maximal absolute residual correlation after conditioning, causality by invariance across the declared regimes; record the four quantities of Definition 3.10 for every candidate that passes.
- (4) *Select.* Rank the passing candidates by the order of Definition 3.10(b) and fix the minimal class; if none passes, the admissible class is empty and the output is that the declared universe does not support the claim at the declared tolerance. Across rebalancing dates apply the hysteresis rule.
- (5) *Estimate and optimize.* On the estimation block regress returns on the selected drivers to obtain  $\hat{B}$ ,  $\hat{V}$  and  $\hat{\Omega}$ , form  $\hat{B}\hat{V}\hat{B}^\top + \text{diag}(\hat{\Omega})$  and the fitted conditional mean at the decision-time state, and solve the classical problem inside the selected class only.

- (6) *Evaluate once.* On the evaluation block report realized risk, the realized-to-predicted variance ratio, out-of-sample residual dependence, turnover and reselections, without returning to step 4.

The comparison with mean–variance practice is then a comparison of two things. The first is the information decision, absent from mean–variance practice: which candidates passed, which class was selected, and whether the class was empty. The second is the portfolio, on the same evaluation block and constraints, against the sample covariance, a shrinkage estimator and a principal-component estimator of matched dimension, on realized variance and on the realized-to-predicted ratio. The theory predicts the pattern. When the selected class satisfies the separator condition the conditional covariance is correctly specified and its ratio should be near one; when no class passes and the procedure is run anyway, the diagonal residual understates risk and the ratio exceeds that of the sample covariance, as Section 4.7 finds.

## 4.6 Controlled experiments

Experiments E1–E4 are controlled, with a known data-generating structure against which every claim can be checked. Figure 7 shows the fact they share. On a single sample the frontiers that dominate in the backtest are those of candidates containing look-ahead information, which do not exist at the decision time. They are implemented as specified in Appendix B, and the code accompanies the paper. The market experiment on public data is reported in Section 4.7.

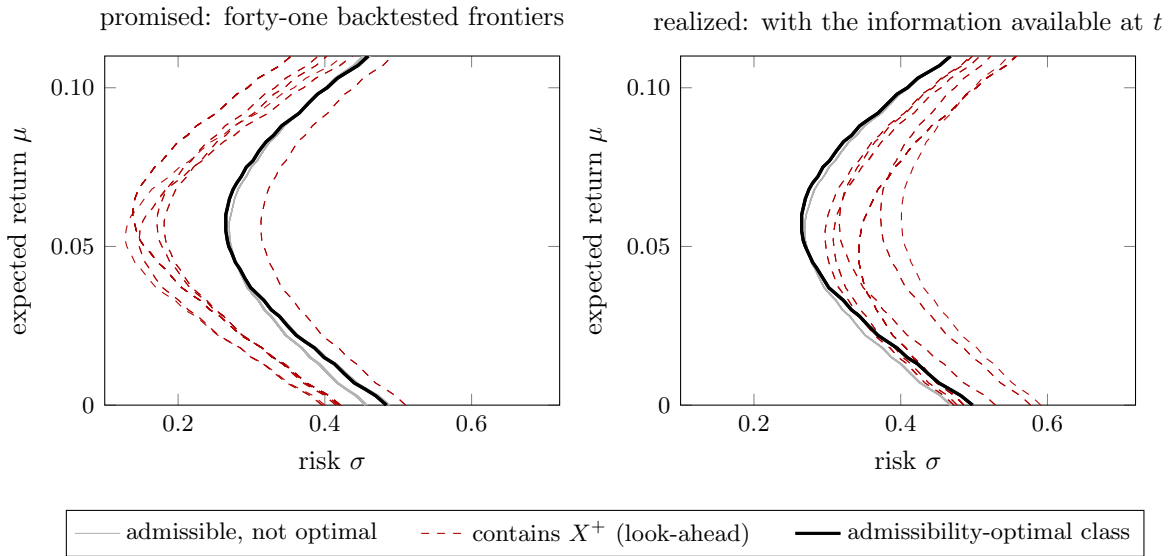


Figure 7: Backtest versus reality on the E1 system, one sample of 2,000 observations. Left: the in-sample frontier of every one of the 41 candidate conditioning sets, at the mean state. The frontiers that dominate are those containing the look-ahead variable  $X^+$  (dashed); the admissibility-optimal class (thick) is not the best backtest. Right: the same portfolios evaluated under the information actually available at the decision time, so that  $X^+$  is replaced by what is known of it, namely nothing. The look-ahead frontiers move to the right of the admissible frontier, whose promised and realized risks nearly coincide. Selecting information on backtest performance would have chosen the frontiers that do not exist; the chronological admissibility condition removes them before any frontier is drawn.

#### 4.6.1 Exact recovery of the admissible class and Markowitz nesting (E1)

A Gaussian structural system with eight assets and two common states generates six candidate variables: the states  $Z_1, Z_2$ , an affine recoding  $\phi(Z_1) = 3Z_1 - 2$ , a noisy proxy  $P = Z_1 + 0.7e$ , an independent noise  $\nu$  and a look-ahead variable  $X^+$  equal to a sum of horizon idiosyncratic shocks. The candidate set contains the 41 nonempty subsets of size at most three; the chronological admissibility condition removes the 16 containing  $X^+$ . The population screening defect  $\varepsilon_t(H)$  is the maximal absolute conditional residual correlation, computed analytically. The four identities hold. The look-ahead candidate has  $\varepsilon_t = 0.435$  and is removed before ranking irrespective of its predictive value. The sets  $\{Z_1, Z_2\}$  and  $\{\phi(Z_1), Z_2\}$  have identical conditional covariances (maximal entrywise difference  $1.1 \times 10^{-16}$ ) and both attain  $\varepsilon_t = 0$  at dimension two, so Stage I returns their common class; every other admissible candidate has  $\varepsilon_t \geq 0.299$  (the proxy pair  $\{P, Z_2\}$ ) or dimension three. The Stage-II frontier computed from the fixed class coincides with Markowitz computed from the same conditional moments to machine precision (maximal difference 0 on a 41-point target grid).

#### 4.6.2 Finite-sample selection and estimation (E2)

The same system is estimated from  $N$  observations split 50/25/25 into discovery, certification and evaluation blocks, with  $\hat{\varepsilon}_t$  the maximal absolute residual correlation on the certification block, tolerance  $0.10 + 2/\sqrt{N_{\text{cert}}}$  (the declared level plus the 97.5% Fisher bound for a correlation; the size-controlled multiplier-bootstrap calibration of the maximum [19] is reported in Appendix B), ties broken by dimension and then by  $\hat{\varepsilon}_t$ , 300 replications. Table 2 and Figure 8 report the Stage-I error rates and the Stage-II frontier error  $\sup_m |\hat{\rho}^*(m) - \rho^*(m)|$  over the central 60% of the target range. Class recovery reaches 0.99 at  $N = 1,000$  and 1 from  $N = 2,000$  (Proposition 4.1). The frontier error with estimated means decreases slowly, the mean-error sensitivity of [20]; with the mean known it halves for each fourfold increase in  $N$ . The last column shows that the look-ahead representation attains lower in-sample decision loss than the true class in essentially every replication. The decision-relevant quantity is the regret  $J(\hat{w}_t; H_t^*) - J(w^*; H_t^*)$  at the declared risk aversion under the true moments, and Table 3 shows it decaying as  $N^{-1}$ , the rate of Proposition 3.25, from 0.035 at  $N = 250$  to 0.0008 at  $N = 8,000$ , with Jorion-type shrinkage [37] of the conditional mean reducing it by 35% at  $N = 250$  and 12% at  $N = 8,000$ . The global-minimum-variance error, a covariance-only quantity, decays at the parametric rate (0.0124 to 0.0013), and the multiplier-bootstrap tolerance reaches class recovery one from  $N = 1,000$ ; both are in Appendix B.

Table 2: E2: Stage-I selection and Stage-II frontier error against sample size (300 replications; medians for frontier errors). Rates carry a binomial Monte Carlo standard error of at most 0.03.

| $N$   | false admission | false exclusion | class recovery | frontier error | frontier error, known mean | look-ahead wins on loss |
|-------|-----------------|-----------------|----------------|----------------|----------------------------|-------------------------|
| 250   | 0.173           | 0.113           | 0.753          | 0.094          | 0.0336                     | 0.997                   |
| 500   | 0.047           | 0.053           | 0.923          | 0.090          | 0.0243                     | 1.000                   |
| 1,000 | 0.003           | 0.007           | 0.993          | 0.078          | 0.0158                     | 1.000                   |
| 2,000 | 0.000           | 0.000           | 1.000          | 0.073          | 0.0103                     | 1.000                   |
| 4,000 | 0.000           | 0.000           | 1.000          | 0.065          | 0.0073                     | 1.000                   |
| 8,000 | 0.000           | 0.000           | 1.000          | 0.054          | 0.0050                     | 1.000                   |

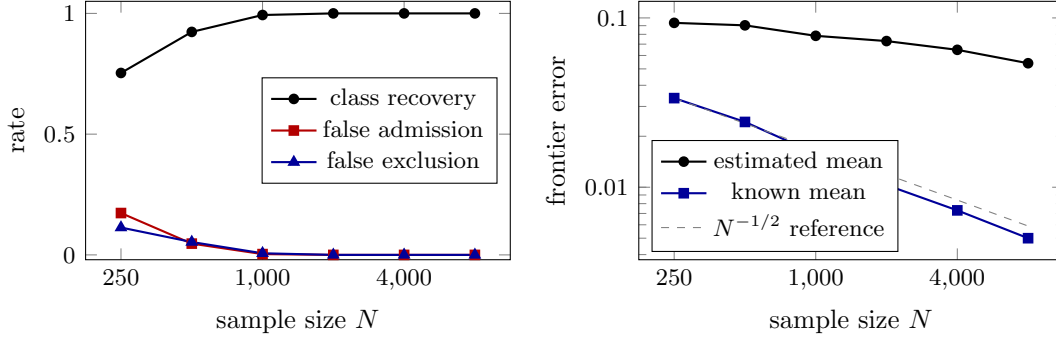


Figure 8: E2. Left: Stage-I error rates against sample size (300 replications). Right: median Stage-II frontier error  $\sup_m |\hat{\rho}^*(m) - \rho^*(m)|$  with the conditional mean estimated and with the mean known; the dashed line is proportional to  $N^{-1/2}$ . Selection is consistent, the covariance-driven part of the frontier error decays at the parametric rate, and the mean-driven part is the classical sensitivity to errors in means.

Table 3: E2: median decision regret  $J(\hat{w}; H_t^*) - J(w^*; H_t^*)$  at  $\gamma = 5$  under the true conditional moments, with the sample mean and with Jorion-type shrinkage toward the grand mean (300 replications).

| $N$   | sample mean | shrunk mean |
|-------|-------------|-------------|
| 250   | 0.0350      | 0.0228      |
| 500   | 0.0159      | 0.0115      |
| 1,000 | 0.0076      | 0.0055      |
| 2,000 | 0.0038      | 0.0028      |
| 4,000 | 0.0017      | 0.0014      |
| 8,000 | 0.0008      | 0.0007      |

#### 4.6.3 Causal admissibility under an intervention (E3)

A one-state system contains the structural state  $Z$  and a proxy  $P = Z + 0.05e$  calibrated so that both have screening defect 0.036 before intervention, indistinguishable at  $N = 4,000$ . The intervention  $\text{do}(P := \tilde{P})$  replaces the proxy by an independent variable with the same marginal and leaves the mechanism from  $Z$  to returns unchanged. The interventional admissibility condition is implemented as invariance of the conditional regression across the two regimes (Definition 3.6); under the null of two observational regimes the same test admits both  $Z$  and  $P$  in every one of 300 replications, so its size at the declared tolerances is below  $1/300$ , and the rejections of  $P$  reported below are not false rejections. Portfolios are the unconstrained conditional mean–variance weights normalized to full investment, a deliberate choice that isolates the effect of the information from that of constraints. Table 4 and Figure 9 report 200 replications. Under a conditional-law contract Stage I retains the proxy in 47% of replications, the two candidates being tied; the interventional admissibility condition admits  $Z$  in every replication and  $P$  in none. After the intervention the proxy’s screening defect rises to 0.563 and its response error to 1.25, while both quantities for  $Z$  are unchanged. The portfolio fixed on proxy-conditioned moments incurs a post-intervention loss of 0.157 against 0.136 for the state-conditioned portfolio; in the loss units of (6) with  $\gamma = 5$  this is a certainty-equivalent return gap of 0.021, about one fifth of the average conditional mean return in the design. The experiment measures the statement of Proposition 3.23(c), that the split of risk

survives the intervention only for the interventionally admissible representation.

Table 4: E3: structural state versus decision-equivalent proxy before and after an intervention severing the proxy channel (means, standard deviations in parentheses, 200 replications).

|                                                               | structural state $Z$ | proxy $P$     |
|---------------------------------------------------------------|----------------------|---------------|
| screening defect $\hat{\varepsilon}_t$ , before intervention  | 0.036 (0.007)        | 0.037 (0.007) |
| screening defect $\hat{\varepsilon}_t$ , after intervention   | 0.036 (0.007)        | 0.563 (0.011) |
| response error $\ \hat{B} - B\ _2$ , after intervention       | 0.035 (0.009)        | 1.248 (0.022) |
| retained by conditional-law Stage I (fraction)                | 0.530                | 0.470         |
| admitted by interventional admissibility condition (fraction) | 1.000                | 0.000         |
| post-intervention loss of the fixed portfolio                 | 0.136 (0.010)        | 0.157 (0.024) |

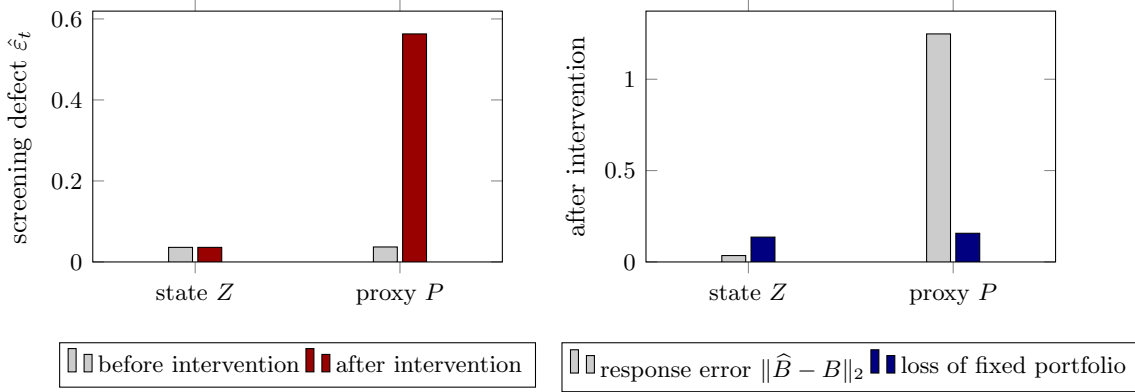


Figure 9: E3. Left: screening defect of the structural state and of the decision-equivalent proxy before and after the intervention severing the proxy channel; indistinguishable before, separated after. Right: post-intervention response error and loss of the portfolio fixed on each representation’s pre-intervention moments. Means over 200 replications.

#### 4.6.4 Reselection across decision nodes (E4)

Forty decision nodes are simulated, each with a fresh sample of 200 observations of which 100 certify; at node 20 the second common state changes from  $Z_2$  to an independent  $Z_3$ , so the population-optimal class changes once. Candidates are the subsets of size at most two of  $\{Z_1, Z_2, Z_3, P, \nu\}$ , scored by  $\hat{\varepsilon}_t + 0.02 \kappa_t$ , a real representation of the admissibility order at the tolerance scale. The hysteresis rule of the hysteresis rule is applied with margins  $\eta \in \{0, 0.02, 0.05, 0.10, 0.20\}$ ;  $\eta = 0$  is reselection by argmin at every node. Figure 10 reports, over 100 replications, the mean number of reselections, the fraction of nodes at which the incumbent is the population-optimal class, and the mean lag in nodes before the new optimal class is adopted after the change. Without hysteresis the incumbent changes 5.7 times on average, all but one driven by estimation noise; a margin of 0.10 reduces this to 1.16 with the incumbent correct at 99.2% of nodes and a lag of 0.12 nodes; a margin of 0.20 suppresses reselection almost entirely but delays adoption of the new class by 0.62 nodes on average. The pattern is that of Proposition 4.6. Reselections not driven by a population change vanish once  $\eta$  exceeds twice the estimation error, at the cost of a lag when the population does change.

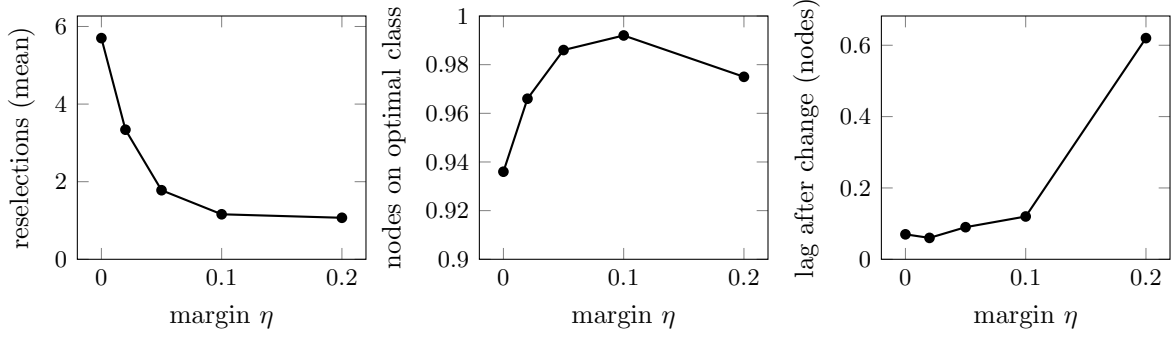


Figure 10: E4. Reselection of the Stage-I class across forty decision nodes with one population change, under the hysteresis rule with margin  $\eta$  (100 replications;  $\eta = 0$  is argmin reselection). Left: mean number of reselections. Centre: fraction of nodes at which the incumbent is the population-optimal class. Right: mean lag before the new optimal class is adopted after the change.

## 4.7 Market data

This subsection applies the procedure to two market panels with the candidate universe declared in advance. The first is twelve U.S. equities, one per sector, and 127 economic and financial series as candidate drivers, daily from January 2001 to July 2023, 5,810 aligned observations after the driver-alignment step. The second is seventeen hedge-fund strategy indices against the same drivers at month end, 267 months. Series that are strictly positive enter as log differences and the rest as first differences. Because the subset lattice over 127 candidates cannot be enumerated, the declared search rule is greedy forward selection: the driver that most reduces the maximal absolute residual correlation is added until the tolerance is met or the declared dimension limit is reached, eight drivers for the equity panel and four for the shorter monthly panel. Windows of 756 sessions stepped by 63 give 81 equity windows and windows of 120 months stepped by six give 25 hedge-fund windows, each split 50/25/25 into certification, estimation and evaluation. Stage II builds  $\widehat{B}\widehat{V}\widehat{B}^\top + \text{diag}(\widehat{\Omega})$  on the estimation block and evaluates its global-minimum-variance portfolio out of sample against the sample covariance, a matched-dimension principal-component estimator with diagonal residual, and Ledoit–Wolf linear shrinkage.

Conditioning removes about half of the cross-asset dependence and does not remove enough. On the equity panel the maximal absolute residual correlation falls from a median of 0.67 before conditioning to 0.33 after selecting eight drivers, and out of sample from 0.67 to 0.46, so the selected representation explains a substantial part of the common variation the sample covariance leaves unmodelled. It clears a per-pair tolerance of 0.20 in only 2 of the 81 windows, and a tolerance calibrated to the maximum over the sixty-six asset pairs, 0.27, in 17% of them; at the stricter reading the admissible class is empty almost throughout, and under the contract the decision is then to enlarge the candidate universe or lower the claim rather than to optimize. On the hedge-fund panel the defect falls from 0.98 to 0.85 against a Šidák-calibrated tolerance of 0.56 and no window is admissible; strategy indices share a common component that four drivers cannot screen off. That panel answers the screening question only: with seventeen series and a thirty-month estimation block the sample covariance is not identified, so no portfolio comparison is reported on it. The selection is economically stable: the volatility index is chosen in every one of the 81 equity windows, followed by equity indices and government yields, and equity-index futures, the Nikkei and a high-yield credit index lead the hedge-fund panel.

Whether the condition can be met at all depends on the asset universe, and in a way the theory

anticipates. Table 5 repeats the exercise on five universes against the same 127 candidates, with the tolerance calibrated to the number of asset pairs by a Šidák correction so that the defect is comparable across universes of different size. On the twelve single names the selected representation removes half of the maximal residual correlation and clears the tolerance in 17% of windows, the first non-empty admissible class in this paper on market data. On the 48 industry portfolios the reduction falls to 17% and on the 25 size and value portfolios to 4%, with no admissible window in either; the European and Developed size and momentum portfolios lie in between. The ordering follows the number of names inside each asset rather than the number of assets, and the design below shows it directly: a sorted portfolio is an average of hundreds of securities whose idiosyncratic components have already cancelled, so what remains is common variation, and a representation cannot screen off a dependence that is the common factor itself. Proposition 3.23 describes this from the other side, and the empirical consequence is that the separator condition is a property of the asset universe as much as of the candidate set: it is attainable on individual securities and, at the tolerances a practitioner would declare, not attainable on pre-diversified portfolios.

Table 5: Statistical admissibility condition across asset universes, 127 candidate drivers, daily 2001–2023, rolling 756-session windows. The tolerance is 0.10 plus a Šidák-calibrated bound for the maximum over the  $n(n-1)/2$  asset pairs. Defects are medians over windows of the maximal absolute residual correlation before conditioning and after greedy forward selection of at most eight drivers.

| universe                     | $n$ | windows | tolerance | before | after | reduction | admissible |
|------------------------------|-----|---------|-----------|--------|-------|-----------|------------|
| 12 US equities               | 12  | 81      | 0.27      | 0.67   | 0.33  | 51%       | 17%        |
| 48 industries                | 48  | 77      | 0.31      | 0.92   | 0.77  | 17%       | 0%         |
| 25 size/value (US)           | 25  | 77      | 0.29      | 0.98   | 0.94  | 4%        | 0%         |
| 25 size/momentum (Europe)    | 25  | 81      | 0.29      | 0.96   | 0.81  | 15%       | 0%         |
| 25 size/momentum (Developed) | 25  | 81      | 0.29      | 0.97   | 0.90  | 8%        | 0%         |

The ordering across universes could reflect anything that distinguishes single names from sorted portfolios. A design that holds everything else fixed isolates the mechanism. From 147 U.S. equities with complete coverage between 2010 and 2024 we form universes of twelve assets in which each asset is an equally weighted basket of  $k$  names, for  $k = 1, 2, 4, 8, 12$ : the same names and the same 127 candidates, aggregated differently, so the only quantity that varies is how much idiosyncratic risk has already been diversified away inside each asset. The defect before conditioning rises monotonically with  $k$ , from 0.68 to 0.88, and the defect attainable after selecting eight drivers rises with it beyond the smallest baskets, from 0.38 at two names per asset to 0.44, 0.55 and 0.63 at twelve, while the fraction of certified windows falls from 8% and 5% to zero (Figure 11). Aggregation is therefore the operative variable, not the identity of the assets: each doubling of the number of names inside an asset removes idiosyncratic variation that the drivers were not needed to explain and leaves common variation that they cannot explain. The single-name design is the one exception to monotonicity after conditioning, and the reason is statistical rather than structural, since the maximum over pairs of an estimated correlation matrix is biased upward and the bias is largest for the noisiest residuals, which are those of individual names.

Whether a larger candidate universe would close the gap has a different answer in each case, and the difference is measurable. Repeating the certification with candidate sets of 4, 8, 16, 32, 64 and 127 drivers drawn at random from the declared universe traces the attainable defect against the size of the search (Figure 12). On the twelve equities the defect falls steadily, from 0.56 with four candidates to 0.33 with all 127, at about 0.07 per doubling of the search; extrapolating the log-

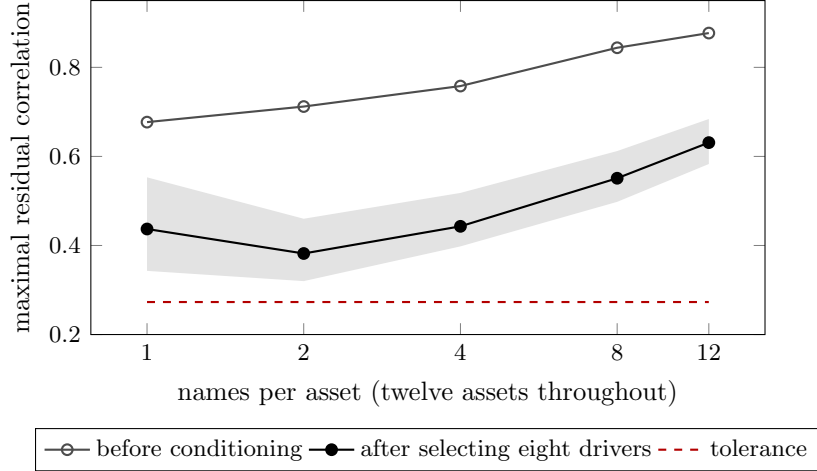


Figure 11: Prior diversification and the attainable defect, with the asset universe held fixed. Twelve assets, each an equally weighted basket of  $k$  names drawn from the same 147 equities, six draws per design, median over windows with the interquartile band of the conditioned defect. Both the dependence to be removed and the dependence that survives conditioning increase with the number of names inside each asset.

linear trend, of the order of three hundred candidates of the same kind would reach the tolerance. On the size and value portfolios the same exercise moves the defect from 0.97 to 0.93, about 0.010 per doubling, and the same extrapolation puts the required search beyond any conceivable universe. The failure of the separator condition is therefore of two different kinds. For individual securities it is a shortage of candidates, and the framework says how many more would be needed. For pre-diversified portfolios it is structural: the residual dependence is the common factor that the sorting has already isolated, and no set of observable drivers of this type removes it.

The portfolio consequences are reported under the protocol of the covariance-estimation literature [23, 40]: out-of-sample volatility, Sharpe ratio, turnover per rebalance and gross leverage of the global minimum variance portfolio, the same under a no-short-sale constraint [36], and paired differences against the conditional estimator tested by a circular block bootstrap. Table 6 gives the equity panel. The conditional covariance is not distinguishable from linear shrinkage on volatility, 14.28% against 14.20% at the median although lower in 69% of windows, and it is lower than the sample covariance, the matched principal-component estimator, nonlinear shrinkage and equal weights; the differences in realized variance are significant at the one per cent level and none of the differences in Sharpe ratio is significant at any conventional level. Where it separates from every alternative is stability: turnover of 0.36 per rebalance against 0.44 for linear shrinkage, 0.54 for the sample covariance and 1.02 for the principal-component estimator, and gross leverage of 1.20 against 1.31, 1.42 and 1.79. A representation of eight drivers selected once per rebalance moves the portfolio less than any estimator that re-reads the whole covariance from returns, which is what a mandate with transaction costs cares about. Two results deserve their own reading. Equal weights attain the highest Sharpe ratio, the finding of [23], but the comparison is between objectives rather than between estimators: minimum-variance portfolios do not use expected returns, and decomposing the difference shows that the advantage of equal weights is an annualised mean higher by 4.8 percentage points, which is not significant at the five per cent level, against a volatility higher by 3.0 points, which is significant at  $10^{-14}$ . On the objective the estimators optimize, the conditional covariance wins; on an objective none of them targets, weighting equally wins, and no covariance

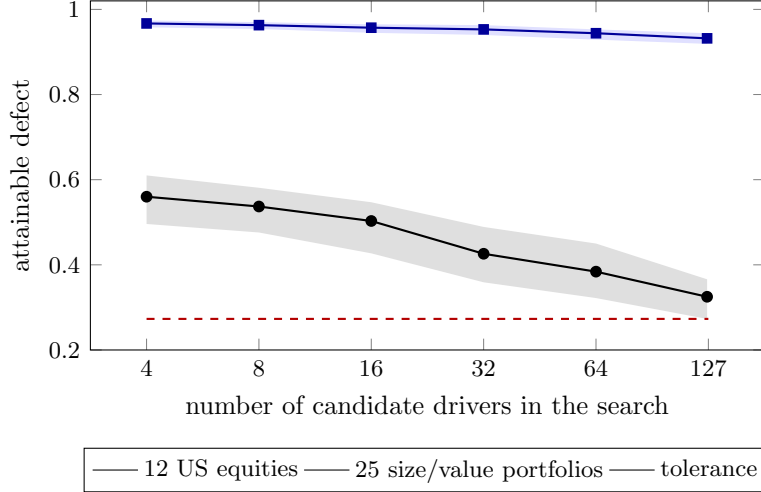


Figure 12: Attainable defect against the size of the candidate search, median over windows and over eight random draws of the candidate set at each size, with the interquartile band. On individual securities the defect declines steadily with the size of the search and the extrapolated trend reaches the tolerance at a few hundred candidates; on pre-diversified portfolios it is almost flat, and no candidate universe of a realistic size would certify them.

estimator can be held responsible for that. Under a no-short-sale constraint the differences narrow sharply, from a range of 1.6 percentage points of volatility to 0.3, which is the sense in which constraints substitute for covariance modelling [36], but they do not vanish and the ordering is preserved. What the constraint does not do is repair the misallocation: the discarded off-diagonal term is 14.7% of predicted variance for the unconstrained portfolio and 14.8% for the constrained one, statistically indistinguishable, because the constraint raises the breadth of the portfolio in (16) at the same time as it lowers the residual share. A practitioner who imposes long-only positions and believes the specification error has been contained is mistaken: the constraint improves the allocation and leaves the reported risk as wrong as it was.

Table 6: Out-of-sample performance of the global minimum variance portfolio on the twelve-equity panel, 81 rolling windows, median across windows. Volatility annualised in per cent, turnover per rebalance, gross leverage  $\|w\|_1$ , and volatility under a no-short-sale constraint. The last two columns give circular block bootstrap  $p$ -values for the paired difference against the conditional estimator in realized variance and in Sharpe ratio.

| estimator                      | vol.  | Sharpe | turnover | leverage | vol. long-only | $p$ var. | $p$ Sharpe |
|--------------------------------|-------|--------|----------|----------|----------------|----------|------------|
| conditional (admissible class) | 14.28 | 0.53   | 0.36     | 1.20     | 14.11          | —        | —          |
| sample                         | 14.55 | 0.59   | 0.54     | 1.42     | 14.33          | 0.009    | 0.640      |
| PCA, matched dimension         | 15.77 | 0.39   | 1.02     | 1.79     | 14.34          | 0.003    | 0.459      |
| Ledoit–Wolf linear             | 14.20 | 0.49   | 0.44     | 1.31     | 14.01          | 0.010    | 0.157      |
| Ledoit–Wolf nonlinear          | 15.07 | 0.53   | 0.66     | 1.51     | 14.34          | 0.005    | 0.376      |
| equal weights                  | 16.55 | 0.74   | 0.00     | 1.00     | 16.55          | 0.004    | 0.283      |

But the risk it predicts is understated, in this market by the mechanism of (16): the residual correlations the representation fails to remove are predominantly positive, with a portfolio-weighted average of 0.05, so  $\bar{c}_w > 0$  and the discarded term is positive for any portfolio whose weighted

exposures are mostly of one sign. The proposition permits either sign; the sign observed here is a property of equity residuals, not a theorem. the ratio of realized to predicted variance has median 1.35 on the equity panel against 1.19 for linear shrinkage. Isolating the restriction at fixed weights makes its size explicit. On the equity panel the discarded term is 14.7% of predicted variance at the median and positive in 80 of the 81 windows, and the Barra bias statistic of the held portfolio, the ratio of realized to predicted volatility whose 95% band is  $1 \pm 0.10$  here, moves from 1.07 with the full residual, inside the band, to 1.16 with the diagonal restriction, outside it. The effect is not an artefact of the optimizer: equal weights give 19.3% and 16,200 random long-only portfolios give 17.6%, positive in every case. Its size tracks the weighted average residual correlation of (16) (rank correlation 0.93) and not the certification defect (rank correlation 0.03), which is what the decomposition predicts and which separates the two questions the framework asks: whether the representation separates, answered by the maximum, and what its failure costs a given portfolio, answered by the quadratic form. Figure 13 shows the two objects. The certification defect stays above the tolerance in almost every window of both panels, so the tolerance path never reaches an admissible class, and the realized-to-predicted ratio of the conditional estimator sits above one, as the diagonal-residual restriction on a non-separating representation requires.

The experiment separates two statements that portfolio practice usually merges. A representation can be good enough to improve an allocation and not good enough to certify the diagonal-residual restriction its covariance imposes, and the framework distinguishes the two before any portfolio is formed: the diagnostic reports the failure, the portfolio is still built if the mandate requires it, and the risk it reports is known in advance to be understated. Closing the remaining dependence requires either a larger candidate universe than the 127 series declared here or an asset universe with fewer common components than twelve sector representatives or seventeen strategy indices, and that is the empirical question the framework poses.

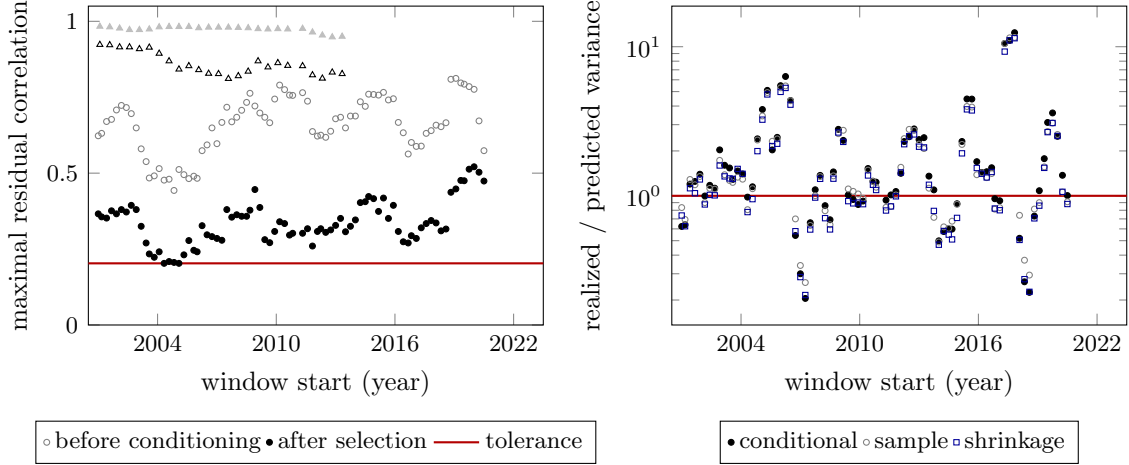


Figure 13: Field experiment on the declared panels, twelve U.S. equities with 81 daily windows (circles and dots) and seventeen hedge-fund strategy indices with 25 monthly windows (triangles), against 127 candidate drivers. Left: maximal absolute residual correlation across assets before conditioning and after greedy forward selection, against the declared tolerance; selection removes about half of the dependence in the equity panel and less in the hedge-fund panel, and in almost every window the result remains above the tolerance, so the admissible class is empty. Right: realized over predicted variance of the global-minimum-variance portfolio on the equity panel for the conditional covariance with diagonal residual, the sample covariance and Ledoit–Wolf shrinkage; the conditional estimator understates realized risk, by the decomposition (16) with the positive average residual correlation this market exhibits.

## 5 Concluding remarks

Portfolio optimization is usually posed after the conditioning information has been chosen, and two errors about that information, using what was not available and treating common variation as idiosyncratic, are then carried into the portfolio without detection. This paper makes the information a decision variable under hard admissibility constraints declared in advance, selects an optimal information class by an order in which no risk–return quantity enters, and solves the classical problem inside it. Classical portfolio theory is preserved as the inner problem; the consequences live in the outer one. The value of information is monotone only under an embedding that portfolio constraints can break, and selection by decision loss therefore admits look-ahead whenever it is offered; the two-stage solution and the joint envelope are the minimal elements of two orders on one set of admissible pairs; under exact separation the diversifiable part of risk is fixed by the admissible class and is stable under interventions only under interventional admissibility; requiring causal identification has an explicit price, traded against search complexity; and the estimator is consistent with second-order regret, with information selection separating across decision nodes.

The scope of what has been established is equally definite. Among the admissibility conditions only the mandate scope is a declaration; the others are certified from provenance, from the enlarged filtration or from returns. The distance bound on the causal price carries an explicit metric in the quadratic case, and the general case identifies the metric the modeller must supply. Portfolio selection with intertemporal costs inside the selected classes is located by the separability result as the dynamic part of the problem and is the natural next step. In the field experiments the condition is met on individual securities and not on pre-diversified portfolios, an ordering the diversification

result predicts and that enlarging the candidate search does not overturn, and where it fails the selected representation still improves the allocation while its reported risk is understated, which separates the quality of a representation for optimization from its adequacy for the restriction its covariance imposes. The question the paper poses is whether optimizing over a certified information–portfolio domain produces decisions more appropriate to the declared claim than optimizing weights after an arbitrary representation has been fixed. The controlled experiments answer affirmatively, and the field experiments show the certification reporting, before any portfolio is formed, a failure whose consequence for reported risk is then observed.

## References

- [1] Aliprantis, C.D. and Border, K.C. (2006). *Infinite Dimensional Analysis: A Hitchhiker’s Guide*, 3rd ed. Springer.
- [2] Acerbi, C. (2002). Spectral measures of risk: a coherent representation of subjective risk aversion. *Journal of Banking and Finance*, 26(7), 1505–1518.
- [3] Anis, H.T. and Kwon, R.H. (2025). End-to-end, decision-based, cardinality-constrained portfolio optimization. *European Journal of Operational Research*, 320(3), 739–753.
- [4] Artzner, P., Delbaen, F., Eber, J.-M. and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, 9(3), 203–228.
- [5] Amendinger, J., Imkeller, P. and Schweizer, M. (1998). Additional logarithmic utility of an insider. *Stochastic Processes and their Applications*, 75(2), 263–286.
- [6] Bareinboim, E., Correa, J.D., Ibeling, D. and Icard, T. (2022). On Pearl’s hierarchy and the foundations of causal inference. In *Probabilistic and Causal Inference: The Works of Judea Pearl*, ACM Books, 507–556.
- [7] Bartlett, P.L. (2008). Fast rates for estimation error and oracle inequalities for model selection. *Econometric Theory*, 24(2), 545–552.
- [8] Best, M.J. and Grauer, R.R. (1991). On the sensitivity of mean-variance-efficient portfolios to changes in asset means: some analytical and computational results. *Review of Financial Studies*, 4(2), 315–342.
- [9] Bermin, H.-P. and Holm, M. (2024). The geometry of risk adjustments. *Decisions in Economics and Finance*, 47, 83–120.
- [10] Bjerring, T.T., Ross, O. and Weissensteiner, A. (2017). Feature selection for portfolio optimization. *Annals of Operations Research*, 256, 21–40.
- [11] Blackwell, D. (1951). Comparison of experiments. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, 93–102. University of California Press.
- [12] Blackwell, D. (1953). Equivalent comparisons of experiments. *Annals of Mathematical Statistics*, 24(2), 265–272.
- [13] Boucheron, S., Lugosi, G. and Massart, P. (2013). *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press.
- [14] Bühlmann, P. (2020). Invariance, causality and robustness. *Statistical Science*, 35(3), 404–426.
- [15] Butler, A. and Kwon, R.H. (2023). Integrating prediction in mean-variance portfolio optimization. *Quantitative Finance*, 23(3), 429–452.
- [16] Boyd, S. and Vandenberghe, L. (2004). *Convex Optimization*. Cambridge University Press.
- [17] Catoni, O. (2012). Challenging the empirical mean and empirical variance: a deviation study. *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques*, 48(4), 1148–1185.
- [18] Chamberlain, G. and Rothschild, M. (1983). Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica*, 51(5), 1281–1304.

- [19] Chernozhukov, V., Chetverikov, D. and Kato, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Annals of Statistics*, 41(6), 2786–2819.
- [20] Chopra, V.K. and Ziemba, W.T. (1993). The effect of errors in means, variances, and covariances on optimal portfolio choice. *Journal of Portfolio Management*, 19(2), 6–11.
- [21] Cont, R. (2006). Model uncertainty and its impact on the pricing of derivative instruments. *Mathematical Finance*, 16(3), 519–547.
- [22] Costa, G. and Iyengar, G.N. (2023). Distributionally robust end-to-end portfolio construction. *Quantitative Finance*, 23(10), 1465–1482.
- [23] DeMiguel, V., Garlappi, L. and Uppal, R. (2009). Optimal versus naive diversification: how inefficient is the  $1/N$  portfolio strategy? *Review of Financial Studies*, 22(5), 1915–1953.
- [24] Delbaen, F. and Schachermayer, W. (1994). A general version of the fundamental theorem of asset pricing. *Mathematische Annalen*, 300(1), 463–520.
- [25] Elmachoub, A.N. and Grigas, P. (2022). Smart “predict, then optimize”. *Management Science*, 68(1), 9–26.
- [26] Elmachoub, A.N., Lam, H., Zhang, H. and Zhao, Y. (2023). Estimate-then-optimize versus integrated-estimation-optimization versus sample average approximation: a stochastic dominance perspective. arXiv:2304.06833.
- [27] Ferson, W.E. and Siegel, A.F. (2009). Testing portfolio efficiency with conditioning information. *Review of Financial Studies*, 22(7), 2735–2758.
- [28] Fishburn, P.C. (1974). Lexicographic orders, utilities and decision rules: a survey. *Management Science*, 20(11), 1442–1471.
- [29] Hansen, L.P. and Richard, S.F. (1987). The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models. *Econometrica*, 55(3), 587–613.
- [30] Ferson, W.E. and Siegel, A.F. (2001). The efficient use of conditioning information in portfolios. *Journal of Finance*, 56(3), 967–982.
- [31] Harvey, C.R. and Liu, Y. (2015). Backtesting. *Journal of Portfolio Management*, 42(1), 13–28.
- [32] Heinze-Deml, C., Peters, J. and Meinshausen, N. (2018). Invariant causal prediction for nonlinear models. *Journal of Causal Inference*, 6(2), 20170016.
- [33] Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301), 13–30.
- [34] Howard, C., Lohre, H. and Mudde, S. (2025). Causal network representations in factor investing. *Intelligent Systems in Accounting, Finance and Management*, 32(1), e70001.
- [35] Jacod, J. (1985). Grossissement initial, hypothèse (H’) et théorème de Girsanov. In *Grossissements de filtrations: exemples et applications*, Lecture Notes in Mathematics 1118, 15–35. Springer.
- [36] Jagannathan, R. and Ma, T. (2003). Risk reduction in large portfolios: why imposing the wrong constraints helps. *Journal of Finance*, 58(4), 1651–1683.
- [37] Jorion, P. (1986). Bayes–Stein estimation for portfolio analysis. *Journal of Financial and Quantitative Analysis*, 21(3), 279–292.
- [38] Kandel, S. and Stambaugh, R.F. (1996). On the predictability of stock returns: an asset-allocation perspective. *Journal of Finance*, 51(2), 385–424.
- [39] Ledoit, O. and Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88(2), 365–411.

- [40] Ledoit, O. and Wolf, M. (2017). Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets Goldilocks. *Review of Financial Studies*, 30(12), 4349–4388.
- [41] Lee, J., Jeon, H., Bae, H. and Lee, Y. (2025). Return prediction for mean-variance portfolio selection: how decision-focused learning shapes forecasting models. *Proceedings of the 6th ACM International Conference on AI in Finance*, 114–122.
- [42] Lim, A.E.B., Shanthikumar, J.G. and Vahn, G.-Y. (2011). Conditional value-at-risk in portfolio optimization: coherent but fragile. *Operations Research Letters*, 39(3), 163–171.
- [43] Mandi, J., Bucarey, V., Mulamba, M. and Guns, T. (2022). Decision-focused learning: through the lens of learning to rank. *Proceedings of ICML*, PMLR 162, 14935–14947.
- [44] Mandi, J., Kotary, J., Berden, S., Mulamba, M., Bucarey, V., Guns, T. and Fioretto, F. (2024). Decision-focused learning: foundations, state of the art, benchmark and future opportunities. *Journal of Artificial Intelligence Research*, 80, 1623–1701.
- [45] Markowitz, H. (1952). Portfolio selection. *Journal of Finance*, 7(1), 77–91.
- [46] Massart, P. (2007). *Concentration Inequalities and Model Selection*. Lecture Notes in Mathematics 1896, Springer.
- [47] Merton, R.C. (1972). An analytic derivation of the efficient portfolio frontier. *Journal of Financial and Quantitative Analysis*, 7(4), 1851–1872.
- [48] Michaud, R.O. (1989). The Markowitz optimization enigma: is ‘optimized’ optimal? *Financial Analysts Journal*, 45(1), 31–42.
- [49] Miller, N. and Ruszczyński, A. (2008). Risk-adjusted probability measures in portfolio optimization with coherent measures of risk. *European Journal of Operational Research*, 191(1), 193–206.
- [50] Pearl, J. (2009). *Causality: Models, Reasoning, and Inference*, 2nd ed. Cambridge University Press.
- [51] Peters, J., Bühlmann, P. and Meinshausen, N. (2016). Causal inference by using invariant prediction: identification and confidence intervals. *Journal of the Royal Statistical Society, Series B*, 78(5), 947–1012.
- [52] Peters, J., Janzing, D. and Schölkopf, B. (2017). *Elements of Causal Inference: Foundations and Learning Algorithms*. MIT Press.
- [53] Reichenbach, H. (1956). *The Direction of Time*. University of California Press.
- [54] Rodrigues, A.F.P., Cavalcante, C.C. and Crisóstomo, V.L. (2019). A projection pricing model for non-Gaussian financial returns. *Physica A*, 534, 122181.
- [55] Rodríguez Domínguez, A. (2023). Portfolio optimization based on neural networks sensitivities from assets dynamics respect common drivers. *Machine Learning with Applications*, 11, 100447.
- [56] Rockafellar, R.T. (1970). *Convex Analysis*. Princeton University Press.
- [57] Rockafellar, R.T. and Uryasev, S. (2000). Optimization of conditional value-at-risk. *Journal of Risk*, 2(3), 21–41.
- [58] Ross, S.A. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, 13(3), 341–360.
- [59] Torgersen, E. (1991). *Comparison of Statistical Experiments*. Cambridge University Press.
- [60] Vapnik, V.N. (1998). *Statistical Learning Theory*. Wiley.
- [61] Yu, B. (1994). Rates of convergence for empirical processes of stationary mixing sequences. *Annals of Probability*, 22(1), 94–116.
- [62] Yang, J., Liang, E., Su, Z., Zou, Z., Zhen, P., Guo, J., Ma, W. and An, K. (2025). DFF: decision-focused fine-tuning for smarter predict-then-optimize with limited data. *Proceedings of AAAI*, 39(25), 26868–26876.

## A Proofs

*Proof of Proposition 3.3.* Since  $H^- \subseteq H^+$ , the law of total covariance gives (2):  $\text{Cov}(R \mid H^-) = \mathbb{E}[\text{Cov}(R \mid H^+) \mid H^-] + \text{Cov}(\mathbb{E}[R \mid H^+] \mid H^-)$ , and  $\text{Cov}(\mathbb{E}[R \mid H^+] \mid H^-) = \text{Cov}(\xi \mid H^-)$  because  $\mathbb{E}[R \mid H^-]$  is  $H^-$ -measurable. Under separation,  $\text{Cov}(R_i, R_j \mid H^+) = 0$  for  $i \neq j$  by conditional independence, so  $\text{Cov}(R \mid H^+)$  is diagonal almost surely and its  $H^-$ -conditional expectation is diagonal. In the affine case  $\xi = B(\zeta - \mathbb{E}[\zeta \mid H^-])$ , whence  $\text{Cov}(\xi \mid H^-) = BVB^\top$ .  $\square$

*Proof of Proposition 3.11.* (a) If  $\phi$  is an invertible measurable recoding,  $\sigma(\phi(H)) = \sigma(H)$ , so  $\phi(H)$  and  $H$  lie in the same information-equivalence class and, by hypothesis, have the same admissibility-condition values and the same four quantities; the order of Definition 3.10 and its set of minimal elements are therefore unchanged, and  $\mathcal{A}_t^*$  contains every element of a class as soon as it contains one. (b) By definition of a least exact separator,  $\varepsilon_t(H_t^*) = 0 \leq \varepsilon_t(H)$  for all  $H \in \mathcal{A}_t$ , and no  $H$  with  $\sigma(H) \subsetneq \sigma(H_t^*)$  satisfies the exact requirement, so no  $H \in \mathcal{A}_t$  has  $\varepsilon_t(H) < \varepsilon_t(H_t^*)$ ; under a conditional-law contract  $\delta_t \equiv 0$ , and under an interventional contract the same argument applies to  $\delta_t$  inside the identified class, so any  $H \prec_{\mathcal{A}} H_t^*$  would have to be strictly smaller in  $(\kappa_t, \iota_t)$  while tying on the scientific coordinates, which is excluded by the minimality of  $H_t^*$  in the information order and the finiteness of  $\iota_t(H_t^*)$ ; hence  $H_t^* \in \mathcal{A}_t^*$ , and  $[H_t^*] \subseteq \mathcal{A}_t^*$  by (a). (c) The four quantities are functions of  $H$  alone; a selection rule depending on  $m$  or on  $w$  is a different order and hence a different contract.  $\square$

*Proof of Theorem 3.12. Step 1 (Stage I).* Write  $a^1, a^2, a^3, a^4$  for the four quantities  $\varepsilon_t, \delta_t, \kappa_t, \iota_t$  of Definition 3.10, in that order. Set  $X^0 := X_t$  and, recursively,  $X^k := \arg \min_{x \in X^{k-1}} a^k(x)$  for  $k = 1, \dots, 4$ . If  $X^{k-1}$  is nonempty and compact, then  $a^k$ , being lower semicontinuous, attains its minimum on  $X^{k-1}$  [1, Thm. 2.43], and  $X^k = X^{k-1} \cap \{a^k \leq \min_{X^{k-1}} a^k\}$  is the intersection of a compact set with a closed set (lower level sets of lsc functions are closed), hence nonempty and compact. By induction  $X^4$  is nonempty and compact. It remains to identify  $X^4$  with the set of  $<_{\text{lex}}$ -minimal elements:  $x \in X^4$  means that  $x$  minimizes  $a^1$ , then  $a^2$  among  $a^1$ -minimizers, and so on, which is precisely the statement that no  $x'$  satisfies  $x' \prec_{\mathcal{A}} x$ ; conversely, if  $x \notin X^4$ , let  $k$  be the first index with  $x \notin X^k$ ; then  $x \in X^{k-1}$  and any  $x' \in X^k$  ties  $x$  on  $a^1, \dots, a^{k-1}$  and is strictly smaller on  $a^k$ , so  $x' \prec_{\mathcal{A}} x$  and  $x$  is not minimal. Hence  $\mathcal{A}_t^* = X^4$  (as a union of classes). *Step 2 (Stage II).* Fix  $H \in \mathcal{A}_t^*$  and a feasible  $m$ . The set  $F_H(m) := \{w \in \mathcal{W}(H) : \mu_H(w) \geq m\}$  is the preimage of the closed set  $[m, \infty)$  under the continuous  $\mu_H$ , intersected with the compact  $\mathcal{W}(H)$ ; it is therefore compact and, by assumption, nonempty. The continuous  $\rho_H$  attains its minimum on  $F_H(m)$ , and  $\mathcal{E}_H(m) = F_H(m) \cap \rho_H^{-1}(\{\min_{F_H(m)} \rho_H\})$  is closed in a compact set, hence compact.  $\square$

*Proof of Theorem 3.13.* For the conditional mean–variance loss  $\frac{\gamma}{2} \text{Var}(w^\top R \mid H) - \mathbb{E}[w^\top R \mid H]$  and a  $\sigma(H_1^-)$ -measurable policy  $\pi$ , the law of total variance gives  $\mathbb{E}[\text{Var}(\pi^\top R \mid H_2)] \leq \mathbb{E}[\text{Var}(\pi^\top R \mid H_1)]$  while the conditional means have the same expectation, so the identity map satisfies the embedding inequality; this is the case used throughout. *Step 1.* For every  $\pi \in \Pi(H_1)$  we have  $\iota\pi \in \Pi(H_2)$  and, by the embedding inequality and monotonicity of expectation,

$$J(\iota\pi; H_2) = \mathbb{E}[\ell(\iota\pi, R; H_2)] \leq \mathbb{E}[\ell(\pi, R; H_1)] = J(\pi; H_1).$$

Hence

$$\inf_{\pi \in \Pi(H_1)} J(\pi; H_1) \geq \inf_{\pi \in \Pi(H_1)} J(\iota\pi; H_2) \geq \inf_{\pi' \in \Pi(H_2)} J(\pi'; H_2),$$

which is  $V(H_1) \geq V(H_2)$ . *Step 2.* Under the additional hypothesis,  $\ell(\iota\pi_1, R; H_2) - \ell(\pi_2, R; H_2) \geq 0$  almost surely and is strictly positive on a set of positive probability, so its expectation is strictly positive:  $V(H_1) = J(\pi_1; H_1) \geq J(\iota\pi_1; H_2) > J(\pi_2; H_2) \geq V(H_2)$ .  $\square$

*Proof of Corollary 3.14.* If  $H_2 \notin \mathcal{A}_t$  then  $V(H_2)$  does not enter (7). If the embedding fails, Step 1 of the proof of Theorem 3.13 has no content, and  $V(H_2) > V(H_1)$  occurs as soon as  $\mathcal{W}(H_2)$  excludes  $w_{H_1}^*$  while the conditional law is unchanged, as in Example 3.17 with a representation-dependent trading restriction.  $\square$

*Proof of Proposition 3.15.* Since  $\sigma(H^-) \subseteq \sigma(H^-, X^+)$  and the feasible sets coincide, the identity map embeds the  $H$  problem in the  $H^\sharp$  problem in the sense of Definition 3.9 (for the conditional mean–variance loss by the law of total variance), and Theorem 3.13 gives  $V(H^\sharp) \leq V(H)$ . For strictness, let  $\pi$  be the  $H^-$ -measurable optimizer and  $\pi^\sharp$  the  $H^\sharp$ -measurable optimizer. Conditional on  $H^\sharp$  the quadratic objective is strictly convex in  $w$  with unconstrained minimizer  $\gamma^{-1} \text{Cov}(R | H^\sharp)^{-1} \mathbb{E}[R | H^\sharp]$ ; on the event where  $\mathbb{E}[R | H^\sharp] \neq \mathbb{E}[R | H^-]$  this minimizer differs from  $\pi$  with positive probability, and since  $\mathcal{W}$  contains a neighbourhood of the optimizer,  $\pi^\sharp$  attains a strictly smaller conditional objective there. Taking expectations gives  $J(\pi^\sharp; H^\sharp) < J(\pi; H)$ , hence  $V(H^\sharp) < V(H)$ . The final statement follows because the minimizer of  $V$  over the candidate set cannot be a candidate strictly dominated by another candidate in the set.  $\square$

*Proof of Theorem 3.18. Step 1 (the graph is compact).* A compact-valued uhc correspondence into a Hausdorff space has closed graph [1, Thm. 17.10], and the image  $\mathcal{W}(X) = \bigcup_{x \in X} \mathcal{W}(x)$  of the compact set  $X$  under a compact-valued uhc correspondence is compact [1, Lemma 17.8]. Hence  $\Gamma$  is a closed subset of the compact set  $X \times \mathcal{W}(X)$  and is therefore compact. *Step 2 (attainment).* A lower semicontinuous function with values in  $\mathbb{R} \cup \{+\infty\}$  attains its infimum on a nonempty compact set [1, Thm. 2.43]; the infimum is finite because  $J \not\equiv +\infty$ . Any point  $(x^*, w^*)$  at which it is attained, together with any representative  $H^* \in x^*$ , is a minimizer of (7) by the quotient invariance of  $V$ .  $\square$

*Proof of Proposition 3.20.* Let  $(H, w) \in \Gamma_t(m)$ . If  $H \notin \mathcal{A}_t^*$ , there is  $H' \in \mathcal{A}_t$  with  $H' \prec_{\mathcal{A}} H$ ; for every  $w' \in \mathcal{W}(H')$  with  $\mu_{H'}(w') \geq m$ , which exists since  $m$  is feasible,  $(H', w') \prec_{\text{APO}} (H, w)$ , so  $(H, w)$  is not minimal. If  $H \in \mathcal{A}_t^*$  and  $w \notin \mathcal{E}_H(m)$ , there is  $w'$  feasible for  $H$  with  $\rho_H(w') < \rho_H(w)$ , and  $(H, w') \prec_{\text{APO}} (H, w)$ . Conversely let  $H \in \mathcal{A}_t^*$ ,  $w \in \mathcal{E}_H(m)$ , and suppose  $(H', w') \prec_{\text{APO}} (H, w)$ . The first alternative in (8) is excluded by minimality of  $H$ ; the second requires  $H' = H$  and  $\rho_H(w') < \rho_H(w)$ , which contradicts  $w \in \mathcal{E}_H(m)$ . Hence  $(H, w)$  is minimal, which proves (9). The statements for  $\preceq_{\text{env}}$  are the definitions of (11) and (7) read on  $\Gamma_t(m)$ .  $\square$

*Proof of Proposition 3.21.* The feasible set  $F(m) = \mathcal{W}(H_t^*) \cap \{w : \mu_H^\top w \geq m\}$  is the intersection of convex sets, hence convex, and closed. The objective is strictly convex since  $\Sigma_H \succ 0$ . If  $w_1 \neq w_2$  were both minimizers, then  $\frac{1}{2}(w_1 + w_2) \in F(m)$  by convexity and  $\rho_H(\frac{1}{2}(w_1 + w_2)) < \frac{1}{2}\rho_H(w_1) + \frac{1}{2}\rho_H(w_2) = \min_{F(m)} \rho_H$  by strict convexity, a contradiction; existence follows from Theorem 3.12 when  $\mathcal{W}(H_t^*)$  is bounded and from coercivity of  $\rho_H$  otherwise. Monotonicity holds because  $F(m') \subseteq F(m)$  for  $m' \geq m$ . Convexity of the value function: for feasible  $m_0, m_1$  and  $\theta \in [0, 1]$ , let  $w_i \in \mathcal{E}_H(m_i)$ ; then  $\mu_H^\top(\theta w_1 + (1-\theta)w_0) \geq \theta m_1 + (1-\theta)m_0$  and  $\theta w_1 + (1-\theta)w_0 \in \mathcal{W}(H_t^*)$ , so  $\rho_t^*(\theta m_1 + (1-\theta)m_0) \leq \rho_H(\theta w_1 + (1-\theta)w_0) \leq \theta \rho_t^*(m_1) + (1-\theta)\rho_t^*(m_0)$  [56, Thm. 29.1].  $\square$

*Proof of Proposition 3.22.* (a) is the definition of the supremum and of the Stage-I class. (b) For the quadratic loss  $J_\gamma(w; H) = \frac{\gamma}{2} w^\top \Sigma_H w - (\mu_H - r_f \mathbf{1})^\top w$  over  $\mathbb{R}^n$  the minimizer is  $w = \gamma^{-1} \Sigma_H^{-1} (\mu_H - r_f \mathbf{1})$  and the conditional minimum value is  $-S_t(H)^2/(2\gamma)$ , so  $V_\gamma(H) = -\bar{S}_t(H)^2/(2\gamma)$ . Theorem 3.13 applied to  $J_\gamma$  gives  $V_\gamma(H_2) \leq V_\gamma(H_1)$  for every  $\gamma > 0$ , hence  $\bar{S}_t(H_2)^2 \geq \bar{S}_t(H_1)^2$ ; the normalization  $\mathbf{1}^\top w = 1$  does not change the maximal slope because the Sharpe ratio is scale-invariant.  $\square$

*Proof of Proposition 3.23.* (a) By Proposition 3.3,  $\text{Cov}(R | H^-) = \mathbb{E}[\text{Cov}(R | H^+) | H^-] + \text{Cov}(\mathbb{E}[R | H^+] | H^-)$ . Under exact separation the components of  $R$  are independent given  $H^+$ , so  $\text{Cov}(R | H^+)$  is diagonal, and the first term, being an  $H^-$ -conditional expectation of

a diagonal random matrix, is diagonal. The second term is the covariance of the response of  $\mathbb{E}[R \mid H^+]$  to the innovation  $\xi_H$  of  $H^+$  relative to  $H^-$ , i.e.  $B_H V_H B_H^\top$  in (13); comparing with the decomposition (13)(a) identifies  $\Omega_H$  with the first term, which is diagonal. For diagonal  $\Omega_H$ ,  $w^\top \Omega_H w = \sum_i w_i^2 \Omega_{H,ii} \leq \max_i \Omega_{H,ii} \sum_i w_i^2$ , which is (15); for  $w = \mathbf{1}/n$ ,  $\|w\|_2^2 = 1/n$ . The identity  $w^\top B_H V_H B_H^\top w = (B_H^\top w)^\top V_H (B_H^\top w)$  is immediate, and if  $w, w'$  satisfy  $B_H^\top w = B_H^\top w'$  the systematic components coincide whatever the spread of weights. (b) Write  $\Omega_H = \text{diag}(\Omega_H) + \Omega_H^{\text{off}}$  and expand the quadratic form; for  $w = \mathbf{1}/n$  the off-diagonal term is  $n^{-2} \sum_{i \neq j} \Omega_{H,ij}$ , which equals the average off-diagonal entry times  $(n-1)/n$  and does not tend to zero unless that average does. (c) By Definition 3.6,  $\mathcal{L}^e(R \mid H) = \mathcal{L}^o(R \mid H)$ , which implies invariance of  $\mu_H$ ,  $\Sigma_H$ , of the decomposition (13) and of its diagonal residual, hence of both components. The final sentence is the observation that a conditional-law-admissible  $H$  that explains covariation through a channel severed by the intervention no longer satisfies the hypotheses of (a) after it, so (b) applies to the post-intervention residual. (d) Write  $\Omega = \text{diag}(s)C \text{diag}(s)$  with  $C$  the residual correlation matrix and  $s_i = \Omega_{ii}^{1/2}$ , and  $v = w \odot s$ . Then  $w^\top \Omega^{\text{off}} w = \sum_{i \neq j} v_i v_j c_{ij}$ , which equals  $D \bar{c}_w$  by the definition of  $\bar{c}_w$ , and  $w^\top \text{diag}(\Omega) w = \|v\|_2^2$ . Dividing by  $w^\top \Sigma_H w$  and writing  $D = (\sum_i v_i)^2 - \|v\|_2^2$  as  $\|v\|_2^2$  times the breadth minus one gives (16). If  $v \geq 0$  then  $D = \sum_{i \neq j} v_i v_j \geq 0$  and  $|\bar{c}_w| \leq \max_{i \neq j} |c_{ij}| = \varepsilon_t(H)$  because  $\bar{c}_w$  is then a convex combination of the  $c_{ij}$ .  $\square$

*Proof of Theorem 3.24.* The lower bound is the inclusion  $\mathcal{A}_t(\text{i}) \subseteq \mathcal{A}_t(\text{c})$ . For the upper bound fix  $G \in \mathcal{A}_t(\text{i})$ . By definition of  $V_i^*$  as an infimum over a set containing  $G$ ,  $V_i^* \leq V(G)$ ; by the Lipschitz property,  $V(G) \leq V(H_0^*) + L d(G, H_0^*)$ . Hence  $\Delta_{\text{ci}} = V_i^* - V(H_0^*) \leq L d(G, H_0^*)$  for every  $G \in \mathcal{A}_t(\text{i})$ , and taking the infimum over  $G$  gives (19). For the quadratic loss,  $|J(w; H) - J(w; G)| \leq \frac{\gamma}{2} \bar{r}^2 \|\Sigma_H - \Sigma_G\|_{\text{op}} + \bar{r} \|\mu_H - \mu_G\|_2 = \bar{r} d(H, G)$  for every  $w \in \mathcal{W}$ , and infima of two functions whose pointwise difference is bounded by a constant differ by at most that constant; the random-state case follows by taking expectations.  $\square$

*Proof of Proposition 3.25.* Write  $r := J(\hat{w}) - J(w^*) \geq 0$  and  $D := J - \hat{J}$ . Since  $\hat{J}(\hat{w}) \leq \hat{J}(w^*)$ ,  $r \leq D(\hat{w}) - D(w^*) = \frac{\gamma}{2} (\hat{w}^\top \Delta_\Sigma \hat{w} - w^{*\top} \Delta_\Sigma w^*) - \Delta_\mu^\top (\hat{w} - w^*)$  with  $\Delta_\Sigma = \Sigma - \hat{\Sigma}$ ,  $\Delta_\mu = \mu - \hat{\mu}$ . The quadratic difference equals  $\frac{\gamma}{2} (\hat{w} + w^*)^\top \Delta_\Sigma (\hat{w} - w^*)$ , so  $r \leq (\gamma \bar{r} \|\Delta_\Sigma\|_{\text{op}} + \|\Delta_\mu\|_2) \|\hat{w} - w^*\|_2 =: \epsilon \|\hat{w} - w^*\|_2$ . Since  $J$  is  $\gamma\lambda$ -strongly convex and  $w^*$  minimizes it over the convex set  $\mathcal{W}$ , the first-order condition gives  $J(w) - J(w^*) \geq \frac{\gamma\lambda}{2} \|w - w^*\|_2^2$  for all  $w \in \mathcal{W}$ , hence  $\|\hat{w} - w^*\|_2 \leq \sqrt{2r/(\gamma\lambda)}$ . Combining,  $r \leq \epsilon \sqrt{2r/(\gamma\lambda)}$ , i.e.  $r \leq 2\epsilon^2/(\gamma\lambda)$ , which is (21). The rate follows since  $\epsilon = O_p(N^{-1/2})$ .  $\square$

*Proof of Proposition 4.1.* Let  $E_N$  be the event  $\{\hat{\mathcal{A}}_t^0 = \mathcal{A}_t\} \cap \{\max_H |\hat{s}_t(H) - s_t(H)| < \Delta_a/2\}$ ; by (i)–(ii) and the union bound,  $\mathbb{P}(E_N) \rightarrow 1$ . On  $E_N$ , for every  $H \notin \mathcal{A}_t^*$  and  $H^* \in \mathcal{A}_t^*$ ,

$$\hat{s}_t(H) > s_t(H) - \frac{\Delta_a}{2} \geq \min s_t + \frac{\Delta_a}{2} = s_t(H^*) + \frac{\Delta_a}{2} > \hat{s}_t(H^*),$$

where the first and last inequalities use  $|\hat{s}_t - s_t| < \Delta_a/2$  and the middle one the definition of  $\Delta_a$ . Hence no  $H \notin \mathcal{A}_t^*$  minimizes  $\hat{s}_t$  on  $E_N$ , so  $\hat{\mathcal{A}}_t^* \subseteq \mathcal{A}_t^*$ . If in addition  $\hat{s}_t$  is constant on  $\mathcal{A}_t^*$ , every element of  $\mathcal{A}_t^*$  is an  $\hat{s}_t$ -minimizer and equality holds.  $\square$

*Proof of Proposition 4.2. Step 1 (value function).* Under (i),  $\rho_t^*$  is finite, nondecreasing and convex on the feasible target interval  $(-\infty, \max_{\mathcal{W}} \mu]$  by the argument of Proposition 3.21 (convexity of  $\rho$  and affinity of  $\mu$  are all that was used), and a finite convex function is continuous on the interior of its domain [56, Thm. 10.1]; by (iii) there is  $\eta_0 > 0$  with  $[m_- - \eta_0, m_+ + \eta_0]$  in that interior, so  $\rho_t^*$  is uniformly continuous there with modulus  $\omega$ . *Step 2 (sandwich).* Fix  $m \in \mathcal{M}$  and  $N$  with  $\eta_N < \eta_0$ , and write  $F(m) = \{w \in \mathcal{W} : \mu(w) \geq m\}$ ,  $\hat{F}_N(m) = \{w \in \mathcal{W} : \hat{\mu}_N(w) \geq m\}$ . Since  $|\hat{\mu}_N - \mu| \leq \eta_N$

on  $\mathcal{W}$ ,  $F(m + \eta_N) \subseteq \widehat{F}_N(m) \subseteq F(m - \eta_N)$ . Minimizing  $\hat{\rho}_N \in [\rho - \eta_N, \rho + \eta_N]$  over these nested sets,

$$\rho_t^*(m - \eta_N) - \eta_N \leq \hat{\rho}_N^*(m) \leq \rho_t^*(m + \eta_N) + \eta_N,$$

and by monotonicity  $\rho_t^*(m - \eta_N) \leq \rho_t^*(m) \leq \rho_t^*(m + \eta_N)$ . Hence  $|\hat{\rho}_N^*(m) - \rho_t^*(m)| \leq \eta_N + \omega(\eta_N)$  uniformly in  $m \in \mathcal{M}$ , and the right-hand side tends to zero in probability, which is (22). *Step 3 (minimizers).* Let  $\hat{w}_N \in \widehat{\mathcal{E}}_N(m)$  and  $\bar{w}$  be an accumulation point along a subsequence (which exists by compactness of  $\mathcal{W}$ ). Along it,  $\mu(\bar{w}) = \lim \mu(\hat{w}_N) \geq \lim(\hat{\mu}_N(\hat{w}_N) - \eta_N) \geq m$  and  $\rho(\bar{w}) = \lim \rho(\hat{w}_N) = \lim(\hat{\rho}_N(\hat{w}_N) \pm \eta_N) = \lim \hat{\rho}_N^*(m) = \rho_t^*(m)$  by Step 2, using continuity of  $\mu, \rho$  and closedness of  $\mathcal{W}$ . Thus  $\bar{w} \in F(m)$  attains  $\rho_t^*(m)$ , i.e.  $\bar{w} \in \mathcal{E}_t^*(m)$ . Conditioning on correct Stage-I selection is legitimate because its probability tends to one, and sample separation ensures that Stage-II performance cannot alter Stage-I eligibility.  $\square$

*Proof of Proposition 4.3.* On  $E_N$ , apply (24) at  $(\widehat{H}, \hat{w})$  and at  $(H^*, w^*)$  and use the optimality of  $(\widehat{H}, \hat{w})$  for  $\widehat{J}_N$ :

$$J(\hat{w}; \widehat{H}) \leq \widehat{J}_N(\hat{w}; \widehat{H}) + \varepsilon_N \leq \widehat{J}_N(w^*; H^*) + \varepsilon_N \leq J(w^*; H^*) + 2\varepsilon_N. \quad \square$$

*Proof of Theorem 4.4.* Set  $x_r := \sigma \sqrt{2 \log(2M_r/\delta)/N}$ , so that  $2 \exp\{-Nx_r^2/(2\sigma^2)\} = \delta/M_r$ . By (26) and the union bound over the  $M_r$  elements of  $\mathcal{A}_r$ ,

$$\mathbb{P}\left\{\max_{H \in \mathcal{A}_r} |\widehat{V}_N(H) - V(H)| > x_r\right\} \leq M_r \cdot \frac{\delta}{M_r} = \delta.$$

On the complementary event, for every  $H^\circ \in \arg \min_{\mathcal{A}_r} V$  (which exists because  $\mathcal{A}_r$  is finite),

$$V(\widehat{H}_r) \leq \widehat{V}_N(\widehat{H}_r) + x_r \leq \widehat{V}_N(H^\circ) + x_r \leq V(H^\circ) + 2x_r,$$

which is (27). For (28),  $\min_{\mathcal{A}_t(i)} V = V_i^*$  and  $V(\widehat{H}_i) - V_c^* = (V(\widehat{H}_i) - V_i^*) + (V_i^* - V_c^*) \leq 2x_i + \Delta_{ci}$  by (27) and (18). Deviation inequalities of the form (26) hold for bounded or truncated losses by Hoeffding's inequality [13, 33]; under a finite-variance assumption only, they hold for the estimated value computed with a robust mean estimator in place of the sample mean [17]. Heavy-tailed portfolio losses such as those of experiment E3 are outside the hypotheses of the theorem unless one of these devices is used, and E3 is reported as a comparison of designs rather than as an instance of the bound.  $\square$

*Proof of Corollary 4.5.* Both bounds are upper bounds on the excess loss over the common oracle  $V_c^*$ ; the conditional-law bound is  $2x_c$  and the interventional bound is  $\Delta_{ci} + 2x_i$ . The inequality  $\Delta_{ci} + 2x_i < 2x_c$  is (29) after substituting  $x_r$ .  $\square$

*Proof of Proposition 4.6.* (a) Reselection means  $\hat{s}_t(\widehat{H}_{t-1}) > \min \hat{s}_t + \eta$ ; since  $\hat{s}_t(\widehat{H}_{t-1}) \leq s_t(\widehat{H}_{t-1}) + \delta_t$  and  $\min \hat{s}_t \geq \min s_t - \delta_t$ , we obtain  $s_t(\widehat{H}_{t-1}) - \min s_t > \eta - 2\delta_t$ . (b) If the incumbent is population-optimal, the left side of (a) is zero, so (a) would require  $\eta < 2\delta_t$ . (c) For  $\widehat{H}_t \in \arg \min \hat{s}_t$  and  $H^* \in \arg \min s_t$ ,  $s_t(\widehat{H}_t) \leq \hat{s}_t(\widehat{H}_t) + \delta_t \leq \hat{s}_t(H^*) + \delta_t \leq s_t(H^*) + 2\delta_t$ .  $\square$

## A continuous candidate class

**Proposition A.1** (A continuous candidate class). *Let  $(R, Z)$  be jointly Gaussian with  $Z \in \mathbb{R}^k$ ,  $\text{Cov}(Z) \succ 0$  and  $\text{Cov}(R | Z)$  diagonal, so that  $Z$  is a separator. For  $\theta$  on the unit sphere  $S^{k-1}$  let  $H_\theta := \sigma(\mathcal{G}_t^0, \theta^\top Z)$  and let  $\mathcal{W}(H_\theta) \equiv \mathcal{W}$  be a fixed nonempty compact set. Then the quotient of  $\{H_\theta : \theta \in S^{k-1}\}$  by information equivalence is compact,  $\theta \mapsto \mathcal{W}(H_\theta)$  is upper hemicontinuous, the screening defect  $\theta \mapsto \varepsilon_t(H_\theta)$  and the conditional value  $\theta \mapsto V(H_\theta)$  are continuous, and both the envelope problem (7) and the Stage-I problem have solutions.*

*Proof.* Information equivalence identifies  $\theta$  with  $-\theta$  and no other pairs, since  $\sigma(\theta^\top Z) = \sigma(\theta'^\top Z)$  for a nondegenerate Gaussian  $Z$  if and only if  $\theta' = \pm\theta$ ; the quotient is the projective space  $S^{k-1}/\{\pm 1\}$ , which is compact. A constant correspondence is upper hemicontinuous. By Gaussian conditioning,  $\text{Cov}(R \mid H_\theta) = \text{Cov}(R) - \text{Cov}(R, \theta^\top Z) \text{Cov}(\theta^\top Z)^{-1} \text{Cov}(\theta^\top Z, R)$ , a rational function of  $\theta$  with denominator  $\theta^\top \text{Cov}(Z) \theta \geq \lambda_{\min}(\text{Cov}(Z)) > 0$ , hence continuous on the sphere, as is  $\mu_{H_\theta}$ ; the maximal absolute off-diagonal correlation is a continuous function of a covariance with positive diagonal, so  $\varepsilon_t$  is continuous, and  $V(H_\theta) = \min_{w \in \mathcal{W}} J(w; H_\theta)$  is continuous by Berge’s maximum theorem [1, Thm. 17.31] since  $J$  is continuous in  $(w, \theta)$ . Theorem 3.18 and Theorem 3.12 apply.  $\square$

## B Procedure, experiment configuration and robustness

### Procedure

Fix the decision contract (universe, horizon, timestamps, constraints, loss, claim level with its regime set, cost model, tolerances). Generate candidates from any source; the source confers no status. Apply the required admissibility conditions, collapse information- and decision-equivalent candidates, estimate the four admissibility quantities on the certification block, form the minimal class, apply pre-declared tie-breaks and fix the result. Estimate the conditional primitives of the fixed class on a disjoint block and compute the efficient set and frontier; under the envelope contract, decision loss on a block disjoint from certification may rank certified classes instead. Evaluate once on untouched data. The declared conditions, tolerances and tie-breaks are part of the model documentation and are reviewed independently of the certification data.

### Reproducibility

All controlled experiments are generated by a single script with fixed seed 20260906 (NumPy default generator); the complete run takes under fifteen minutes on one processor core, and every number quoted in Section 4 is read from the script’s output file rather than transcribed. Synthetic data are reproducible from the script; the market panel of the pre-registered design is licensed and is not redistributed.

### Robustness

Three further sampling seeds with the same data-generating structure give, for E2 at 150 replications, class recovery in  $[0.91, 0.93]$  at  $N = 500$  and  $[0.98, 0.99]$  at  $N = 1,000$ ; for E3 at 100 replications, proxy retention by the conditional-law contract in  $[0.43, 0.49]$ , admission of  $Z$  and  $P$  by the interventional admissibility condition equal to one and zero in every seed, and post-intervention losses in  $[0.155, 0.156]$  for the proxy-fixed and  $[0.136, 0.137]$  for the state-fixed portfolio. Halving the declared tolerance to 0.05 raises false exclusion to 0.25, 0.19 and 0.05 at  $N = 500, 1,000, 2,000$  with recovery 0.75, 0.81 and 0.95; raising it to 0.15 gives recovery 0.95, 1 and 1 with false admission 0.047, 0 and 0. Calibrating the tolerance to the null distribution of the maximum over the 28 asset pairs (Šidák multiplier  $3.12/\sqrt{N}$  in place of  $2/\sqrt{N}$ ) removes false exclusion entirely but admits low-dimensional inadmissible candidates at small samples, with false admission 0.72, 0.14, 0.007 and 0 at  $N = 250, 500, 1,000, 2,000$ ; a percentile bootstrap of the maximum statistic is biased upward and excludes the true class at every sample size considered. The multiplier-bootstrap calibration of the maximum [19], with candidate-specific critical values at level 0.025 and 100 multiplier draws, gives class recovery 0.07, 0.69 and 1 at  $N = 250, 500, 1,000$ , with false exclusion zero throughout and

false admission 0.93 and 0.31 at the two smallest sizes: it controls the exclusion error by construction and pays for it in admissions at small samples, exactly as the Šidák rule does. The per-pair Fisher tolerance of the main text is therefore a compromise between the two error types, and the sensitivity of the rates to it is the finite-sample content of the gap condition in Proposition 4.1.

## E1–E2

Eight assets, two common states  $Z \sim N(0, I_2)$ ; loadings  $B \in \mathbb{R}^{8 \times 2}$  drawn once from  $N(0, 0.6^2)$ ; idiosyncratic scales  $\sigma_i \sim U[0.5, 1]$ ; intercepts  $\mu_{0,i} \sim U[0.02, 0.10]$ ; candidate variables  $(Z_1, Z_2, \phi(Z_1) = 3Z_1 - 2, P = Z_1 + 0.7e, \nu, X^+ = u_1 + u_2)$  with  $e, \nu \sim N(0, 1)$  independent; candidate set: all nonempty subsets of size  $\leq 3$  (41 candidates); chronological admissibility condition from timestamps; defect  $\varepsilon_t =$  maximal absolute conditional residual correlation (analytic in E1; sample correlation of least-squares residuals in E2); tie-break by dimension then  $\hat{\varepsilon}_t$ ;  $\gamma = 5$ ; E2 sample sizes  $N \in \{250, 500, 1000, 2000, 4000, 8000\}$ , 300 replications, 50/25/25 split, tolerance  $0.10 + 2/\sqrt{N_{\text{cert}}}$  (declared level plus the Fisher 97.5% bound), frontier error on 21 targets covering the central 60% of  $[\min \mu_0, \max \mu_0]$ , conditional mean evaluated at  $Z = 0$  through each representation’s variables.

## S1: risk contract at fixed information (results)

Returns follow a one-state model with Student- $t_5$  innovations, state-dependent scale and negative jumps on three assets; the Stage-I class is fixed at the true state, and Stage II is solved under variance and under expected shortfall at level 0.05 by the linear program of [57] at three targets, then evaluated on 20,000 fresh draws from the same conditional law. Table 7 shows that the two contracts produce different portfolios ( $\ell_1$  weight distance 0.06–0.17) and that the expected-shortfall optimizer does not attain lower out-of-sample expected shortfall than the variance optimizer at either training size. This is the known fragility of sample expected-shortfall optimization [42], and it is the point of the experiment: changing the risk contract inside a fixed admissible class changes the portfolio but confers no dominance, so the risk contract and the information contract are orthogonal.

Table 7: S1: out-of-sample mean, variance and expected shortfall at level 0.05 (ES) of variance-optimal and expected-shortfall-optimal portfolios in the same fixed information class.

| $N_{\text{train}}$ | target | variance-optimal |          |       | expected-shortfall-optimal |          |       |
|--------------------|--------|------------------|----------|-------|----------------------------|----------|-------|
|                    |        | mean             | variance | ES    | mean                       | variance | ES    |
| 2,000              | -0.033 | -0.036           | 0.190    | 0.994 | -0.037                     | 0.191    | 1.002 |
| 2,000              | 0.045  | 0.041            | 0.185    | 0.932 | 0.041                      | 0.188    | 0.949 |
| 2,000              | 0.274  | 0.264            | 0.382    | 1.203 | 0.264                      | 0.392    | 1.238 |
| 10,000             | -0.042 | -0.039           | 0.185    | 1.001 | -0.040                     | 0.188    | 1.026 |
| 10,000             | 0.027  | 0.029            | 0.183    | 0.951 | 0.029                      | 0.187    | 0.983 |
| 10,000             | 0.278  | 0.281            | 0.426    | 1.275 | 0.280                      | 0.438    | 1.331 |

## S1 (configuration)

One state,  $z_0 = 0.5$ ; scale  $\exp(0.3z_0)$ ; Student- $t_5$  innovations; negative exponential jumps of mean  $3\sigma_j$  with probability 0.04 on assets 1, 4, 7; training sizes 2,000 and 10,000; test size 20,000; targets at the 0.35, 0.6, 0.85 quantiles of the estimated means; weights in  $[-1, 1]$ , full investment; expected shortfall at level 0.05 by the Rockafellar–Uryasev linear program.

### E3

One state  $Z \sim N(0, 1)$ , proxy  $P = Z + 0.05e$ ;  $N = 4,000$  per regime; intervention replaces  $P$  by an independent  $N(0, 1)$ ; interventional admissibility condition: maximal absolute change of conditional slopes across regimes below  $6\sqrt{\log n/N}$  and change of  $\hat{\varepsilon}_t$  below 0.10; portfolios  $w \propto \hat{\Sigma}_H^{-1} \hat{\mu}_H$  normalized to full investment from pre-intervention moments; post-intervention loss  $\frac{\gamma}{2} w^\top \Sigma w - \mu_0^\top w$  under the true post-intervention conditional covariance ( $\text{diag}(\sigma^2)$  given  $Z$ ;  $bb^\top + \text{diag}(\sigma^2)$  given the uninformative proxy); 200 replications.

### E2 (regret addendum)

Regret of the estimated conditional mean–variance portfolio at  $\gamma = 5$  under the true moments, on the evaluation block of the true class, with the sample mean and with Jorion shrinkage toward the GMV-implied grand mean, intensity  $(n + 2)/((n + 2) + T(\hat{\mu} - \hat{\mu}_g)^\top \hat{\Sigma}^{-1}(\hat{\mu} - \hat{\mu}_g))$ ; 300 replications.

### Market data (Section 4.7)

Equity panel: twelve U.S. equities, one per sector, and 127 economic and financial series, daily 2001–2023, 5,810 aligned observations; asset prices enter as log differences, drivers as log differences when strictly positive and first differences otherwise. Windows of 756 sessions stepped by 63 (81 windows), split 50/25/25 into certification, estimation and evaluation. Statistical condition: maximal absolute residual correlation across assets after least-squares conditioning, tolerance 0.10 plus a Šidák-calibrated bound for the maximum over the  $n(n - 1)/2$  pairs (0.273 for  $n = 12$ ); candidate search by greedy forward selection up to eight drivers. Stage II:  $\hat{B}\hat{V}\hat{B}^\top + \text{diag}(\hat{\Omega})$  from the estimation block; global-minimum-variance portfolios, unconstrained and long-only (projected gradient with exact simplex projection); baselines sample covariance, matched-dimension principal components with diagonal residual, Ledoit–Wolf linear and nonlinear shrinkage, equal weights; paired differences tested by circular block bootstrap (block length 5, 999 draws). Hedge-fund panel: seventeen strategy indices, monthly, 267 months, drivers aggregated to month end, windows of 120 months stepped by six, forward selection up to four drivers, screening only. Universe grid: 48 industry, 25 size/value, 25 European and 25 Developed size/momentum portfolios from Kenneth French’s library, aligned to the driver calendar. Aggregation design: 147 equities with complete coverage 2010–2024, twelve assets each an equally weighted basket of  $k \in \{1, 2, 4, 8, 12\}$  names, six random draws per design. Candidate-search design: random candidate subsets of size 4 to 127, eight draws per size. Lag test: drivers shifted by one, two and five sessions.

### E4

Forty nodes; per node  $N = 200$  with the first 100 observations for certification; loadings as in E1–E2 with the second state switched from  $Z_2$  to an independent  $Z_3 \sim N(0, 1)$  at node 20; candidates: subsets of size  $\leq 2$  of  $\{Z_1, Z_2, Z_3, P, \nu\}$ ; score  $\hat{\varepsilon}_t + 0.02\kappa_t$ , the weight being one fifth of the declared tolerance, the smallest value at which one additional dimension outweighs a defect difference inside the tolerance band; with weights 0.01 and 0.05 in place of 0.02 (60 replications) the reselection counts at  $\eta \in \{0, 0.10, 0.20\}$  are 5.8, 1.2, 1.1 and 6.3, 1.3, 1.1, the fraction of nodes on the optimal class 0.93, 0.99, 0.97 in both cases, and the lags 0.07, 0.13, 0.67 and 0.08, 0.13, 0.67, so the conclusions do not depend on the weight; margins  $\eta \in \{0, 0.02, 0.05, 0.10, 0.20\}$ ; 100 replications; reported: mean reselections, fraction of nodes on the population-optimal class, mean lag after the change.